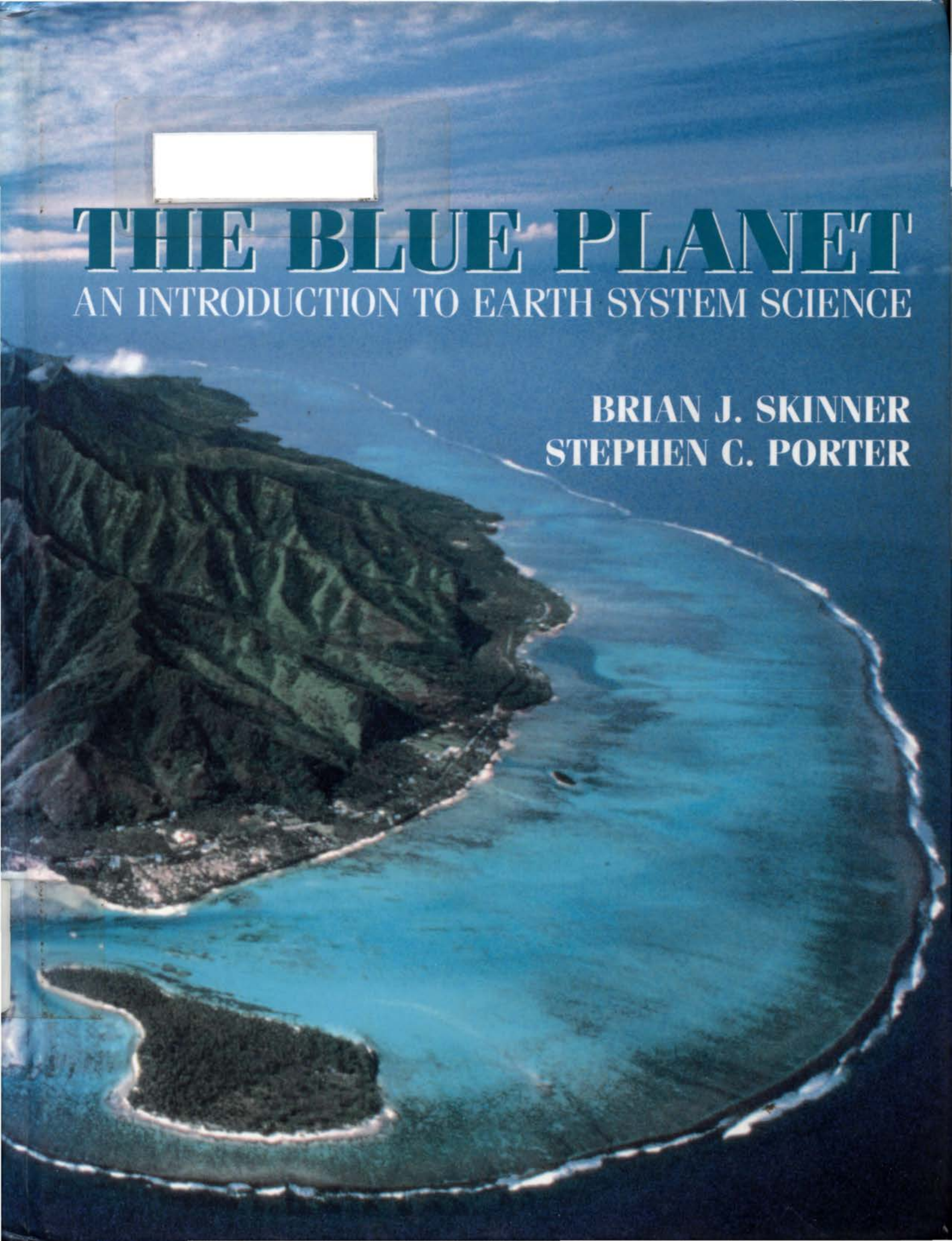
THE BLUE PLANET

AN INTRODUCTION TO EARTH SYSTEM SCIENCE

BRIAN J. SKINNER
STEPHEN C. PORTER

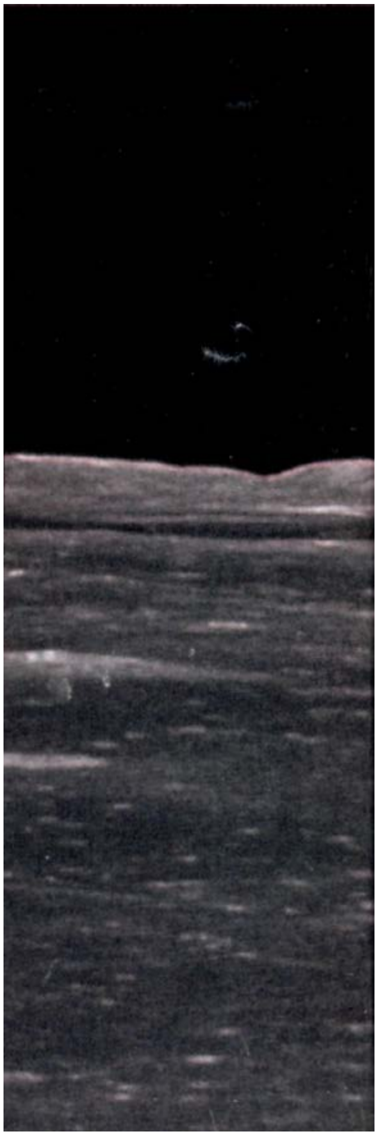


The Blue Planet

John Wiley & Sons, Inc.

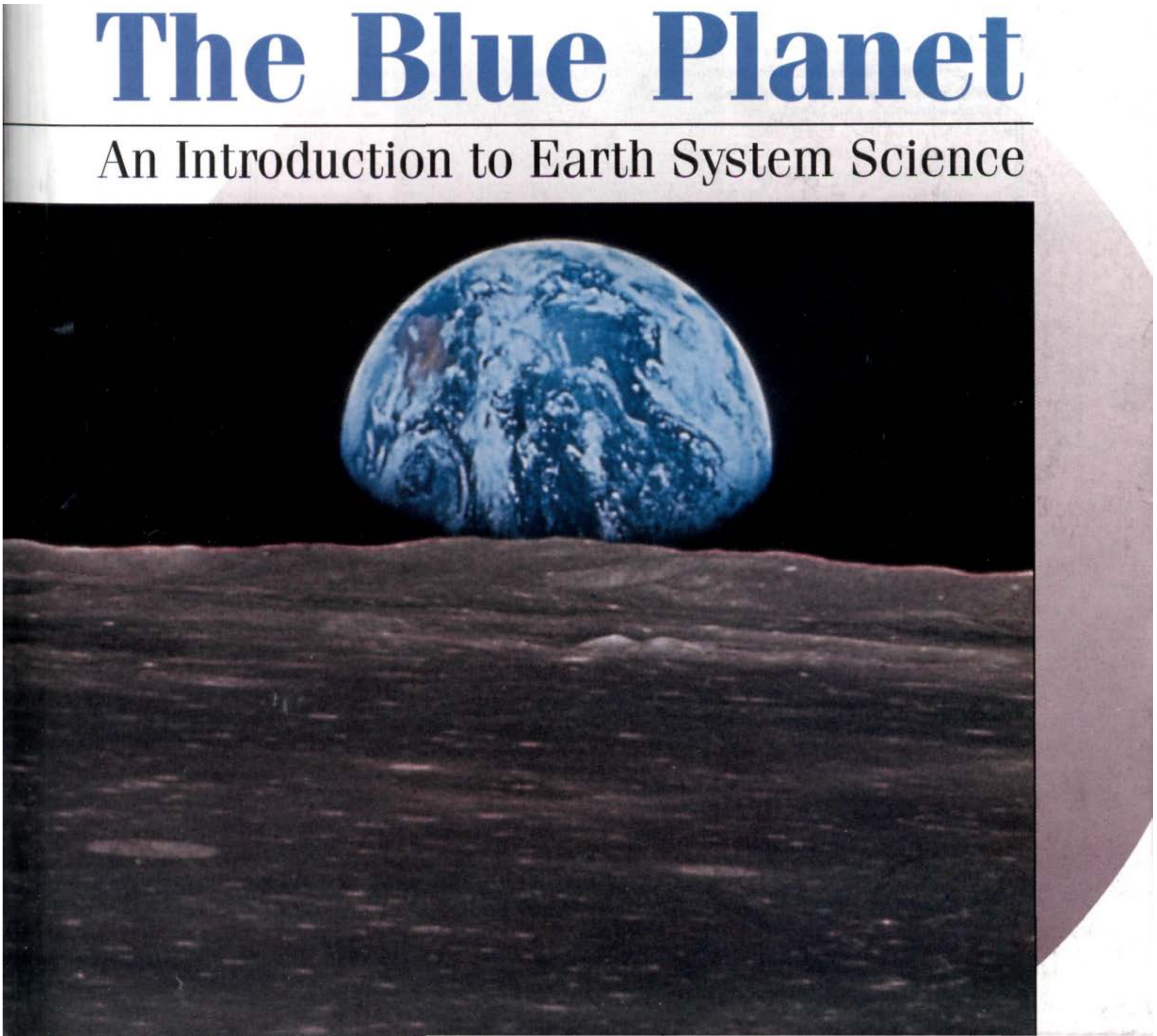
New York
Chichester
Brisbane
Toronto
Singapore

Photographed by an Apollo astronaut, Earth, the blue planet, rises over the stark, gray, lunar landscape. The Earth appears to be blue just as the sky seems blue on a bright, sunny day. The cause is the same in both cases, an optical effect of the atmosphere. We know that air is colorless, not blue, but the atmosphere appears to be blue when viewed from a distance, due to scattering of light rays by molecules of gas in the atmosphere. Exactly how the scattering occurs is discussed in Chapter 8.



Brian J. Skinner
Yale University

Stephen C. Porter
University of Washington



ACQUISITIONS EDITOR:	Christopher Rogers
DEVELOPMENTAL EDITOR:	Barbara Heaney and Irene Nunes
MARKETING MANAGER:	Catherine Faduska
SENIOR PRODUCTION EDITOR:	Sandra Russell
SENIOR DESIGNER:	Kevin Murphy
MANUFACTURING MANAGER:	Susan Stetzer
DIRECTOR OF PHOTO DEPARTMENT:	Stella Kupferberg
PHOTO EDITOR:	Alexandra Truitt
PHOTO ASSISTANT:	Michelle Orlans
SENIOR ILLUSTRATION COORDINATOR:	Sigmund Malinowski
ILLUSTRATIONS:	J/B Woolsey Associates
COVER PHOTO:	© David Hiser/Photographers/Aspen.

This book was set in 10/12 ITC Garamond Book by Ruttle, Shaw & Wetherill, Inc. and printed and bound by Von Hoffman Press, Inc. The cover was printed by The Lehigh Press, Inc. The color separations were prepared by Color Associates, Inc.

John Wiley & Sons, Inc., places great value on the environment and is actively involved in efforts to preserve it. Currently, paper of high enough quality to reproduce full-color art effectively contains a maximum of 10% recovered and recycled post-consumer fiber. Wherever possible, Wiley uses paper containing the maximum amount of recycled fibers. In addition, the paper in this book was manufactured by a mill whose forest management programs include sustained yield harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

TOTAL 10% RECYCLED PAPER
ALL POST-CONSUMER WASTE

Library of Congress Cataloging in Publication Data:
Skinner, Brian J., 1928—
The blue planet : an introduction to earth system science / Brian J. Skinner, Stephen C. Porter.
Includes bibliographical references.
ISBN 0-471-54021-8
1. Earth. 2. Earth sciences. I. Porter, Stephen C. II. Title.
QB631.S57 1994

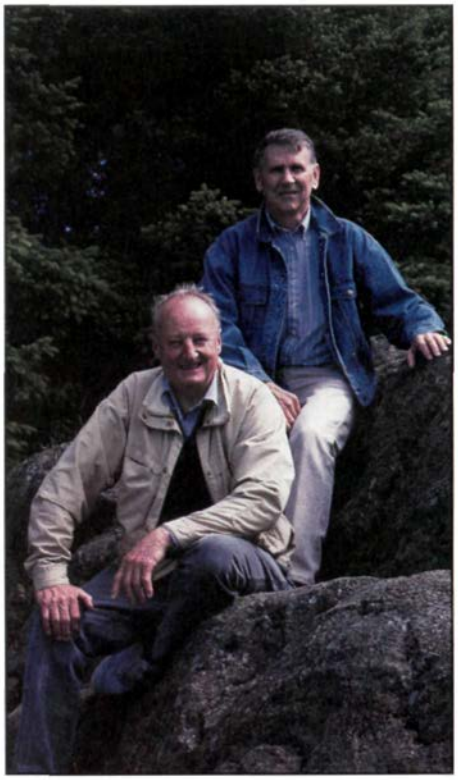
550—dc20
94-25686
CIP

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ABOUT THE AUTHORS

1 was born and raised in small country towns in Australia and my early rural experiences did a lot to shape the way I think about the Earth and the way I address problems. I learned to look closely at the world around me, for example, and I saw the giant changes in the landscape as unspoiled bushland was cleared for sheep pastures.

By my twelfth birthday I knew that I wished to be a scientist and the science I fixed my mind on was chemistry. Accordingly, when I entered the University of Adelaide in South Australia in 1946, I selected a course of studies leading to a degree in chemistry. Things never go exactly as planned. Because I had a great-grandfather who had what seemed to me a very mysterious profession, mining engineering, and who had worked at the famous copper mines at Moonta in South Australia, I elected to take a course in geology. My motivation was to find out what had interested my ancestor. What I quickly discovered was my own interest in the subject. I completed my degree in chemistry but I also completed the requirements for graduation in geology.



Brian J. Skinner Stephen C. Porter

After experience as a member of an exploration team looking for lead and zinc deposits, and employment as a mining geologist at a tin mine in Tasmania, I entered graduate school at Harvard University. When I emerged $3^1/_2$ years later with a PhD I returned to Australia and started teaching and research on the origin of mineral deposits, a subject that continues to occupy much of my time and energy. An unexpected offer to become a member of the research staff of the U.S. Geological Survey brought me back to the United States in 1960.

After a number of years in the USGS I moved to the Department of Geology and Geophysics at Yale University and there I have stayed and worked to the present. It was at Yale that I found opportunities to explore the wider aspects of the earth sciences that increasingly occupied my thoughts. For a number of years I was involved in the space program, first the lunar landing experiments and later the unmanned landing experiments on Mars. With Yale colleagues I worked on problems involving oceanography and climatic change, and on such diverse topics as volcanic gases and economic models of resource depletion. Through exposure to such a diverse range of topics I came to understand that everything on the Earth is interconnected and that the way to appreciate how this wonderful planet of ours works is to understand the system by which the disparate parts interact with each other. It was my search for understanding that led to my participation in the preparation of this book about Earth System Science.

B.J.S.

VJrowing up on the dynamic coast of southern California, I was introduced to geology in action at an early age, an introduction that included being thrown out of bed and across my room by a major earthquake that jolted Santa Barbara in the early 1940s. Although earth science was not part of the curriculum when I attended high school, sumnmers spent trekking in the High Sierra and the Rocky Mountains awakened in me an interest in mountains that gained momentum in college when rock climbing and mountaineering occupied much of my spare time. At Yale, a distinguished faculty introduced me to the science of geology, and after serving aboard a Navy destroyer in the Pacific Fleet, I returned to Yale for graduate study. There I focused on glacial and Quaternary geology under the guidance of Professor Richard F. Flint, for many years a prominent author of Wiley geology textbooks. I had the chance to "prove" myself during three long summers of field work in the Brooks Range of Arctic Alaska where I worked at unraveling the puzzling complexities of a thick section of intensely deformed sedimentary rocks, without the benefit of plate tectonics to guide the way, and also studied the glacial history of the range.

A faculty position at the University of Washington provided an ideal setting for someone with my interests in glaciers and the ice ages. Within a day's drive of the campus, our earth science students can be introduced in the field to nearly every topic they will hear or read about in class.

The research on mountain glaciation I began in western North America expanded to include many of the world's glaciated mountains: the Alps, the Andes, the Himalaya, New Zealand's Southern Alps, and tropical Mauna Kea on the island of Hawaii. In addition to glacial-geologic studies, my work has also involved analyzing rockfall hazards in the Italian Alps, prehistoric cave sites in northern Spain that served as home to ice-age hunters, and prehistoric eruptions of Cascade and Hawaiian volcanoes. My current research includes collaborative studies with Russian colleagues on mountain glaciation in northeastern Siberia and with Chinese colleagues on the monsoon history of central China and Tibet recorded in deposits of wind-blown dust and ancient soils. Such experiences, together with my duties as editor of *Quaternary Research,* an international interdisciplinary journal on the glacial ages, enable me to keep in touch with other earth scientists throughout the world and help me stay abreast of new advances in a rapidly moving scientific field.

S.C.P.

PREFACE

When historians of the future consider the most important achievements of the 20th century, the chances are that one of the selections will be the development of a holistic view of the Earth. Such a view has come about as a result of the technological and scientific advances that have made it possible to measure the many ways that the different parts of the Earth can interact. For example, how events deep inside the Earth can influence events on the Earth's surface, or how small changes in the water temperature of the ocean can bring about changes in the distribution of land plants and how those changes can lead to the evolution of new plant species.

One of the discoveries that arise from the holistic view of the Earth is that our modern industrial society and our huge population are changing the Earth. We can now measure in real time the innumerable ways by which we humans are changing the global environment as a result of our collective activities.

This book is an introduction to the holistic view of the Earth. It is about the interactions between the different parts of the Earth—the atmosphere, hydrosphere, biosphere, and the solid Earth—and about the balance in the global environment that exists as a result of those interactions. It is a book that presents a new view of the Earth that has come to be called *Earth System Science.*

Purpose of the Book

Earth system science is rapidly changing the way we study and think about the Earth and as a result it is changing the way earth science courses are being taught. We have written this book in order to introduce students to the science of the Earth system.

Courses about the Earth are being taught with increasing frequency. Such courses may have titles such as global change, earth science, biospherics, or even the global environment, but the approach is increasingly that of Earth System Science.

The Book's Organization

The text begins with a discussion of the Earth's place in the solar system and contrasts the Earth's appearance and structure with those of neighboring planetary bodies. A chapter is devoted to the Sun, not only because it is the dominant feature of the solar system, but because it supplies the energy that drives most of the surface processes on our planet and that permits life to exist. Next, we discuss the solid Earth, the minerals and rocks that comprise it, the nature of processes operating deep within the Earth that are inferred indirectly, and the dynamics of the crust that are explained in terms of a relatively new, comprehensive theory, the theory of plate tectonics. Having explored the solid Earth beneath our feet, we next examine the layers of water and ice that cover much of its surface: the oceans, streams, groundwater, snow, glaciers, sea ice, and frozen ground, and we explore how some of these different agents erode and shape landscapes on which we live. In the third part of the book we explore the atmosphere, weather, and climate, and examine the evidence of past changes in climate on various time scales. Having discussed the aspects of the Earth that have made it a habitable planet, we next look at the diversity and dynamics of plants and animals comprising the biosphere and the evidence of biological evolution through Earth history that is recorded in fossiliferous rocks. In the final section of the book, we look at natural resources that have permitted the development and growth of modern civilization, and ways in which human activities contribute to global changes in our environment.

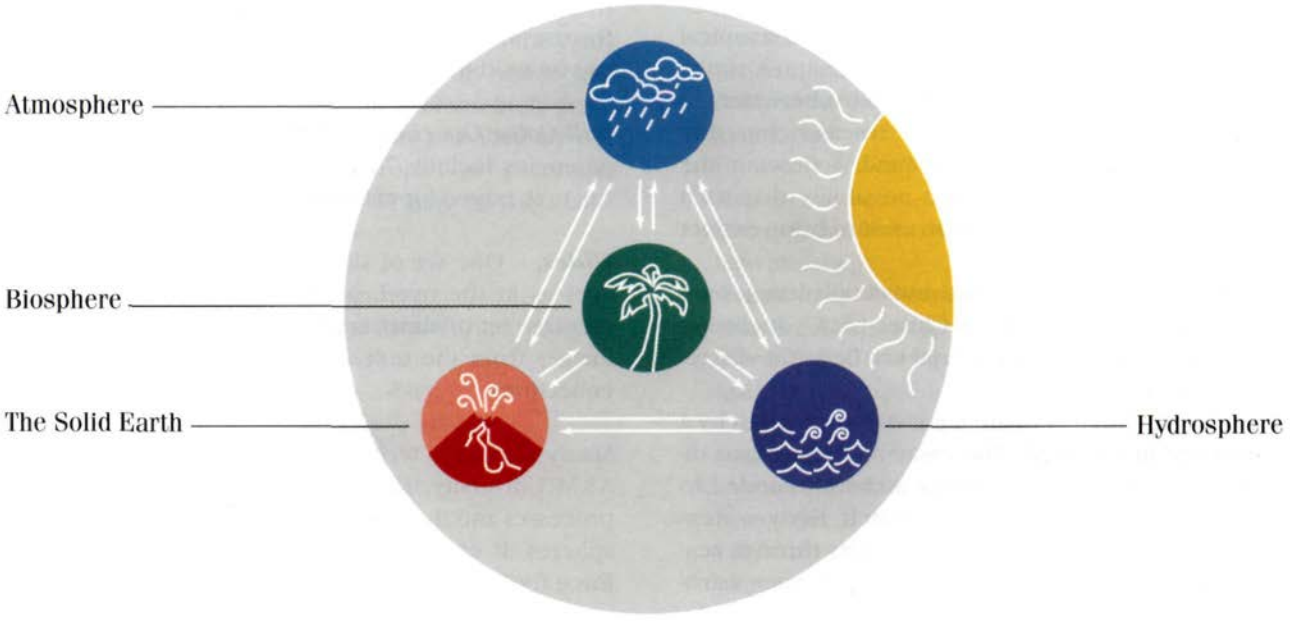
While we have given careful consideration to the organization of the book we realize that not all instructors will favor the one we have adopted. Therefore, the parts and chapters have been written so that some reorganization of topics is possible without serious loss of continuity. For example, the chapters on the solar system could be assigned toward the end of a course, rather than at the beginning, and the chapters on the biosphere could be shifted, or omitted. Where aspects of astronomy or biology are important in discussion of the solid Earth and its surface environments, they are included there as well.

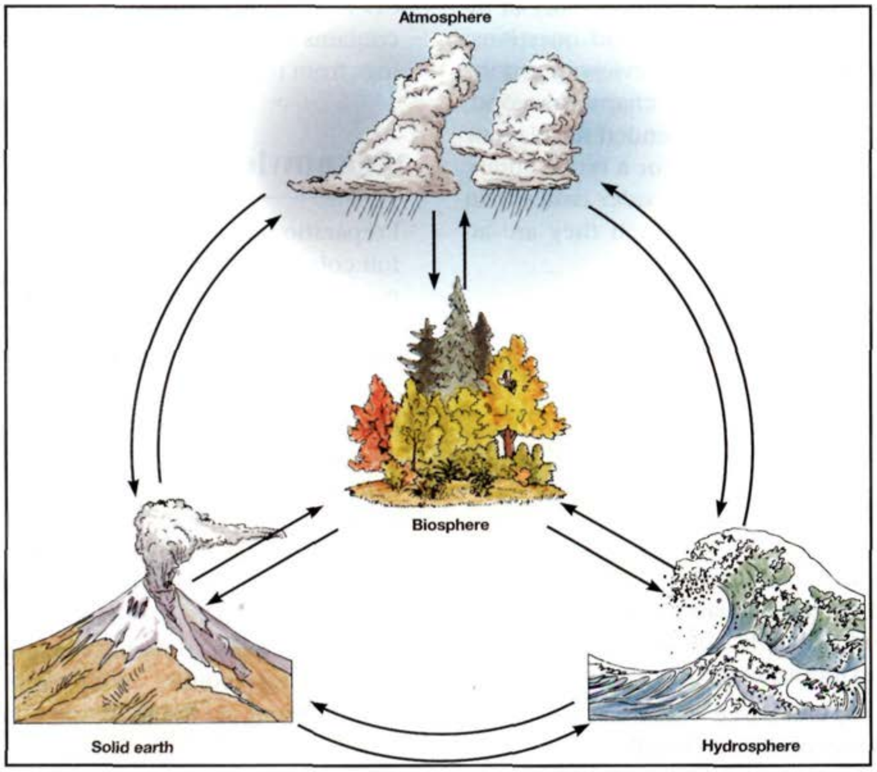
The Artwork

Special attention has been devoted to producing artwork and photographs that illuminate discussions in the text. Because no continent or country holds a monopoly on relevant and interesting examples, we have attempted to provide photographs, maps, and illustrations from around the world to provide a global perspective of Earth System Science. The art program has benefitted from talented artists who have worked closely with the authors to make their illustrations both attractive and scientifically accurate. Two hundred and twelve photographs and line drawings, all in full color, have been selected.

THE SYSTEMS APPROACH

The key to understanding the Earth system is an appreciation of the interactions between the spheres. Interactions are emphasized by the four icons shown below, each of which represents a part of the system. The icons appear throughout the book and help the reader identify sections in the text where interactions are discussed. For example, in the discussion of streams and drainage systems in Chapter 9, the icons for water, air and land are introduced to emphasize the interactions between those three parts of the Earth system.





Features

Each of the six parts in the book opens with a brief essay on a special aspect of the topics covered within the chapters of the part. The intent of the essays is to emphasize that the different parts of the Earth system are interdependent. For example, Part V, *The [Dynamics of Life,* opens with an essay on how some migrating birds find their way by using the stars while others employ the magnetic field and still others use the polarization of sunlight.

Each chapter in the book also opens with a topical essay dealing with research on the chapter topic. Chapter 16 on the evolution of the biosphere, for example, opens with an essay on research into the preservation of DNA in fossils and, following the theme from *furassic Park,* the possibility that such ancient DNA might some day be used to bring extinct species back to life.

Within chapters specialized and detailed topics are boxed under the heading "A Closer Look". Inclusion or deletion of the boxed material can be at the discretion of the instructor.

Each chapter closes with a guest essay written by a researcher in the field. The essay subjects relate directly to material in the chapter and are intended to provide insights into on-going research. Essay writers range from scientists working in industry through academic and government scientists to a former astronaut.

Finally, each chapter closes with a summary of in-chapter material, a list of key terms, and questions. The questions are of two kinds: first, review questions relate strictly to the material in the chapter; second, discussion questions, which are intended for class or section discussion, sometimes call for a bit of library research, and in most cases raise broader issues than those in the specific chapter to which they are attached.

Supplements

A **full** range of supplements to accompany *The Blue Planet* is available to assist both the instructor and the student.

Laboratory Manual Written by Marcia Bjornerud of Miami University, Ohio, the laboratory manual is divided into six modules with labs available for every part of the text, flexibly arranged to cover all topics. It includes many computer-based exercises and hands-on activities that students can do locally.

Instructor's Manual and Test Bank. Written by Barbara Murck, of Toronto University, this guide includes a table of contents, chapter summaries references, and approximately 85 test questions per chapter.

Computerized Test Bank. A computerized test bank is available in both IBM-compatible and Macintosh versions. This easy-to-use test generating program enables instructors to choose test questions from the printed test bank, print the completed tests for use in the classroom, and save the tests for later use on modification.

Full-Color Overhead Transparencies. The transparencies include 75 line drawings and tables from the text, edited for maximum classroom effectiveness.

Slides. One set of slides includes the same 75 images as in the overhead transparency set; a supplementary set of slides is also available, containing 150 images from the text and from the authors' private collections.

Study Guide. Written by Michael Jordan, of Texas A&M University, Kingsville, the study guide stresses processes and the interconnections among the Earth's spheres. It contains questions and exercises to reinforce the systems approach.

CD-ROM. A CD-ROM is available to adoptors and contains a compilation of photographs and line drawings from the Skinner & Porter texts.

Acknowledgments

Preparation of a book, especially the first edition of a full color text, needs the help and expertise of many people. We are greatly indebted to our publisher and our editor, initially Barry Harmon and subsequently Christopher Rogers, for supporting us, for understanding when other demands meant we had to put *Blue Planet* aside for a time, and for maintaining an encouraging level of confidence that we would finally produce.

Above all we are indebted to Brent Peich for his skillful coordination and editing of the Guest Essay our freelance developmental editor, Irene Nunes, who ironed out our syntax and undangled our dangling participles; to Barbara Heaney, our in-house developmental editor, who kept things flowing smoothly and did the thousand things we overlooked.

Many of our professional colleagues have provided a great deal of help and support. Two must be singled out for especial thanks. Bryan Gregor of Wright State University provided valuable help with the chapters on the biosphere and geochemical cycles, and Jill Schneiderman, initially at Pomona College and now at Vassar, read all the galleys, catching many errors in the process and allowing us to concentrate on illustrations and other unfinished business.

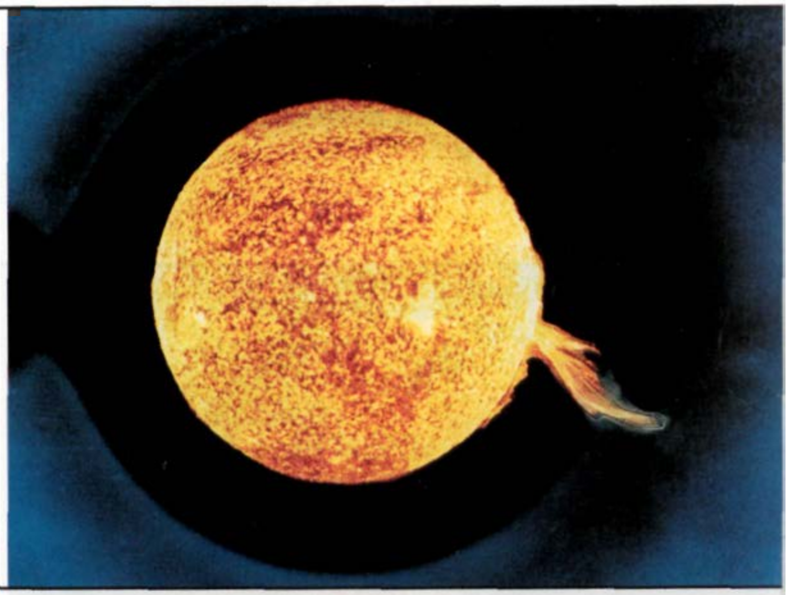
Our many colleagues who prepared part or chapter closing essays did so with grace and professional acumen. We are very grateful to them. Besides the essay writers, each of whom is identified by name and photo adjacent to their entry, we are extremely grateful for the guidance and judgment provided by colleagues who discussed this project in encounter groups and who reviewed all or part of the manuscript. They are:

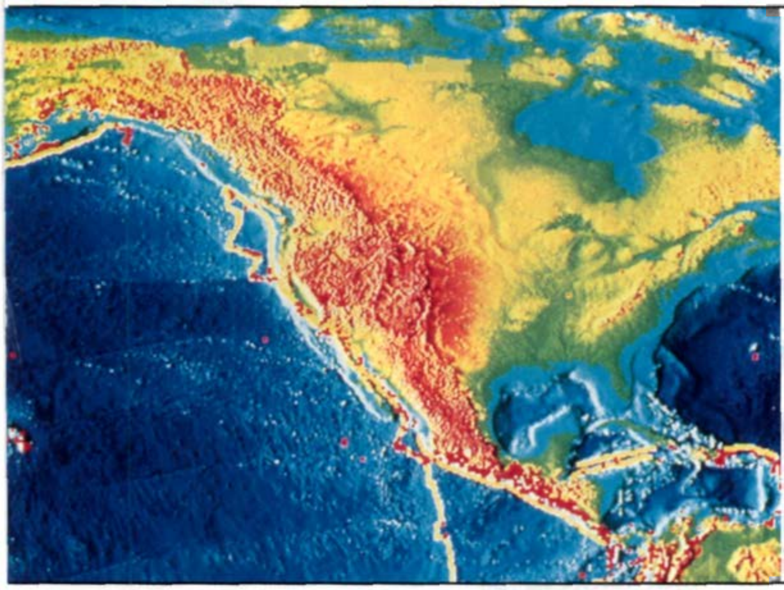
Bud Alcock
Pennsylvania State University, Ogontz
DeWayne Backhus
Emporia State University
Kathe Bertine
San Diego State University
Marsha Bollinger
Winthrop University
Phillip Boger
State University College, Geneseo
John A. Campbell
Fort Lewis College
Lindgren Chyi
University of Akron
George S. Clark
University of Manitoba
Brent Dugolinsky
State University of New York at Oneonta
John Ernissee
Clarion University
Lynnette Flann
Wichita State University
Katherine Freeman
Pennsylvania State University
James Henry
Tennessee State University
Norman A. Higginbotham
Southeastern Louisiana University
Robert Howe
Indiana State University
Rudi Kiefer
University of North Carolina, Wilmington
Robert Lawrence
Oregon State University
John Madsen
University of Delaware
Charles F. Martin
Bowie State University
Rama Murthy
University of Minnesota
John Mylroie
Mississippi State University
James Shea
University of Wisconsin, Parkside
John Southam
University of Miami
Brooke Towery
Pensacola funior College
Felicia West
University of North Florida

BRIEF CONTENTS

CONTENTS

PART ONE





PART TWO

PART FIVE

The Dynamics of Life /388

PART SIX

Living on Earth /442

GUEST ESSAYISTS

Introduction
William S. Fyfe,
University of Western Ontario
Toward Global Responsibility: Earth Sense

Chapter 1
Harrison H. Schmitt,
Consultant
"The First Stepping Stone in Space"

Chapter 2
Barbara L.Weither,
Harvard-Smithsonian Center for Astrophysics
"Profile: Maria Mitchell"

Chapter 3
Jeffrey Park,
Yale University
"Rethinking Earthquake Prediction"

Chapter 4
Jill S. Schneiderman,
Vassar College
"Asbestos: To Understand the Science Is to Understand the Politics"

Chapter 5
Kathleen Crane,
Hunter College, Lamont-Doherty Earth Observatory
"Lake Baikal, One of the World's Underwater Wonders"

Chapter 6
Charles Drake,
Dartmouth College
Profile: "Harry Hess and Global Tectonics"

Chapter 7
Paul D. Lowman,
NASA
"Plate Tectonics and Continental Drift: A Skeptic's View"

Chapter 8
Victoria Kaharl,
Woods Hole Oceanographic Institution
"Where the Sun Never Rises, Seeing Is Believing"

Chapter 9
W. R. Osterkamp,
U.S. Geological Survey
"River Aesthetics: A Janus Perspective"

Chapter 10
Lonnie Thompson,
The Ohio State University
"Ice Core Archives: The Keys to Our Future Are Frozen in Our Past"

Chapter 11
A. W. Ward,
U.S. Geological Survey
"Erosion on Mars"

Chapter 12
James C. G. Walker,
University of Michigan
"Atmospheres on Other Planets"

Chapter 13
Howard B. Bluestein,
University of Oklahoma
"Tracking Tornadoes Through the Southern Plains of the United States"

Chapter 14
Lisa Graumlich,
University of Arizona
"Is Our Climate Changing? Tree-ring Records Can Put Extreme Weather into Context"

Chapter 15
Compton Tucker,
NASA, Goddard Space Flight Center
"Mapping the World's Vegetation from Space"

Chapter 16
Andrew Knoll,
Harvard University
"The Maturation of Earth and Life"

Chapter 17
Pinar Yilmaz,
Exxon Production Research Council
"The Pursuit of Undiscovered Fuel"

Chapter 18
Thomas J. Crowley,
Applied Research Corporation, College Station, Texas
"Assessing Projections for Future Climate Change"

The Blue Planet

Introduction

Human activities influence the atmosphere. Noxious gases belching from a carbon-black plant in Romania.

"The Year Without a Summer"

The winter months of 1783-1784 were bitterly cold in Europe. The previous summer had also been unusually cool because much of the time the sky had been partially obscured by a persistent hazy blue fog. The ever-thoughtful Benjamin Franklin wondered if there might be a connection between the two. Franklin was in Paris at the time, serving as ambassador from the newly formed U.S. republic to the court of Louis XVI, king of France. In a letter written in 1784 to the Literary and Philosophical Society of Manchester, England, Franklin suggested that both the summer haze and the subsequent cold winter might have resulted from an eight-month-long volcanic eruption in Iceland:

> *During the summer months of the year 1783, when the effect of the sun's rays to heat the earth in these northern regions should have been greatest, there existed a constant fog over all of Europe, and a great part of North America. The fog was of a permanent nature; it was dry, and the rays of the sun seemed to have little effect toward dissipating it, as they easily do a moist fog, arising from water. They were indeed rendered so faint in passing through it, that when collected in the focus of a burning glass, they would scarce kindle brown paper. Of course, their summer effect in heating the earth was exceedingly diminished.*
>
> *Hence the surface was early frozen.*
>
> *Hence the first snows remained on it unmelted, and received continual additions.*
>
> *Hence the air was more chilled, and the winds more severely cold.*
>
> *Hence perhaps the winter of 1783-84 was more severe than any that had happened for many years.*

Franklin then suggested two possible causes for the fog. His first suggestion, a meteor or a comet passing through the upper atmosphere, has not stood the tests of time and investigation. His second, however, is a brilliant example of a scientific hypothesis. The haze was caused, he proposed, by "the vast quantity of smoke, long continuing to issue during the summer from Hecla in Iceland, and that other volcano which arose out of the sea near that island, which smoke might be spread by various winds over the northern part of the world." The 1783 summer eruption of Hecla, now referred to as the Laki eruption, produced the largest lava flow in historical times.

Franklin's deduction has been proved correct in an interesting way. The Laki eruption emitted a great quantity of sulfurous gases, and it was condensed droplets of sulfur compounds that caused the blue haze. These droplets were acidic, and wherever rain or snow fell, traces of the acid were brought down to the ground. Snow that fell in Greenland during the winter of 1783-1784 left a distinctly acidic layer in the Greenland ice sheet. Scientists recently drilled cores and identified this layer at several places in Greenland. By counting downward through the annual layers in the ice sheet, they have been able to prove that the acidic layer formed during the winter of 1783-1784, thus confirming Franklin's suggestion.

Great volcanic eruptions have apparently influenced the Earth's climate many times during the Earth's long history. The greatest change in historic times was not caused by the Laki eruption, but rather by the eruption of Tamboro, a volcano in Indonesia. Tamboro differs from Icelandic volcanoes in that it emits a great deal of volcanic dust in addition to volcanic gases. It is estimated that when Tamboro erupted in 1815 a total of 150 cubic kilometers (36 mi^3) of volcanic debris was blasted high into the atmosphere. So much sunlight was blocked by the dust and acid droplets that 1816 was a year of massive crop failures around the world and became known as "the year without a summer." Several years were to pass before the effects of the eruption were no longer apparent.

EARTH SYSTEM SCIENCE

Volcanism is a process controlled by events that happen deep inside the Earth. Franklin's hypothesis that something happening inside the solid Earth can dramatically change our climate is one of the earliest recognitions of what today we call Earth system science. When we study the Earth carefully, we discover that it can be treated as a system of many separate but interacting parts. Examples of the parts are the ocean, atmosphere, continents, lakes and rivers, soils, plants, and animals; each can be studied separately, but each is more or less dependent on the others. Further investigation reveals that there are numerous interactions between the parts. Just as the solid Earth, via the Laki and Tamboro eruptions, changed the atmosphere and thereby the climate, so can a change in any one part of the system produce changes in any or all of the other parts. **Earth** system science, then, is the science that studies the whole Earth as a system of many interacting parts and focuses on the changes within and between these parts.

The Scientific Method

Science is a method of learning and understanding. It advances by application of the **scientific method,** the basis of which is the use of evidence that can be seen and tested by anyone with resources who cares to do so. Although not always a clearcut process, the scientific method can be viewed as consisting of the folio wing steps.

1. *Observation.* Evidence that can be measured and observed.
2. *Formation of a hypothesis.* Scientists try to explain observations by developing a **hypothesis**—an unproved explanation for the way things happen. Franklin's two explanations for the cold winter of 1783-1784 were hypotheses.
3. *Testing of hypotheses and formation of a theory.* When a hypothesis has been examined and found to withstand numerous tests, scientists become more certain about it and it becomes a **theory.** That large volcanic eruptions can influence climate has been demonstrated several times since Franklin's hypothesis in 1784. The idea that volcanic eruptions play a major role in determining the Earth's climate is now a theory.
4. *Formation of a law.* Eventually, a theory or a group of theories may be formulated into a scientific **law.** A law is a statement that some aspect of nature is always observed to happen in the same way and that no deviations have ever been seen. An example of a law is the statement that heat always flows from a hotter body to a cooler one. No exceptions have ever been found.
5. *Continual reexamination.* The assumption that underlies all of science is that everything in the world around us is governed by scientific laws. Because even theories and laws are open to question when new evidence is found, hypotheses, theories, and laws are continually reexamined.

Modern science, which is based on the scientific method, started in Europe several hundred years ago. Initially, there were no specialties among scientists. By the middle of the nineteenth century, however, so many diverse topics had come under investigation that specialization appeared. Physicists investigated the physical properties of matter and phenomena such as light and magnetism, chemists studied how materials react, biologists studied living things, astronomers the stars, geologists the solid Earth, meteorologists the weather, oceanographers the ocean—on and on it went.

Study of the Earth as a system involves all these specialties. Separate investigations of the oceans, the atmosphere, and the solid Earth are no longer practical. When oceanographers go to sea in research vessels, they are accompanied by geologists who study the sea floor, biologists who study the aquatic life, meteorologists who study the way wind affects the sea surface, and chemists who measure water properties. When remote-sensing measurements are made from satellites, scientists use the data in many areas of specialization. Satellite observations, above all other ways of gathering evidence, continually remind us that each part of the Earth interacts with, and is dependent on, all other parts. Earth system science was born from the realization of that interdependence.

The Four Reservoirs of the Earth System

A convenient way to think about the Earth as a system of interdependent parts is to consider it as four vast reservoirs of material with flows of matter and energy between them. The four reservoirs are

1. The **atmosphere**, which is the mixture of gases—predominantly nitrogen, oxygen, carbon dioxide, and water vapor—that surrounds the Earth.
2. The **hydrosphere,** which is the totality of the Earth's water, including oceans, lakes, streams, underground water, and all the snow and ice, but exclusive of the water vapor in the atmosphere.
3. The **biosphere,** which is all of the Earth's organisms as well as any organic matter not yet decomposed.
4. The solid Earth, which is composed principally of **rock** (by which we mean any naturally formed, nonliving, firm coherent aggregate mass of solid matter that constitutes part of a planet) and **regolith** (the irregular blanket of loose, uncemented rock particles that covers the solid Earth).

Using this concept of four reservoirs and the interflow of materials and energy, we can represent the Earth system as shown in Figure I.1. This book is about the four reservoirs, and it is organized so that one reservoir can be discussed without losing sight of the roles played by the flow of materials and energy to and from other reservoirs. Icons based on Figure I.1, used at the beginning of this section are used throughout this book as a guide to the study of the Earth system.

Seawater provides an example of the way we can think about the Earth as a system of reservoirs and flows (Fig. I.2). Water leaves the ocean by evaporation and forms water vapor, which then mixes with the other gases of the atmosphere. Thus, water vapor moves from the hydrosphere reservoir to the atmosphere reservoir. As water vapor in the atmosphere rises, it cools and condenses to form clouds and eventually rain or snow, which falls on either the land or the sea. Thus, water flows from the atmosphere reservoir to the hydrosphere reservoir and from the atmos-

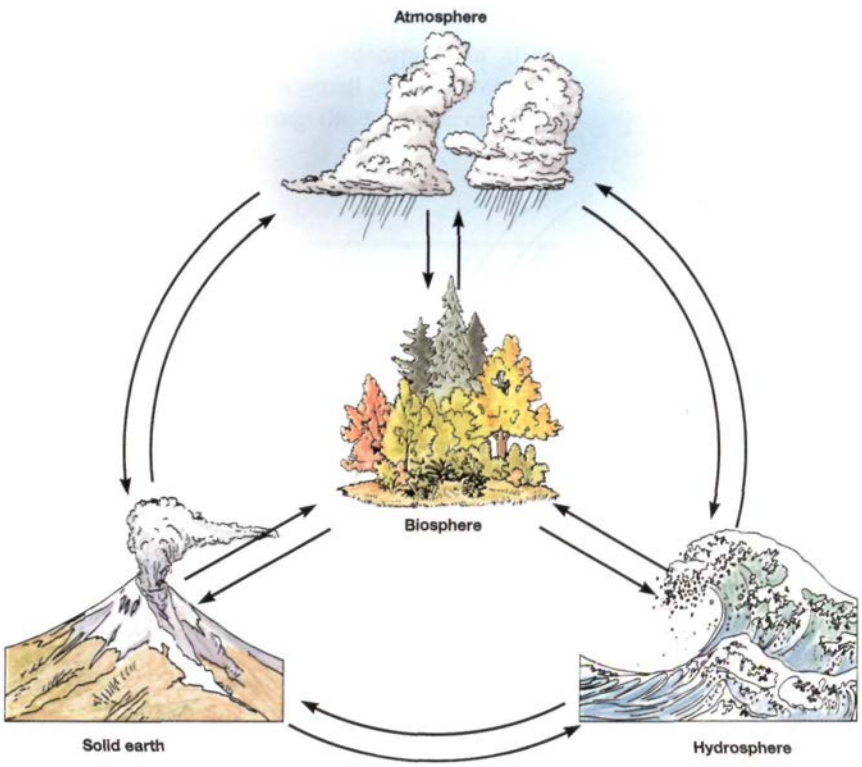


Figure I.1 Diagrammatic representation of the Earth as a system of interacting parts. Each character represents a reservoir, and each arrow a flow of energy or materials.

phere reservoir to the solid Earth reservoir. The water that falls on the land can either evaporate again, be taken up by plants in the biosphere reservoir (in both cases, water vapor is added to the atmosphere and eventually forms clouds and rain), run back to the sea, or seep into the ground. *Transpiration* is the name given to the passage of water vapor from a living body through a membrane or pore. This means water flows from land to atmosphere and from land to ocean. Snow that falls on the sea melts and mixes back into the ocean. Snow that falls on the land will also eventually melt, but most of the snow that happens to fall in Greenland, Antarctica, or high mountains may become part of an ice sheet or mountain glaciers. It could be hundreds or even thousands of years before melting occurs and the water flows back into the sea again.

The seawater flows depicted by Figure I.2 are not isolated events. For example, if rain didn't fall, trees could not grow and there would be no streams in which fish and frogs could live. Much of the biosphere therefore depends on the flow of water from the atmosphere to the land and the ocean. Consider, too, what happens when rainwater falls on the land; the water dissolves small amounts of various salts from the regolith and carries them, via streams and rivers, to the sea. (It is these salts that maintain the saltiness of seawater.) In this way, material in the regolith moves from the solid earth reservoir to the hydrosphere reservoir. The movement represented by the arrows in Figures I.1 and I.2 may be fast or slow, and so an essential part of Earth system science is the measurement of rates of movement. Flows between the reservoirs, and even between parts of the same reservoir, never cease, but the rates of flow may change, and when this happens, volumes must change too. One of the keys to understanding the Earth is therefore an appreciation of why and how reservoir volumes change.

We can observe that rivers flow continuously to the sea, that rain falls with some regularity, and that clouds are always forming in the atmosphere, which means that evaporation and transpiration never stop. If the rates of any of the flows in Figure I.2 changed markedly for a long period, the reservoirs would change in volume. In fact, world sea level is essentially constant on a time scale of several decades. Therefore, we conclude that the volume of the ocean reservoir is nearly constant and that the different flows must be very nearly in balance. But a short-term balance does not mean that changes never happen; changes do indeed occur. During glacial ages, for example, glaciers around the world grow larger. Because water to make the ice comes from the ocean, the ocean volume shrinks, leading to a fall in sea level. At the end of an ice age, the opposite happens. Ice in the glaciers melts quickly, the melt water flows back to the ocean, and sea level rises.

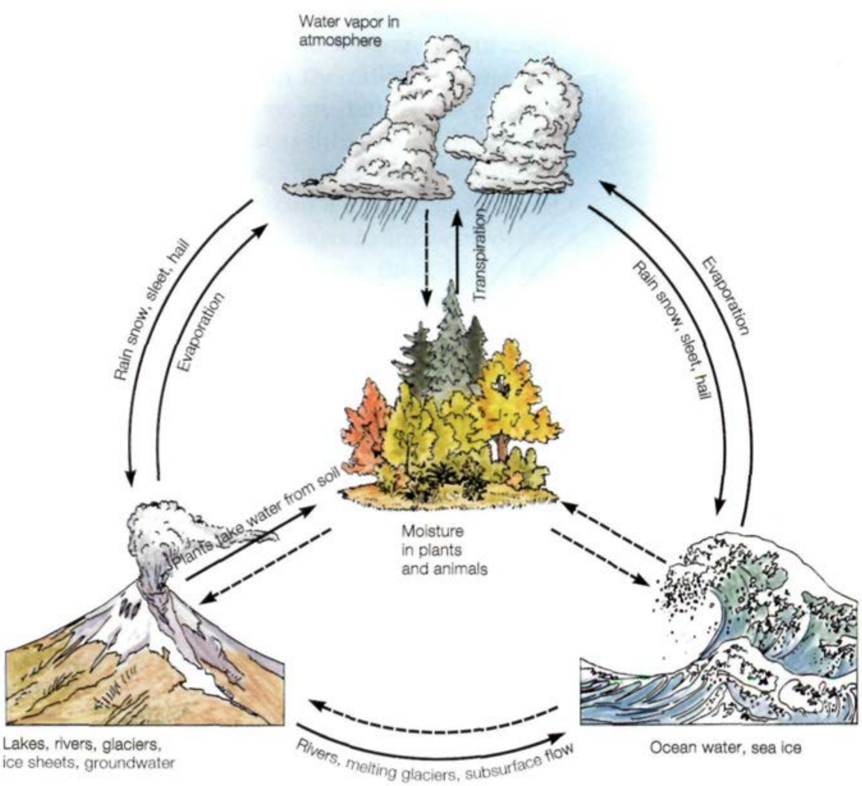


Figure I.2 The flows influencing ocean volume. Water flowing to and from the atmosphere and the solid Earth keeps the ocean volume approximately constant on a time scale of a few decades. Major flows are shown as solid arrows, minor ones as dashed arrows.

Different time scales are involved in the examples just described. Evaporation of water from the ocean, the formation of clouds, and the falling of rain or snow all take place in a few days or weeks. In contrast, the buildup or meltdown of glaciers is a much slower process that may require hundreds or thousands of years. An approximate balance may therefore be maintained on a short time scale, even though changes are slowly occurring on a long time scale. A scientific investigation of the Earth, then, is concerned with both fast and slow rates (that is, with events that happen on both short and long time scales). Rates in the atmosphere and biosphere tend to be rapid and to occur on short time scales. Rates in the solid Earth tend to be slow and to operate on time scales of thousands or millions of years, and rates in the hydrosphere vary from rapid (as in flowing streams) to slow (as in the flow of water deep beneath the ground surface) (see Table I.1).

UNIFORMITARIANISM

Among the many important questions asked by scientists is the question of the relative importance of cumulative small, slow changes like the washing away of soil by an ordinary rainstorm, in contrast to massive, drastic changes like earthquakes and floods. Massive changes are relatively infrequent, but they cause rapid, dramatic changes to the landscape. People remember the floods, hurricanes, landslides, and other great events that change the landscape, but they quickly forget the innumerable small rain showers between the great events. During the seventeenth and eighteenth centuries, before the power of the scientific method became widely appreciated, people suggested that all the Earth's features—mountains, valleys, and oceans—had been produced by a few great catastrophic events. The catastrophes were thought to be so huge they could not be explained by ordinary processes, and so the supernatural was called upon. This concept came to be known as **catastrophism.** Not only were the catastrophes thought to be gigantic and sudden, but also some people believed they had occurred relatively recently and fit a chronology of catastrophic events recorded in the Bible.

The Rise of a New Theory

During the late eighteenth century, the concept of catastrophism was reexamined, compared with geological evidence, and found wanting. The person who used the scientific method to assemble the evidence and propose a counter theory was James Hutton (1726-1797), a Scottish physician and 'gentleman farmer'. Hutton was intrigued by what he saw in the environment around him, especially in Edinburgh, where he lived and studied. He wrote about his observations, offered hypotheses, and then used tests and observational evidence to develop theories that were supported by the evidence. Hutton is widely regarded today as the father of the scientific specialty we now call geology. In 1795 he published a two-volume work titled *Theory of the Earth, with Proofs and Illustrations* in which he introduced his counter theory to catastrophism.

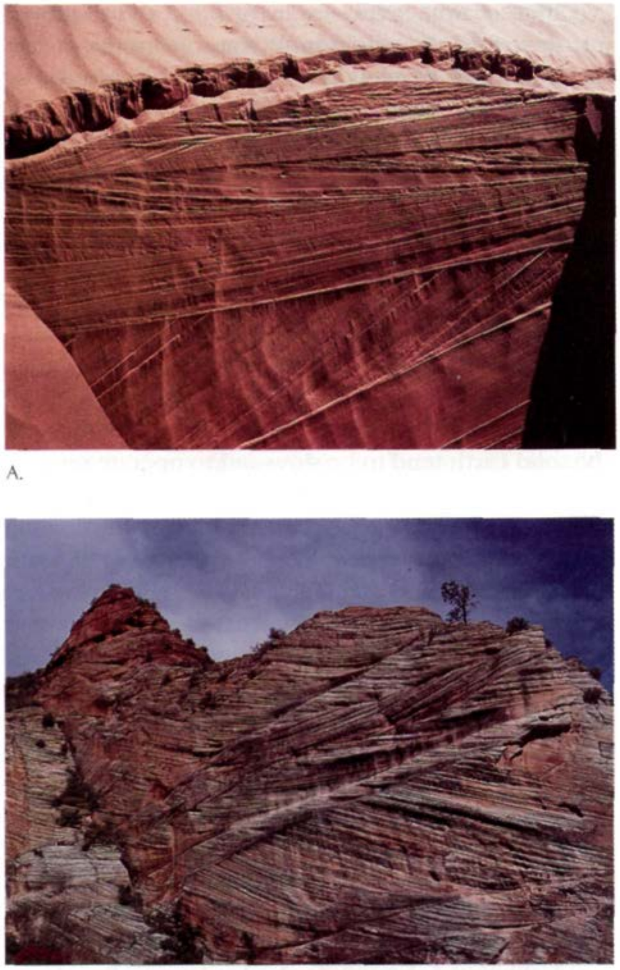
We refer to the complex group of related processes by which rock is broken down and the

Table I.1
Examples of Fast and Slow Rates

	Event	**Time**
ATMOSPHERE	Formation of clouds (fast)	Minutes
	Tornado (fast)	Hours
	Hurricanes (fast)	Days
	Duration of an ice age (slow)	Thousands of years
HYDROSPHERE	Flash flood in a desert stream (fast)	Minutes to hours
	Flood in a great river system (fast)	Days to months
	Circulation of deep ocean water (slow)	Years
BIOSPHERE	Lifetime of a grass (fast)	Months
	Lifetime of a redwood (slow)	Hundreds of years
	Growth of a forest (slow)	Hundreds to thousands of years
SOLID EARTH	Landslide (fast)	Minutes
	Volcanic eruption (fast)	Hours
	Elevation of a mountain range (slow)	Tens of millions of years

products moved around as **erosion.** Hutton observed the slow but steady effects of erosion: rock particles are carried great distances by running water and ultimately deposited in the sea. He reasoned that mountains must slowly but surely be eroded away, that new rocks must form from the debris of erosion, and that the new rocks in turn must be slowly thrust up to form new mountains. Hutton couldn't explain what causes mountains to be thrust up, but everything, he argued, moves slowly along in repetitive, continuous cycles. His ideas evolved into what we now call the **Principle of Uniformitarianism,** which states that the same Earth processes we recognize in action today have been operating throughout the Earth's history. We can examine any rock, however old, and compare its characteristics with those of similar rocks forming today in a particular environment. We can then infer that the old rock very likely formed in the same sort of environment. In short, *the present is the key to understanding the past.* For example, in many deserts today we can see gigantic sand dunes formed from sand grains transported by the wind. Because of the way they form, the dunes have a distinctive internal structure (Fig. I.3A). Using the Principle of Uniformitarianism, we infer that any rock composed of cemented grains of sand and having the same distinctive internal structure as modern dunes (Fig. I.3B), is the remains of an ancient dune.

Hutton was especially impressed by evidence he saw at Siccar Point in Scotland (Fig. I.4). Sandstones are formed by the cementation of sands into solid rocks. Wherever Hutton observed sand being deposited, it formed horizontal layers: he realized that all sandstones must have originally been laid down as horizontal layers. At Siccar Point, however, Hutton could see ancient sandstone layers standing vertical and capped by gently sloping layers of younger sandstone. The boundary between the layers, he pointed out, was an ancient surface of erosion. The now-vertical layers are composed of debris that was eroded, millions of years ago, from an ancient mountain range, transported by streams, deposited on the sea floor, and there formed into new rocks. As a result of mechanisms we will study in Chapter 7, the newly formed rock layers were uplifted, tilted to their present position, and eroded. When erosion had formed a flat surface on the tops of the vertical sandstone layers, a pile of younger erosional debris was deposited on the new surface. Eventually, the younger debris became rock and uplift occurred again, although not much tilting was in evidence during this second stage of uplift. The cycle of uplift, erosion, transport and deposition, solidification into rock, and renewed uplift that could be deduced from this visible evidence impressed Hutton immensely. There is, he wrote, "no vestige of a beginning, no prospect of an end" (1795, *Theory of the Earth)* to the Earth's geological cycles.



A.

B.

Figure I.3 The internal structure of sand dunes, ancient and modern, demonstrates the power of uniformitarianism. A. A distinctive pattern of wind-deposited sand grains can be seen in a hole dug in this dune near Yuma, Arizona. B. The same distinctive pattern in rocks in Zion National Park, Utah, lets us infer that these rocks, too, were once sand dunes.

Geologists who followed Hutton have been able to explain the Earth's features in a logical manner by using the Principle of Uniformitarianism. In so doing, they have also made an outstanding discovery—the Earth is incredibly old. It is clear that most erosional processes are exceedingly slow. An enormously long time is needed to erode a mountain range down, for instance, or for huge quantities of sand and mud to be transported by streams, deposited in the ocean, then cemented into new rocks and the new rocks deformed and uplifted to form a new mountain. Slow though it is, this cycle has been repeated many times during the Earth's long history.

Figure I.4 Siccar Point, Berwickshire, Scotland. The vertical layers of sedimentary rock on the right, originally horizontal, were lifted up into their vertical position. Erosion developed a new land surface that became the surface on which the now gently sloping layers of younger sediments were laid. The gently sloping layers, which are named the Old Red Sandstone, are 370 million years old. At this locality, in 1788, James Hutton first demonstrated that the cycle of deposition, uplift, and erosion is repeated again and again.

Although Hutton never used the term *Earth system,* he described parts of the Earth system in ways that show he understood the concept. His concept of a cycle of erosion, transport, deposition, formation of new rock, and uplift is just another way of discussing flows of materials between reservoirs.

The concept of uniformitarianism is very important to all branches of science, not just geology. For example, astronomers have developed a powerful theory about the way stars form, pass through a long life cycle, and then die. Because the lifetime of a star is measured in billions of years, it is not possible to make all needed observations by watching a single star. Instead, astronomers study the billions of stars in the sky, observe examples at various stages of development, and find that the cycle of birth, growth, and death follows a predictable pattern. Whenever a new star is examined, uniformitarianism allows the observer to use previous observations to estimate where in its life cycle the new star is.

Rare Events and the Reconsideration of Catastrophism

Uniformitarianism is a powerful principle, but should we abandon catastrophism as a totally incorrect hypothesis? Recent discoveries of thin but very unusual rock layers at many places around the world suggest that rare, random, catastrophic events have indeed caused massive changes in the geological record. However, these are not the catastrophes perceived by seventeenth-century biblical scholars, who had to call on supernatural forces to explain things. Rather, they are events that can be readily explained but are so large and damaging that they caused catastrophic change.

One rare event suggested by the unusual rock layers is a huge meteorite striking the Earth (Fig. I.5). The peculiar rock layers mentioned above are rich in the uncommon metal iridium, which is much more abun-

Figure I.5 Meteor Crater, near Flagstaff, Arizona. The crater was created by the impact of a meteorite about 50,000 years ago. It is 1.2 km in diameter and 200 m deep. Note the raised rim and the blanket of broken rock debris thrown out of the crater. Many impacts larger than the Meteor Crater event are believed to have occurred during the Earth's long history.

dant in meteorites than in the Earth's common rocks. The iridium-rich rocks have been discovered in Italy, Denmark, and other places around the world (Fig. I.6). The hypothesis is that a massive impact did occur about 66 million years ago, perhaps in the Yucatan area of present-day Mexico, and that, as a result, many forms of life, including the dinosaurs, became extinct. The suggestion is that the impact threw so much debris into the atmosphere that the air temperature plummeted, just as during the mighty Tamboro eruption in 1815. In the impact case, however, the temperature drop was much steeper and much faster. Consequently, most animals and many plants could not survive. When the debris settled, it formed a thin, iridium-rich layer derived from the iridium-rich meteorite, wherever sediments were being deposited around the world. This hypothesis is still being tested, and many confusing bits of evidence remain to be explained.

Even more dramatic extinctions than the one *66* million years ago have occurred at other times in the past. The geologic record indicates that one about 245 million years ago sent almost 90 percent of all living plants and animals to extinction. No evidence suggests that a meteorite impact caused this great extinction. To the contrary, fragmentary evidence indicates that slow but drastic climate changes resulting from the break up of a huge supercontinent may have caused it. When we view the Earth's history as a combination of endless small changes as well as a series of such repeated but rare events, we have to conclude that uniformitarianism can describe even the rare events and that there is absolutely no reason to believe that similar events will not occur again. Indeed, there are good reasons to believe just the opposite. Astronomers have already identified a comet that will come close enough to the Earth at some time during the next 1500 years and possibly cause an event as big as the one 66 million years ago.

A fascinating but frightening suggestion has been made that a disaster of a different kind may already be happening. Our collective human activities may be changing the Earth so rapidly and so significantly that we may be living through a change similar in magnitude to some of the major ones in the geological record. At present, the suggestion is only a hypothe-

Figure I.6 This thin, dark layer of rock (marked by the coin) is rich in the rare chemical element iridium and looks out of place in the thick sequence of pale-colored limestones above and below. The iridium-rich layer, here seen in the Contessa Valley, Italy, has been identified at many places around the world and is believed to have formed as a result of a world-circling dust cloud formed by a great meteorite impact about 66 million years ago.

sis; it remains to be tested and thereby proved or disproved. Nevertheless, the very fact that serious scientists are concerned that the hypothesis might prove to be true emphasizes an important fact: human activities are an important part of Earth system science, and changes to the Earth and the welfare of the human race are indissolubly linked.

THE HUMAN DIMENSION

Notice in Figures I.1 and I.2 that the biosphere lies at the center of the Earth system diagram. It is placed there for a special reason. Significant changes are now taking place in many of the flows between the biosphere and the other reservoirs, and as a result the reservoirs are changing in many unexpected ways. Some of the changes have become daily news—the ozone hole, the increase of carbon dioxide in the atmosphere, the dispersal of pesticides throughout the ocean, the rate at which we are consuming nonrenewable resources such as oil, and the extinction of plant and animal species, to name several examples.

We, the human population, are the cause of these and other recent changes. Until a few centuries ago, the human population was small (Fig. I.7), and even though humans have always changed their local environments, a small population causes changes so slowly that the Earth system is not thrown out of balance. Now the population is large and growing ever larger. At the time these words are being written, in 1993, the world's population is 5.5 billion and increasing by 95 million each year. There are now so many of us that we are changing the Earth just by being alive and going about our business.

Many kinds of large animals have, at various times, lived on the Earth. Throughout all of the Earth's long history, however, there has never been such a huge number of large animals as in the human population today. Our collective activities have become so pervasive that there is no place on the Earth we haven't

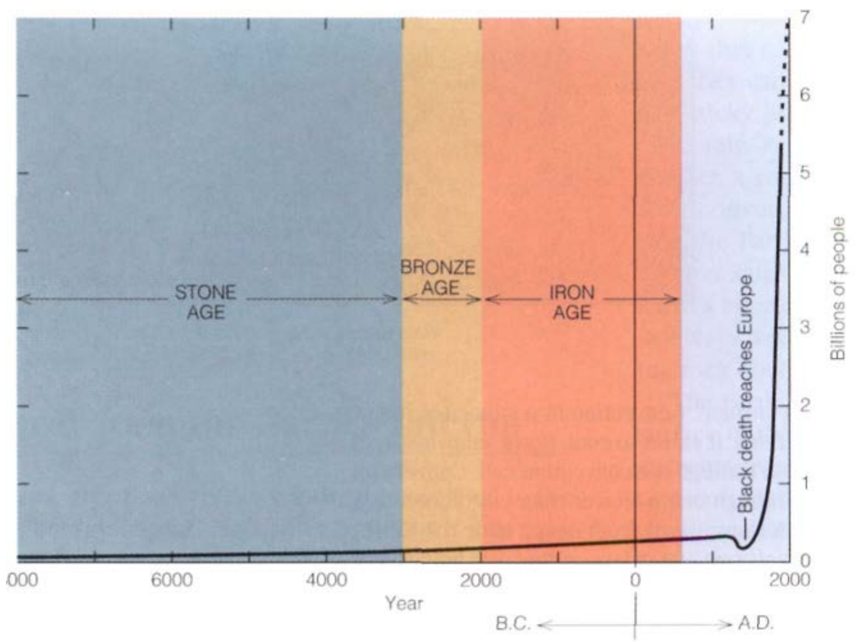


Figure I.7 Human population growth from ancient times to the present. Growth was slow up to the time of the Industrial Age (about 1750), except for setbacks caused by such disasters as the Black Death, which spread from Asia and reached Europe in 1348.

changed. We go almost everywhere to seek the resources we need. In the process, we have made rainfall more and more acidic, we have caused fertile top soil to blow away, and we have changed the composition of the soil that remains. We have caused deserts to expand, and we have changed the composition of the atmosphere, the ocean, streams, and lakes. Even the snowflakes that fall on Antarctica bear the imprint of our activities. In short, we are influencing all of the reservoirs and many of the flows, and thereby changing our own environment. And we continue to do so at ever faster rates. We have even coined a special term to describe the changes produced in the Earth system as a result of human activities: **global change**. Measuring, monitoring, and understanding global change is now a topic of intense study by many scientists. Once again, uniformitarianism is their guide: the present is not only the key to understanding the past, but it is also the key to understanding the future.

Global change should not be viewed as necessarily negative. Most human activities have made the world a nicer and friendlier place in which to live. No one could deny that building cities and clearing land for farms causes large changes in the environment, but who would argue that a beautiful city like Paris is not a proud achievement? Think, too, of the abundant food that flows from modern agriculture. Our ancestors had a much harder time feeding themselves than we do today. To be sure, we have witnessed some changes to the environment that may be dangerous; most of these changes happened accidentally because we didn't understand the Earth system sufficiently well. When we started burning coal 300 years ago, for instance, no one had the slightest idea that someday (i.e., today) the atmosphere would be changed as a result. If, say, the climate becomes warmer because of these changes to the atmosphere, ice in Antarctica might melt, the sea level might rise, and cities might be flooded. Surely those are important consequences that we must consider, but note that we say *might* happen—might, because we do not yet understand the Earth system in enough detail to be sure. A necessary part of Earth system science is therefore an investigation of how the collective actions of the human population are changing the reservoirs and flows, and what the consequences of those changes will be. We address the issues of human influences on the Earth system at many places throughout this book.

ENERGY AND THE EARTH SYSTEM

Everything that happens in and on the Earth requires energy. The flows of material between the reservoirs of the Earth system all involve energy, particularly heat energy. Heat can be transferred in three ways:

1. **Conduction** is the process by which heat can move through solid rock, or any other solid body, without changing the shape of the solid. Conduction is the way heat moves along the metal handle of a hot saucepan, but it does not cause the move-

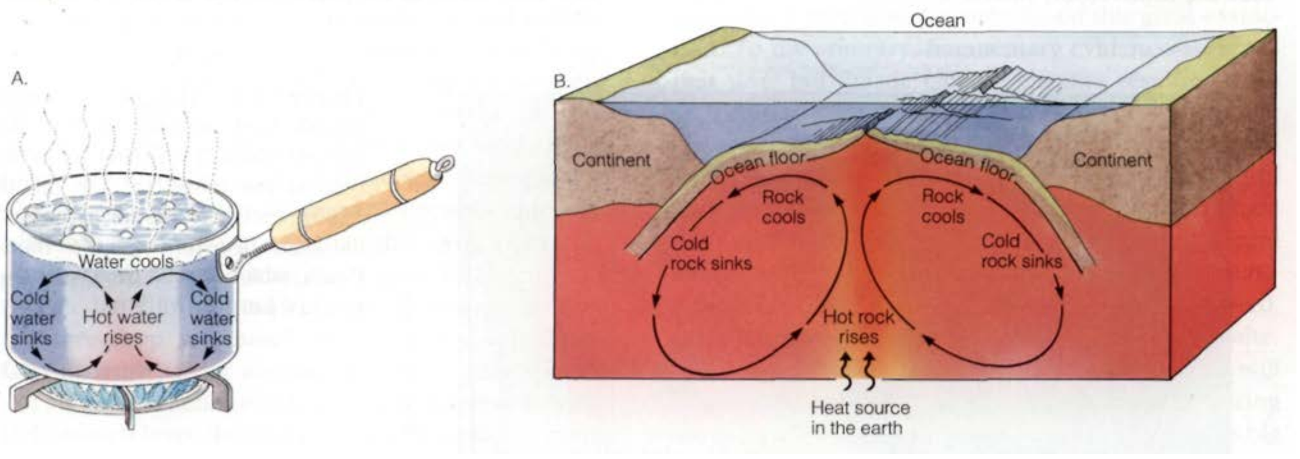


Figure I.8 Convection shapes the Earth's surface. Convection in a saucepan full of water: heated water expands and rises. As it rises, it starts to cool, flows sideways, and sinks, eventually to be reheated and pass again through the convection cell. Convection as it is hypothesized to occur in the Earth. Though much slower than convection in a saucepan, the principle is the same. Hot rock rises slowly from deep inside the Earth, cools, flows sideways, and sinks. The rising hot rock and sideways flow are believed to be the factors that control the positions of ocean basin and continents.

ment of hot material from one place to another.

2. **Convection** is the process by which hot, less dense materials rise upward and are replaced by cold, downward-flowing and sideways-flowing materials to create a **convection current** (Fig. I.8).
3. **Radiation** is the process by which heat passes through a gas, a liquid, or even a vacuum.

Heat energy must come from somewhere. On the Earth, there are three main sources of energy: the Sun, the interior of the Earth, and the tides. We'll deal with the first two sources here. For more details on these sources and the tides, see "A Closer Look."

Energy from the Sun

Energy from the Sun reaches the Earth as radiation. Approximately 70 percent of the Sun's radiation that reaches the Earth is absorbed by the land, the sea, or the atmosphere. The remaining 30 percent is simply reflected back into space.

The radiation absorbed by the sea warms the water and causes evaporation. The resulting water vapor forms clouds and eventually rain, snow, sleet, or hail. The radiation absorbed by the land warms the exposed rocks and regolith. Warmed air expands, becomes less dense, and rises convectively. Then cool air flows in to take the place of the rising air. Flowing air is wind, and as winds blow over the sea, they create waves. Thus, the familiar everyday processes that happen at the Earth's surface—rain, streams, winds, waves, even glaciers—are produced by the Sun's energy. All of the Earth's *external processes,* as the processes of weather, climate, and erosion are called, are driven by energy from the Sun. When we discuss the energy from the Sun in the earth system, we will use the sun icon shown above.

Energy from the Earth's Interior

Volcanic eruptions, unlike winds, are unrelated to the Sun's energy output. No matter how hot it gets on a summer's day, the Sun's heat is insufficient to melt rocks, and even frigid Antarctica has active volcanoes. We therefore hypothesize that the heat energy needed to form the molten lava that spews from a volcano must come from somewhere inside the Earth.

This hypothesis is not difficult to test. It you went down into a mine and measured rock temperatures, you would find that the deeper you went, the higher the temperature would become. The increase in temperature as you go deeper is called the **geothermal gradient**. We use this gradient to make a deduction based on the scientific law that heat always flows from a warmer place to a cooler one. We deduce that heat energy must be flowing outward from the hot interior of the Earth toward the cool surface. Careful measurements made in mines and drill holes around the world show that the geothermal gradient varies from place to place, ranging from 15° to 75^{c}C/km, (95°F/mi. to 269°F/mi.) but becomes less pronounced with depth so that far inside the Earth the gradient is only 1 or 2°C/km (55°F/mi. to 58°F/mi.). By extrapolation, we calculate that the temperature at the center of the Earth must be at least 5000°C (9032°F). Measurements also establish that the heat flow is greatest in those places where there is volcanic activity. We can conclude, therefore, that volcanism is indeed caused by the Earth's internal heat energy.

The heat energy that flows out through solid rocks shown at the beginning of this section. From the Earth's interior does so by conduction. But because volcanoes obviously involve the movement of hot material, we have to conclude that at least some heat energy moves inside the Earth by convection.

The hypothesis that flow occurs in a seemingly rigid solid body like the Earth may seem odd, but flow in solids can be observed in any glacier. Glaciers flow slowly downvalley partly because solid ice at the bottom is deformed by the weight of ice above. Tests show that rocks, like glaciers, don't have to melt before they can flow. Rocks, if sufficiently hot, can flow like sticky liquids, although the rates of flow are exceedingly slow. The higher the temperature, the weaker a rock is and the more readily it will flow. Slow convection currents of rock are possible deep inside the Earth because the interior is very hot. Convection currents bring hot rocks upward from the Earth's interior. The hot rock flows slowly up, spreads sideways, and eventually sinks downward as the moving rock cools and becomes more dense.

The Earth's internal convection currents shape the surface of the Earth. We see the effects of convection currents everywhere around us. Because of convection currents, mountains are thrust upward, and continents move slowly and are sometimes split asunder, forming new ocean basins. The most important effect of convection currents inside the Earth is **plate tec-**

A Closer Look

Energy

Energy is the capacity to do work or, to say it more familiarly, the capacity to make things happen. The surface of our planet is a place of intense activity solely because of the large amount of energy it receives. Without energy, the Earth would be a lifeless planet.

Energy is measured in terms of the work performed. As a result of scientific and engineering specialization over the past two centuries, many different energy units have been formulated, some familiar, some unfamiliar. In this book, all units used, including those of energy, are SI (Système International d'Unités) units. For a fuller discussion of SI units, please see Appendix A.

The SI unit of energy is the joule, a unit named for James P. Joule (1818-1889), an English physicist who made discoveries about heat and electricity. A **joule (J)** is defined as the work done when a force of one newton acts over a distance of one meter. The joule may be an unfamiliar unit to you. The most familiar energy unit to most people is the **calorie (C),** which is the amount of heat energy needed to raise the temperature of one gram of water by one degree Celsius. A calorie is equal to 4.186 J. Be careful not to confuse a so-called big Calorie (note the capital C), the kind counted by weight watchers, with the "little" calorie of science; a Calorie is equal to 1000 calories.

Often it is more informative to consider the *rate* at which energy is available than to count the *total* amount of energy. For example, the energy that reaches the Earth from the Sun does so at a fixed rate. What is important in understanding the effect of the Sun's energy on the Earth is this rate of arrival, not the total amount of energy that has ever reached the Earth. The most familiar unit of energy rate (or, as we say more commonly, the work rate) is the horsepower. This ancient unit was defined in 1 766 by James Watt (1736-1819), inventor of the steam engine that ushered in the Industrial Revolution. Watt needed a way to evaluate the rate at which his steam engine did work. In his day, most of the heavy work was done by horses, and so he compared the working rate of his engine with the rate at which a horse could perform the same tasks. The definition of a horsepower is the work done when a weight of one pound is raised at a rate of one foot per second.

Because horsepower is a cumbersome unit, the energy rate unit used in this book is the watt, named for the same Watt who coined the term *horsepower.* A **watt (W)** is defined as 1 joule/second. The relationship between watts and horsepower is 1 horsepower equals 746 watts.

The total energy used by humans in 1992 (counting all sources, such as oil, gas, wood, hydroelectricity, nuclear, wind, and solar) was estimated to be 2.8×10^{20} J, which equates to a usage rate of 9×10^{12} watts.[1] To put such a huge number into perspective, consider that a healthy, active human can work at the rate of about 100 watts but only for 8 hours a day. Therefore, considering our ability to work in terms of a 24-hour day, we humans are at best only 33-watt machines!

Figure CI.1 shows that energy reaches the Earth from three sources: the Sun, the interior of the Earth, and the tides. Tidal energy is the least important of the three. Gravitational attraction by the Moon and the Sun creates two tidal bulges in the ocean. No significant amount of energy would be involved were it not for the Earth's rotation about its axis. As a result of this rotation, the positions of the tides in the ocean change continuously. Because the Earth makes a complete rotation every 24 hours and because there are two tides, each place in the ocean has two high and two low tides a day. Tides cannot move around the Earth unhindered because the continents get in the way. In effect, the continents, in their daily rotation on the Earth's axis, run into a mass of water piled up by the tide twice a day. Every time a collision between water and continent takes place, the Earth's rotation is slowed a tiny amount. Fortunately, the rate of slowing is small (the length of the day is increasing by 0.002 second a century as a result), but the net result is that a small portion of the Earth's rotational energy is transferred to the Earth's surface. The rate of transfer is only 2.7×10^{12} watts, but eventually, some billions of years from now, the tides will bring the Earth's rotation to a stop.

Energy from the Earth's interior reaches the surface by conduction and convection. Conduction is the more important process of the two, but also the more diffuse. The average loss of heat energy by conduction through rocks is 4.2×10^{-6} watts/cm^2. Since the surface area of the Earth is 510×10^{16} cm^2, the rate of conducted heat flow is 21×10^{12} watts. Convective heat flow at the Earth's

tonics—the slow, lateral movement of segments (plates) of the Earth's hard, outermost shell. The movement of plates splits and moves continents (Fig. I.9), forms mountains, triggers earthquakes, and causes volcanoes to be where they are. Convection currents, via plate tectonics, continuously shape and change the face of the Earth. The processes driven by energy flowing out from the solid Earth are called *internal processes.*

To understand what a remarkable and dynamic planet the Earth really is, look at the photograph of the Moon in Figure I.10. The Moon is the Earth's near-surface is mainly a result of volcanism and hot springs, a great deal of which takes place under the sea. The total is 11.3×10^{12} watts.

The Sun makes the other energy sources pale by comparison. Approximately a third of the 17.3×10^{16} watts of incoming solar energy is reflected back into space by clouds and the surface of the land and sea. The remainder follows various paths through the atmosphere and hydrosphere. Evaporation of water to form clouds uses energy at a rate of 4.0×10^{16} watts, which is huge compared with the rate at which energy is used by green plants in photosynthesis—0.004×10^{16} watts. Nevertheless, because a small amount of organic matter is always being buried in mud and other sediment and therefore not decaying, a significant amount of ancient solar energy is buried in the Earth. If all of this buried matter were dug up and burned, an estimated 10^{26} J of energy would be released. This seems like a huge amount, but it equates to only 20 years of solar energy. Most of the organic matter buried in the Earth is so dispersed that it takes more energy to dig it up than can be gained when it is burned. The estimated amount of *economically* recoverable energy from buried organic matter is only equivalent to 18 days' worth of solar energy!

[1] There are 3.2×10^7 s in one year, so the arithmetic is

$$2.8 \times 10^{20}\ \text{J/y} = 2.8 \times 10^{20}\ \text{J} / 3.2 \times 10^7\ \text{s} = 9 \times 10^{12}\ \text{J/s} = 9 \times 10^{12}\ \text{W}.$$

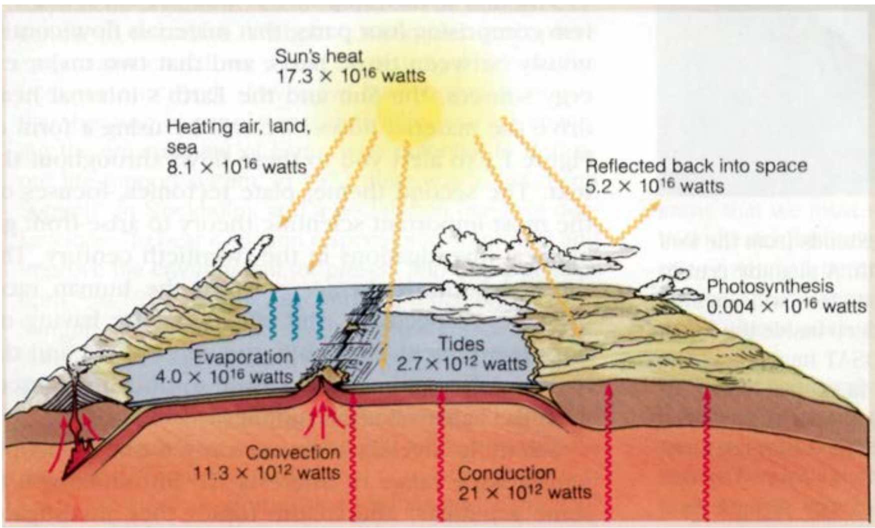


Figure CI. 1 Energy reaches the Earth's surface from three sources: the Sun, the Earth's interior, and the tides. The Earth is such a dynamic planet because these three energy sources drive different activities. Internal heat drives all of the solid Earth's internal activities, such as mountain building and volcanism. The Sun's energy and, to a much smaller extent, the tides drive all of the external activities, such as erosion, wind, ocean currents, and the growth of green plants.

est neighbor in the solar system, but it has an ancient, seemingly changeless surface. The surface does not change because the supplies of internal heat energy are so run down that no new mountain ranges have been formed for more than 2 billion years. Mars, the planet nearest to the Earth, does reveal evidence of volcanism at some time during the past billion years, but otherwise it seems to be a dead planet. It seems likely that plate tectonics and most of the wonderful landscape-forming processes that make the Earth such a dynamic and interesting planet are not active on our closest neighbors in space.

Figure I.9 The African Rift Valley extends from the Red Sea in the north to Malawi in the south. A gigantic rent in the Earth's surface, hundreds of kilometers wide, marks the place where convection currents deep inside the Earth are splitting Africa in two. This LANDSAT image is of the eastern side of the Rift Valley (green) in central Kenya. To the east (right) a high plateau (red) marks the eastern edge of the Rift Valley. The dark lines in the valley are elongate fractures in the Earth's outermost layer. The five round features in the valley are volcanoes formed as a result of lava rising up the fractures.

Figure I.10 Photograph of the Moon's surface. The circular features are meteorite impact craters. Note that there is no indication of continents or ocean basins.

ABOUT THIS BOOK

Although this text takes all of Earth system science as its scope, it emphasizes four themes:

1. The interdependence of the Earth's four major reservoirs—the solid Earth, the atmosphere, the hydrosphere, and the biosphere.
2. The connective link between internal convection and the Earth's external features through plate tectonics.
3. The fact that the human race is causing measurable changes in some of the Earth's reservoirs and is influencing the flows of material and energy between them.
4. The need for humans to use the Earth's limited store of natural resources wisely and to understand how human activities change the environment.

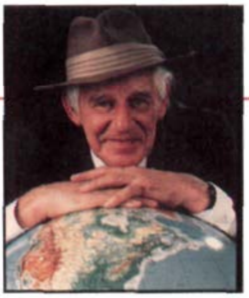
The first theme emphasizes that the Earth is a system comprising four parts, that materials flow continuously between those parts, and that two major energy sources, the Sun and the Earth's internal heat, drive the material flows. We will be using a form of Figure I.1 to alert you to these flows throughout the text. The second theme, plate tectonics, focuses on the most important scientific theory to arise from geological investigations in the twentieth century. The third and fourth themes concern the human race; they are, respectively, the effects we are having on our environment, especially the atmosphere and the hydrosphere, and, society's need to obtain resources from the Earth's limited supplies.

Scientific investigations are carried out by people with a wide range of interests. To introduce you to some scientists, and to the topics they investigate, brief "Guest Essays" written by working scientists appear in each chapter. As you read the essays, you will observe that some of the writers offer hypotheses that differ from ours. That is how science progresses—by questioning and by allowing everyone to draw conclusions and develop hypotheses.

Each chapter contains features titled "A Closer Look," essays that provide more quantitative and numerical data about a topic than are mentioned in the text. Three additional features, found at the end of each chapter, are designed to help the reader assimilate the material that has been covered in the chapter: a brief summary, a list of important words and terms that should be remembered because they will be used at various places in the book, and a series of questions based on material in the chapter.

Guest Essay

Toward Global Responsibility: Earth Sense



William S. Fyfe earned his B.Sc, M.Sc, and Ph.D degrees in chemistry and geology at the University of Otago, New Zealand. Recently, he has served as the dean of the Faculty of Science of the University of Western Ontario. His teaching and research in geochemistry, global tectonics, resource development and conservation, agriculture geochemistry, environmental geochemistry, and nuclear waste disposal has earned him many honors from numerous academic, professional, and government institutions.

During this century, we have seen a remarkable increase in our ability to observe natural systems. We can look into space and watch the behavior and evolution of galaxies. From space, we can watch the surface of our planet in remarkable detail and witness the birth of a hurricane, the erosion of soil, the destruction of a forest, and dust spreading from a volcano. And, at the atomic-molecular level, we can observe the interactions of electrons and atomic nuclei in all the materials that make up **our** environment and biological species. From such power of observations, we have come to appreciate the vast interactive complexity of the Earth System, and we marvel at the systems that make the existence of organisms like us possible. We have learned that, without photosynthetic organisms, there would not be enough oxygen for us to survive. Without our complex atmosphere and magnetic field, we would be destroyed by radiation from the Sun. And, slowly, we have begun to realize that, because of our clever technologies, we are changing the environment of Earth, with potential to destroy our life support system. In 1972 there was a first "Earth Summit" in Stockholm, and a declaration (we love declarations) "to bear a solemn responsibility to protect and improve the environment for present and future generations." In 1993 we had a second and much bigger Earth Summit, in Rio de Janeiro, with more complex declarations. But did things improve between 1972 and 1993? Did actions match words?

At the period when Christ was supposed to have been born, we think there were about 200 million humans on our planet. Today, we add to that population every two years. We are now approaching 6 billion and, barring almost unthinkable catastrophes, human population will reach 10 billion during the next century. People speak of potential catastrophes, but **for** the 2 billion **or** so who today suffer from extreme malnutrition, or the 40,000 young children who die each day, the catastrophe is now.

What elements comprise our life support system? And, within this system, how many humans can live well on this planet (with hope for the long-term future), given our present technologies? We depend on climate, which never was and never will be constant, and the chemistry and physics of our atmosphere, which we are now changing very quickly. Energy from the Sun, which, along with our atmosphere, keeps the planet comfortable, provides a constantly renewable resource. At this time, however, the energy we use for our technologies, transportation, industry, and farming is 90% derived from rapidly declining, nonrenewable fossil carbon in oil, coal, and gas. We rely on the availability of usable water, and now, in at least forty nations, there are crises concerning clean water supply. The concrete, copper, steel, and phosphate fertilizers that we mine from the minerals and rocks near the Earth's surface amount to an annual consumption rate of twenty tons of rock per person. Finally, soil, the thin surface layer of Earth that supports much of the photosynthetic biosphere, combines with water and climate to provide our food. We now know that we must sustain a large food surplus, yet the Worldwatch Institute tells us that we are losing soil at a rate of nearly 1 percent per year. If this statistic is true, it indicates that we are headed for a food disaster (and in many places, such as the Sudan and Somalia, it is here now).

Today, there exist new and growing waste disposal problems. Every day, we read about the problems and costs of new landfill sites. In the developed world, human activities produce complex wastes (make a list of **your** wastes), which total 2 kg per day, per person. We worry about nuclear wastes, and we continue to search for the best disposal methods and sites. And we worry about problems related to our health and the byproducts of our technologies, which include herbicides, pesticides, and exhaust gases from our transportation.

We now know that we cannot tolerate careless technology. I am always reminded of a lecture I heard from Max Perutz, who received the Nobel Prize for his work on the structure of hemoglobin. He stressed that all life has a genetic code, based on the same molecular building blocks, and thus, any chemical, any change, that influences (destroys) one organism may well do similar things to all organisms. For example, the pesticide DDT was designed to eliminate the scourge of malaria by killing the mosquitoes that transmitted the disease. However, its effects on many other species like birds was catastrophic, and as a result, in most of the world DDT is banned.

Is there hope? The answer is certainly yes. If we can use our knowledge with wisdom—common sense—we can live well on a sustainable basis. However, we urgently need new technologies. We are not short of energy—the Sun and the deep Earth (geothermal energy) can supply our needs without vast pollution. We can stop soil erosion and forest destruction. We can stop the pollution and waste of precious water. But, in order to stop such waste, all people on this planet must be eco-geo-literate.

I am writing these notes in the summer of 1993. We have just witnessed the greatest floods in recorded history in the Mississippi region and in northern India. Why? We are not sure. But every cook knows that as a liquid gets hotter, the vapor pressure increases in the pot. We are getting warmer, and, as the oceans get a little warmer, there will be more water in the atmosphere and more rain somewhere. And every cook knows that if the pot gets too hot, it may splatter and boil over. As systems get hotter, they become more turbulent, and their behavior becomes more erratic and less predictable. We know that burning fossil carbon may accelerate all of these effects. As Wallace Broecker of Columbia University recently said, we are playing Russian Roulette with the planet—with our environment. Do we have Earth Sense? Ask yourself some questions: Will I drive a small fuel-efficient or an electric car? Will I support an efficient transportation system? Will I stop waste? If the answers are positive, life on this planet can be richer in all ways. For a moment, just think what might happen to this planet if all 10 billion behaved like North Americans. I very much liked some recent words from Sir Christopher Ball of England: "Success will come to those who are able to design strategies that recognize the realities of today and tomorrow, not of yesterday."

Summary

1. Earth system science is the study of the whole Earth viewed as a system of many interacting parts and focuses on the changes within and between the parts.
2. Science is a system of learning and understanding that advances by application of the scientific method: observation, formation of a hypothesis, testing, formation of a theory, more testing, and, in some cases, formation of a law.
3. The Earth can be considered as a system of four vast, interdependent reservoirs: the solid Earth, the atmosphere, the hydrosphere, and the biosphere.
4. Material moves back and forth from one reservoir to another. Some rates of movement are fast, others slow. If a rate of movement changes, the volumes of the reservoirs adjust in response.
5. The Principle of Uniformitarianism states that the internal and external processes operating today have been operating throughout Earth's history.
6. Random, massive, but rare events, such as gigantic meteorite impacts, appear to have played an important role in the Earth's history. These events cause catastrophic change in the Earth's appearance but are not attributed to supernatural forces the way the events of the outdated concept called catastrophism were.
7. The Earth's surface is a place of dynamic interaction between two vast energy sources. The Sun's heat energy drives the Earth's external processes, which involve erosional activity by the atmosphere, hydrosphere, and biosphere. The Earth's internal heat energy drives internal processes in the solid Earth, such as the moving of tectonic plates.
8. Plate tectonics is the slow lateral movement of segments (plates) of the Earth's hard, outermost shell as a result of slow convection currents deep inside the Earth.
9. There are three main methods of heat transfer: radiation (which is the way heat from the Sun reaches the Earth), conduction, and convection.
10. Internal heat reaches the Earth's surface by both conduction and convection. The slow convective motions inside the Earth drive plate tectonics and determine the shapes and locations of the Earth's surface features.

Important Terms to Remember

Terms in italics are defined in A Closer Look.

atmosphere (p. 5)
biosphere (p. 5)
calorie (p. 14)
catastrophism (p. 7)
conduction (p. 12)
convection (p. 13)
convection current (p. 13)
Earth system science (p. 4)
erosion (p. 8)
geothermal gradient (p. 13)
global change (p. 12)
hydrosphere (p. 5)
hypothesis (p. 4)
joule (p. 14)
law (scientific) (p. 4)
plate tectonics (p. 13)
radiation (p. 13)
regolith (p. 5)
rock (p. 5)
scientific method (p. 4)
theory (p. 4)
Uniformitarianism, Principle of (p. 8)
watt {p. 14)

Questions for Review

1. What is the scientific method." Illustrate your answer with an example of the scientific method in practice.
2. How does Earth system science differ from physics, biology, or any other specialized area of science?
3. How does the Principle of Uniformitarianism help us to understand the history and workings of the Earth? Explain why this principle can also be used to understand the solar system and the universe.
4. Suggest three human activities that affect the Earth's external activities in a noticeable manner.
5. Identify three human activities in the area where you live that are causing big changes in the environment.
6. What are the principal energy sources that control the Earth system?
7. How does the Earth's internal heat energy influence the Earth's surface features? Explain why a body such as the Moon or Mars, with a much smaller internal heat source than the Earth's, should have an exterior very different from that of the Earth.

Questions for A Closer Look

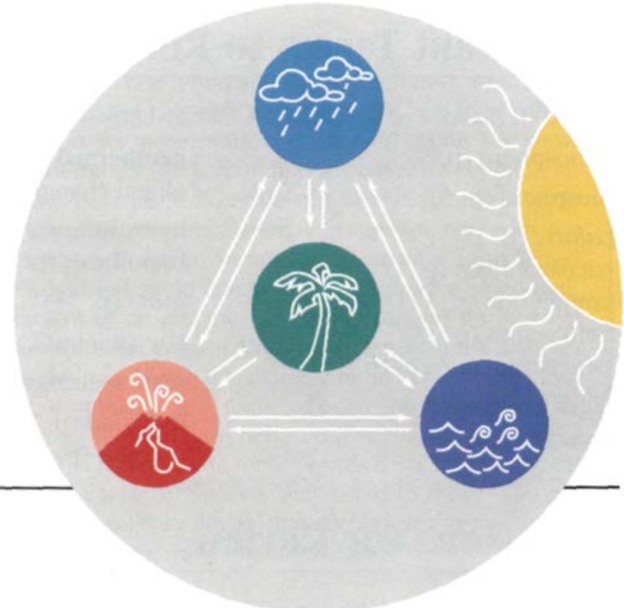
1. Define two different units by which energy is measured. What is the relationship between the two units you have chosen? Why are there different energy units?
2. What is the relationship between power and energy? What are the common units by which power is measured?
3. What are the three sources of energy that control the Earth's activities?
4. The Earth's rate of rotation is thought to be slowing down. What causes the slowdown?
5. How does heat energy inside the Earth reach the surface?

Questions for Discussion

1. Scientists are currently tracking asteroids (small rocky masses that orbit the Sun) and comets because they are concerned that an asteroid or comet might collide with the Earth sometime over the next few hundred years. What effects might the impact of an asteroid or comet have on the Earth? Which branches of Earth science do you imagine might be most involved in the work on asteroids and comets?
2. Is the suggestion that the extinction of the dinosaurs was due to the impact of a large meteorite a hypothesis or a theory? Research some alternative suggestions about the extinction of the dinosaurs. Which of the suggestions would you call uniformitarian, which catastrophism?
3. There is currently a vigorous scientific debate about whether human activity is causing global warming. Research some of the hypotheses about global warming and analyze them in terms of the scientific method.

PART ONE

The Earth in Space



Venus: An Earthlike Planet

When the sun sets and the sky starts to darken, a "star" with a beautiful silvery glow appears above the horizon where the Sun has just disappeared. Except for the Sun and the Moon, this first star of the evening is the brightest object in the sky. Long an object of veneration, this star is actually the planet Venus, named for the goddess of love.

Venus is the nearest planet and the one most like Earth in size and density. In fact, astronomers have long considered Venus and the Earth to be nearly identical twins, and their thoughts have encouraged writers such as Jules Verne to look on Venus as a potential place for human habitation. How wrong the writers and astronomers turned out to be. Venus is actually a terrifyingly hostile place.

Venus is completely shrouded by clouds, but unlike the clouds above the Earth those around Venus never part. The Venus that we see in the sky is simply sunlight reflected off the cloud cover. We had to await the development of radar systems capable of "seeing through" the clouds in order to find out just how different Venus is from the Earth.

The atmosphere of Venus is largely carbon dioxide, and it is so dense and oppressive that the pressure at the surface of this planet is ninety times greater than the pressure at the surface of the Earth. Even worse than the pressure is the temperature. The dense atmosphere makes an effective greenhouse. Radiation from the Sun can penetrate the atmosphere and heat the surface, but the heat given off by the surface cannot quickly escape back into space. As a result, the temperature at the surface of Venus is about 500° C (932°F). The only way anyone could land on Venus would be in a refrigerated spacecraft. Even flying through the upper atmosphere of Venus would be difficult because the clouds are not Earthlike clouds; rather, they are clouds that contain droplets of sulfuric acid!

A full understanding of just how different Venus is has only recently been realized as a result of data from a recent spacecraft visit. On May 4, 1989, a spacecraft named *Magellan* was launched, and on August 10,

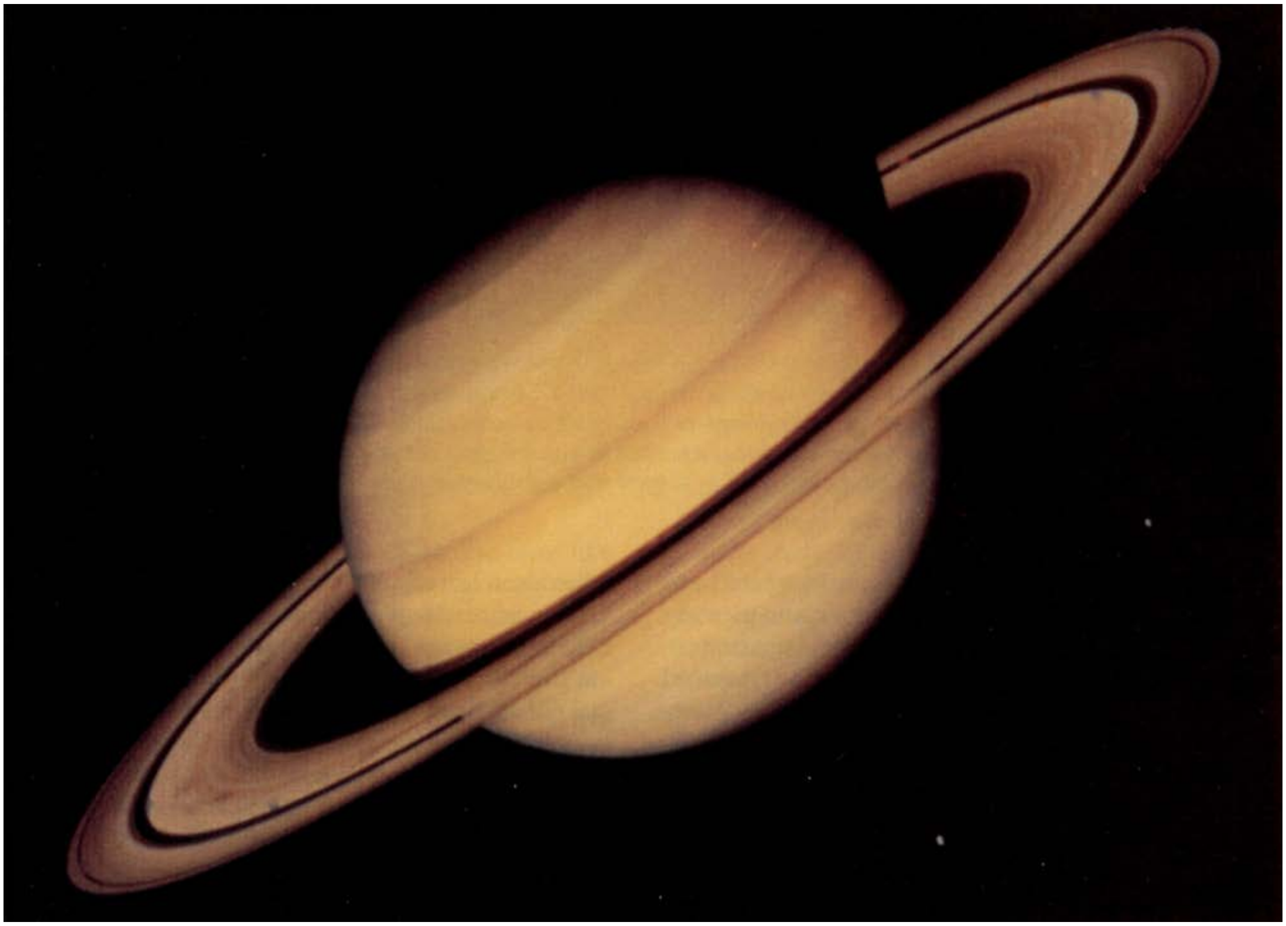
The atmosphere in ultraviolet light as seen by Mariner I in February 1974 (left). Cloud shapes are a consequence of the planet's rotation. The dense atmosphere is opaque to visible and near-visible radiation, (right) A false-color radar image of the surface of Venus. Radar waves penetrate the hostile atmosphere of Venus and reveal a surface of volcanic features and meteorite impact craters. No evidence has been found to suggest that organisms ever lived on Venus.

1990, *Magellan* reached Venus, went into orbit, and commenced mapping the surface with cloud-piercing radar. The *Magellan* mission is one of the greatest triumphs of the space age, and Venus is now revealed for what it is—a planet with an incredibly hostile environment, a planet of vast volcanic plains and mountains higher than Mount Everest, a planet so unlike the Earth that we now have to ask how two planets of about the same size and density could be so close in space and yet be so different. A complete answer to the question cannot, as yet, be given, but part of the answer lies in the biosphere. On the Earth, plants and other organisms in the biosphere remove carbon dioxide released into the atmosphere by volcanoes. Living organisms incorporate carbon as carbon dioxide in limestone or as organic matter buried in rocks in the lithosphere. Because Venus lacks a biosphere, all of its carbon dioxide remains in the atmosphere.

An appropriate place to begin an examination of the Earth and of Earth system science is the solar system. By comparing Earth with the other planets in the solar system, it becomes apparent that the Earth works the way it does precisely because it is the size it is, because of where it is in the solar system, and because of the way the different parts of the Earth system interact.

CHAPTER 1

Fellow Travelers in Space: Earth's Nearest Neighbors



Saturn and its remarkable rings. Note the shadow of the rings on Saturn and the gap, called the Cassini Division. This image was made by the spacecraft Voyager 1.

The Jewel of the Solar System

There are nine planets in our solar system. Seen through a telescope, each of the eight we can see from the Earth has a beauty of its own, but the most beautiful by far is Saturn.

Saturn is one of the brighter objects in the sky. It has a distinctive pale yellow glow and is surrounded by a remarkable ring. One of the first persons to see the ring was Christian Huygens (1629-1695), a Dutch physicist famous for his work with lenses. When Huygens viewed Saturn through his telescope in 1659, he was amazed to see, as he wrote, that the planet is "surrounded by a thin, flat ring." The ring, we now know, is 10,000 km (6214 mi) wide but no more than 10 km (6 mi) thick. When telescopes improved, it became apparent that the ring is really a disc made up of many rings, all lying in the same plane and all centered on Saturn. Just how complex and remarkable the disc of rings is became apparent only following visits from two *Voyager* spacecraft in the early 1980s.

The disc consists of at least a thousand rings, some as small as 2 km (1.2 mi) wide, and the rings form groups with visible gaps between. Five of the ring groups are visible from the Earth: in the *Voyager* image of the rings on the facing page, three of the five can be seen. Within many rings, numerous smaller ringlets can be seen, and even in the largest of the seemingly empty gaps (known as the Cassini Division), faint rings are visible.

Measurements made from the two *Voyager* missions resolved many long-standing debates about the nature of the rings. They are not solid, like a bracelet; rather, each ring is composed of millions of individual particles like beads in a poorly sorted necklace, ranging in size from dust grains to boulders 10 m (11 yd) in diameter. Because each particle is in orbit around Saturn, each is a tiny moon. The particles seem to be mostly ice, but because the rings are pale brown, the ice is thought to be stained by iron oxide. Note that the image of the rings shown on the facing page is a false-color image resulting from computer-aided image processing. The image processing emphasizes color differences between the groups of rings. The color differences probably indicate slight compositional differences, but what the differences are remains a mystery.

The origin of the rings is also a mystery. In addition to the rings, Saturn has 18 large moons with orbits far beyond the outermost ring. It is now known that it is the opposed gravitational pulls of Saturn and the 18 moons that keep the rings in place and prevent the ring particles from accreting to form another moon. However, no one has been able to explain how the rings formed. Whatever their beginnings, we do know that some of the other planets also have rings. Because there are faint rings around Jupiter, Uranus, and Neptune, astronomers conclude that ring formation, whatever the explanation, is apparently a normal part of planetary development.

ASTRONOMY AND THE SCIENTIFIC REVOLUTION

Why does the Sun rise each day and disappear each evening? For much of human history, people believed that the Sun revolved around the Earth. They believed, too, that the Moon, the stars, and the planets also revolve around the Earth. A universe in which a stationary Earth sits at the center and everything else revolves around it is called a **geocentric** universe. Today we are all taught from childhood that the Moon revolves around the Earth and the Earth revolves around the Sun.[1] Proving that what we are taught is correct is a challenging task, and the search for proof that the Earth really does revolve around the Sun was a major factor in the rise of modern science.

Figure 1.1 The celestial spheres.

Ideas from Antiquity

The Greek civilization of antiquity flowered for 800 years from about 650 B.C. to A.D. 150 and spawned many famous philosophers. The most influential of the philosophers, Aristotle (384-322 B.C.), espoused a geocentric universe. He pictured the Sun, the Moon, and the five visible planets as being suspended on concentric, hollow spheres that rotate about an imaginary axis extending outward from the two poles of the Earth, with the stars on the outermost sphere. The star sphere had to be outermost because star positions were fixed relative to each other, but, day by day, the Sun, Moon, and planets could be seen to move in front of the stars. Beyond the star sphere, and invisible to humans, was the realm of the gods (Fig. 1.1).

A few people in Aristotle's time realized that a geocentric universe is not the only way to explain what is seen. The apparent movement of the star sphere across the sky could also be explained if the stars were fixed and the Earth rotated on its axis once every day. Similarly, the fact that there are seasons could be explained if the Earth revolved in an orbit annually around the Sun. One Greek philosopher in particular, Aristarchus (312-230 B.C), favored a Sun-centered, or *heliocentric,* system. Aristarchus used two of the branches of mathematics discovered by the Greeks, geometry and trigonometry, to determine the relative sizes of the Sun, Moon, and Earth. His measurements indicated a huge Sun, a small Earth, and a tiny Moon. It did not make sense to Aristarchus that a huge Sun should rotate around a small Earth. He was unable to convince people that his heliocentric hypothesis might be correct, and so the concept of a geocentric universe continued to be widely accepted until the middle of the sixteenth century, more than fifteen hundred years after the death of Aristotle. In fact, belief in a geocentric universe came to be accepted in most Christian religions as a divine fact.

The Challenge by Copernicus

The most difficult question that a geocentric universe has to answer concerns the motions of the planets. The five visible planets—Mercury, Venus, Mars, Jupiter, and Saturn—look like stars, but they are stars with a difference because they seem to wander. Indeed, the very name planet comes from *planetai,* Greek for wanderers. The paths of the wanderers, measured against the background of fixed stars, are odd. They move a bit farther east each evening, but periodically they slow down and briefly reverse direction before once again resuming their eastward motion. The temporary reversal of direction is known as *retrograde* motion (Fig. 1.2).

The geocentric explanation for retrograde motion is as follows: each planet revolves in an orbit around the Earth and *also* follows a small circular orbit (called an *epicycle)* around an imaginary point, as shown in Figure 1.3. The larger the epicycle, the greater the amount of retrograde motion. The person who worked out the geometry of epicycles in greatest detail and used them to predict planetary positions was

[1] In this chapter we use two words for circular motion: revolve and rotate. *Revolve* means a body moving in an orbit around some central point external to the body; *rotate* means a body spinning around an axis through the body. The Earth revolves once around the Sun each year and rotates once on its axis in 24 hours.

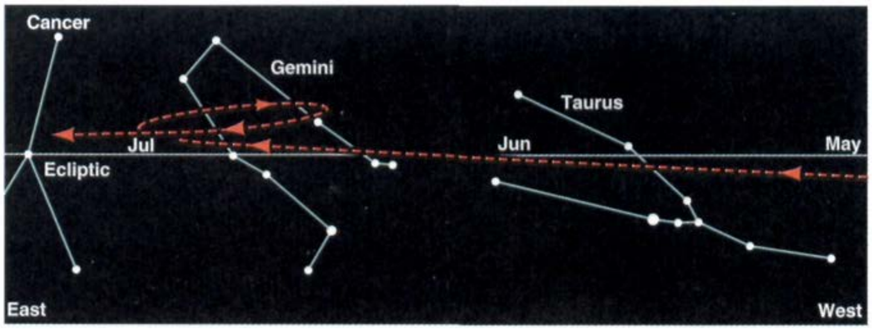


Figure 1.2 The retrograde motion of Mars during June and July 1993. First, the planet moved steadily eastward, passing in front of the constellations Taurus and Gemini. At the eastern edge of Gemini, it suddenly reversed direction and moved westward for a few days before reversing again, passing once more in front of Gemini and continuing in an easterly direction.

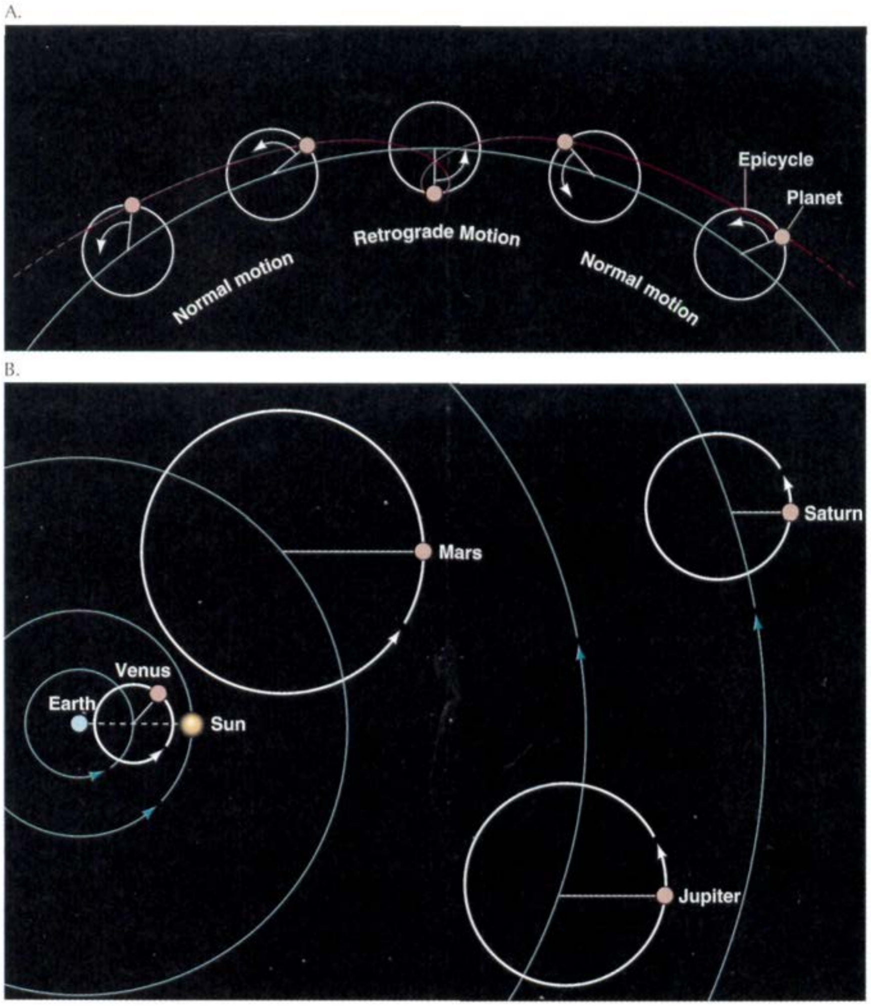


Figure 1.3 The geocentric universe of Aristotle and Ptolemy. A. The planets are imagined to orbit the Earth and also to move in smaller orbits, called epicycles, around an imaginary point on the celestial sphere. B. The size of the epicycle determined the amount of retrograde motion.

Claudius Ptolemy, a man who lived and worked in Alexandria, Egypt. About A.D. 150, Ptolemy published the results of his work in the *Almagest,* one of the most important works we have inherited from the ancient world.

When Nicolaus Copernicus (1473-1543) was a student at the University of Bologna, Italy, in the 1490s, he read a Latin translation of the *Almagest* and decided that the heliocentric system of Aristarchus was more attractive than the geocentric system of Aristotle and Ptolemy. Copernicus recognized that the retrograde motions of planets could be explained in a heliocentric system as a result of differences between the time it takes the Earth to orbit the Sun and the time it takes for any other planet to orbit the Sun, as shown in Figure 1.4. Furthermore, Copernicus sug-

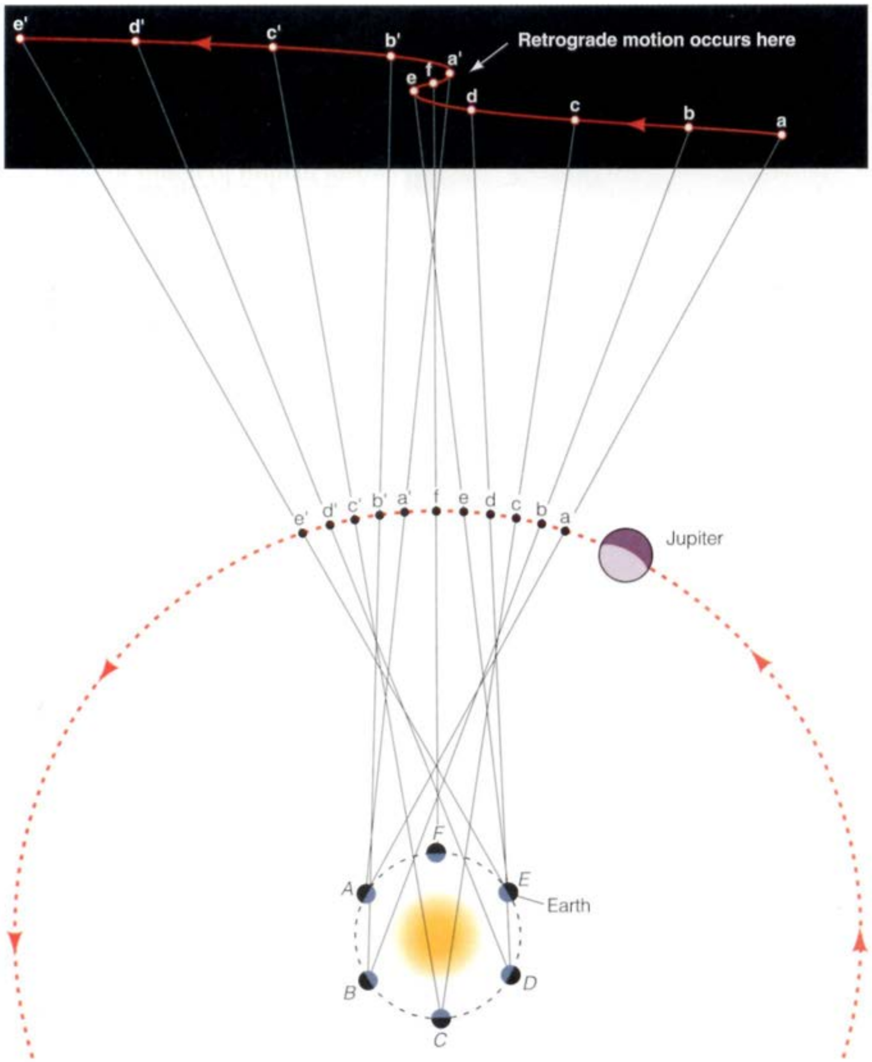


Figure 1.4 The retrograde motion of Jupiter as explained by Copernicus. As Jupiter moves from point *a* to e' in its orbit, the Earth moves counterclockwise from *A*, completely around the Sun and back to *A*, then on to point *E*. An observer on the Earth watching the position of Jupiter against the background of fixed stars would see the green curve as the path of Jupiter across the sky.

gested that because Mars has a larger retrograde motion than Jupiter or Saturn, it must be the closest of the three planets to the Earth, while Saturn, with the smallest motion, must be the most distant.

Copernicus also offered two other major hypotheses. First, he suggested that the positions of the planets at any given time in the future could be predicted by assuming they move in circular orbits around the Sun. As we will discuss, this prediction was not quite correct, but it led to a correct hypothesis that was another stepping stone in the scientific revolution. In the second hypothesis, Copernicus suggested that the Earth spins on its axis. It proved to be quite difficult to establish that the Earth really does spin, and a firm demonstration was not achieved until Jean Foucault (1819-1868), a French physicist, did so in 1851.

By espousing a heliocentric system for the planets, a Moon that orbits the Earth, and an Earth that rotates on its axis, Copernicus offered a direct challenge to the Catholic Church. The Church had built the idea of a geocentric universe into its official doctrine. Looking back, we can see that, by using the scientific method to question a topic accepted as doctrine for over a millennium, Copernicus sowed the seeds that finally separated science from religion and spawned the continuing scientific revolution that has shaped the society in which we live today. It is especially noteworthy that modern science has its roots in astronomical studies and in particular in studies of the motions of the Earth and planets. Earth science is a founding member of modern science.

Kepler and the New Astronomy

When the ideas of Copernicus were published in 1543, they convinced most intellectuals but not all. Among the skeptics was Tycho Brahe (1546-1601), a Dane. In 1572, with funds from King Frederick II of

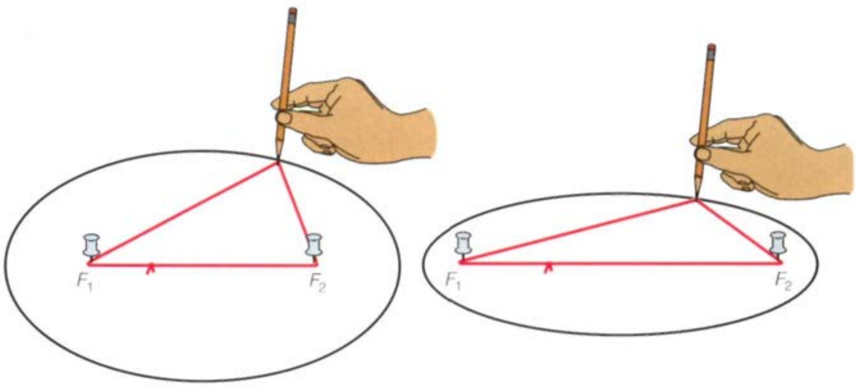


Figure 1.5 How to draw an ellipse. Pin a piece of string at two places and draw a closed figure by pulling the string taut. The two pin points (F_1 and F_2) are the foci of the ellipse. The closer F_1 and F_2 are to each other, the closer an ellipse approaches a circle. The farther apart F_1 and F_2, the greater the eccentricity of the ellipse. Kepler calculated that the orbit of Mars is an ellipse with the Sun at one focus.

Denmark, Tycho built the first modern astronomical observatory on the Danish island of Hven, naming it Uraniborg, or Castle of the Heavens. Optical telescopes had not been invented in Tycho's day, and so all observations were made with the naked eye. Tycho's measurements of planetary positions, made in part to prove Copernicus wrong, were by far the most accurate made up to that time.

Frederick II died in 1588 and Tycho, disliked by Frederick's successor, fell from favor. In 1597 Tycho moved to Prague and while there he hired a young German mathematician, Johannes Kepler (1571-1630), to do his astronomical calculations. After Tycho died, Kepler continued to have access to his numerous measurements of planetary positions. The opportunity was a fortunate one. Kepler, unlike Tycho, thought Copernicus might be right, and he also gave a great deal of thought to a problem Copernicus had not treated—what is the nature of the force that keeps the planets moving around the Sun? Why do they revolve in orbits instead of moving in straight lines out into space?

Because the planets closest to the Sun move faster than those far away, Kepler suggested that a mysterious force must reside in the Sun and have a greater effect on closer objects. Today we know that the force is gravity, but in Kepler's day gravity was an unknown concept. Kepler suggested that magnetism might be that force.

Try as he would, Kepler could not make planetary positions calculated from circular orbits agree with Tycho's measurements. Eventually, Kepler tried calculating the position of Mars based on an elliptical orbit (Fig. 1.5).

Kepler discovered three laws that describe planetary motions:

1. *The law of ellipses.* The orbit of each planet is an ellipse with the Sun at one focus.
2. *The law of equal areas.* A line drawn from a planet to the Sun sweeps out equal areas in equal times (Fig. 1.6). One consequence of the law of equal areas is that orbital speeds are not uniform but instead change in regular ways. A planet moves rapidly when close to the Sun and slowly when far away from the Sun.

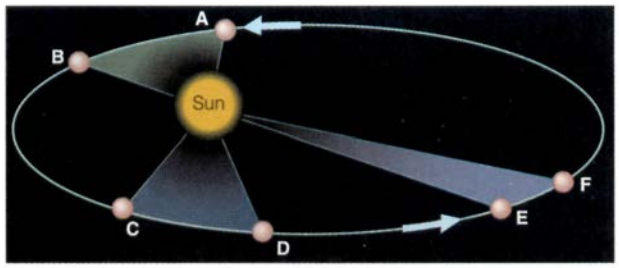


Figure 1.6 Kepler's law of equal areas: Because the orbital speed of any planet varies during the time it takes to complete one orbit, a line connecting the planet and the Sun sweeps out equal areas in equal times. The time it took the planet whose orbit is drawn here to travel from *A* to *B* is exactly the same as the time to travel from *C* to *D* and from *E* to *F*.

3. *The law of orbital harmony.* For any planet, the square of the orbital period in years is proportional to the cube of the planet's average distance from the Sun. The period is the time a planet takes to make one complete revolution around the Sun. (For example, the period of the Earth is 365.24219 days.) Kepler sought to formulate a law to explain that distant planets have long periods while those close to the Sun have short periods (Fig. 1.7). This third law, which describes what Kepler considered to be a cosmic harmony among the planets, can be expressed as

$$p^2 = kd^3$$

where p is the period, k is a constant, and d is the average distance between the planet and the Sun.

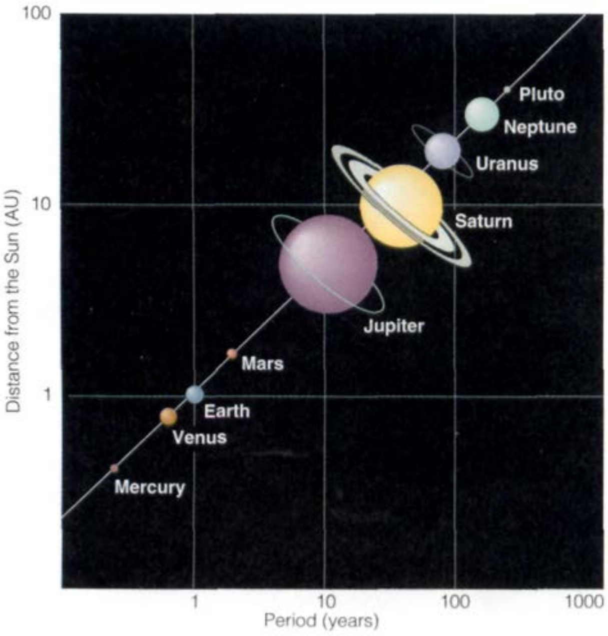


Figure 1.7 Kepler's third law relates the distance of a planet from the Sun to the period of the planet's orbit around the Sun. The unit of distance is the *astronomical unit,* the average distance from the Earth to the Sun. Note the gap between Mars and Jupiter. This is the place in the solar system where the asteroids are found, and some scientists believe that the asteroids are simply rocky fragments that did not accrete to form a planet.

Newton and Galileo

Galileo Galilei (1564-1642) was an extraordinary man who made a great many scientific discoveries. In 1609 he constructed a small telescope with a magnification of thirty times. Turning his telescope to the sky, Galileo viewed mountains on the Moon, discovered that the Milky Way is a dense mass of stars rather than a band of luminous gases, observed four moons in orbit around Jupiter, and saw that Venus, like the Moon, goes through phases from crescent to full. These last two observations sealed the fate of the geocentric universe. Because the moons revolved around Jupiter, the Earth could not be the center around which all objects in the universe revolved. The fact that Venus has phases, and also changes greatly in size, could best be explained if Venus and the Earth are in orbit around the Sun. When the Earth and Venus are on the same side of the Sun, Venus is seen as a crescent. When the Earth and Venus are on opposite sides of the Sun, Venus is seen as a hill disc, but it is only one-seventh the diameter of the crescent because it is so far away. In the old Aristotelian-Ptolemaic system, with Venus in orbit around the Earth, the size of Venus should change very little.

Galileo also made major contributions to our understanding of moving bodies. Motion, he asserted, is due to **a** force, and once a body is moving, it will stop or change direction only in response to another force. If you drop a ball, it falls because some force (gravity) exerted by the Earth pulls it down. Galileo concluded that gravity pulls all falling bodies with the same acceleration. Uniform acceleration means that, in the absence of air resistance, all falling bodies, regardless of their mass, reach the same speed and fall the same distance in the same time. This conclusion reversed a very ancient and deep-seated belief that heavy bodies fall faster than light ones. Most important, it provided one of the key steps by which the force of gravity was discovered and the motions of planets were finally explained. The person who pulled all the pieces together was Isaac Newton.

Isaac Newton (1642-1727) was a genius. First turning his attention to mechanics (the branch of physics that deals with mass, velocity, and acceleration), Newton considered the problem of the force that pulls objects down so that they fall to the Earth when released. Legend has it that he started to think about gravity when he saw an apple fall from a tree. His great leap of insight was to realize that, if the force of gravity acts on an apple, that force must also, as he wrote, extend "to the orb of the Moon." The Moon revolves around the Earth instead of moving through space in a straight line because the force exerted by the Earth's gravity continuously exerts a small pull on the Moon (Fig. 1.8). The force that Kepler had misidentified as magnetism was actually gravity. Newton, through his own insight, discovered one of the universal laws of nature, the **law of gravitation,** which states that *every body in the universe attracts every other body.*

Newton's discovery of the law of gravitation greatly advanced scientific understanding. No longer

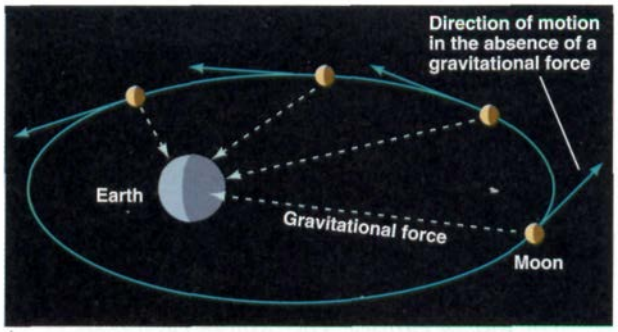


Figure 1.8 The gravitational pull exerted by the Earth on the Moon continuously diverts the moon's direction of motion from a straight line to a closed ellipse.

A Closer Look

Testing Newton's Law of Gravitation

Newton lacked precise measurements; nevertheless, he managed to test the law of gravitation from simple observations and straightforward reasoning.

First, from astronomical measurements he knew that the distance from the center of the Earth to the center of the Moon is approximately 60 R, or 60 Earth radii. Second, he reasoned that the Earth's gravitational force on an apple at the Earth's surface, essentially 1 R from the Earth's center, would be stronger than the Earth's gravitational force on an apple at 60 R. He made the shrewd guess that the force weakened inversely as the square of the distance. Thus, the force on an apple at 60 R would be only $(1/60)^2 = 1/3600$ as strong as the force at 1 R (Fig. C1.1). Third, Newton made use of Galileo's discovery that acceleration does not depend on mass. The acceleration at the Earth's surface due to the force of gravity had been measured at 9.8 meters per second^2 (m/s^2)(32 ft/s^2). Newton hypothesized that the Earth's gravitational force caused the Moon to accelerate at a rate of 9.8 m/s^2 divided by 3600, or 2.7×10^{-3} m/s^2.

Newton tested his hypothesis in the following way. A body moving in a circular orbit—say, a ball rolling inside a cylinder—has an inward acceleration pulling it toward the center of the circle. This inward acceleration is

where v is the linear speed of the body and R is the radius of the cylinder or orbit. Newton knew that the radius of the Moon's orbit around the Earth is 3.84×10^8m. Knowing the time it takes for the Moon to orbit once around the Earth, he calculated the Moon's average linear speed to be 1.02×10^3 m/s. Using these values, he calculated the inward acceleration of the Moon to be 2.7×10^{-3} m/s^2, which was identical to his predicted value and therefore validated his hypothesis.

Note that, by making use of Galileo's discovery, Newton was able to test the law of gravitation without knowing the masses of either the Moon or the Earth.

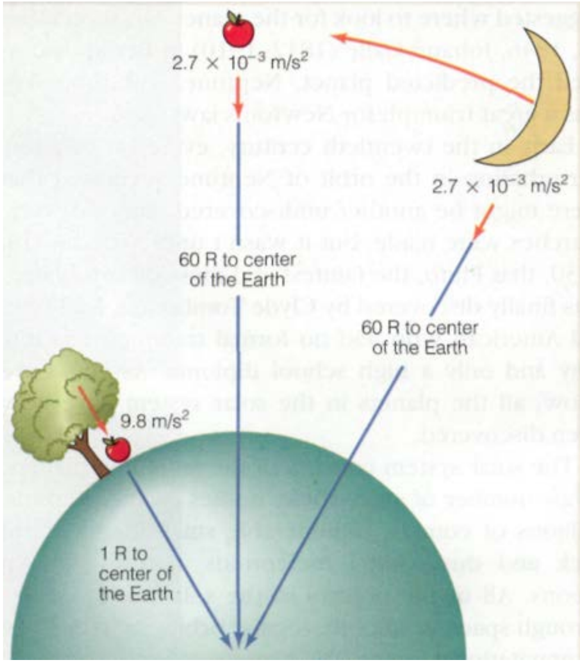


Figure C1.1 Newton reasoned that the force pulling an apple down is the same force that keeps the Moon revolving around the Earth and that the pull is inversely proportional to the square of the distance between the bodies. An apple, very near the Earth's surface, because it is one radius from the Earth's center, falls with an acceleration of 9.8 m/s^2, whereas the Moon, 60 radii away, is pulled toward the Earth with an acceleration that is $1/60^2$ times as large.

was it necessary to call on different forces to describe the fall of apples and the motions of planets. More important, Newton managed to put the law into algebraic form:

$$F = G\frac{M_1 . M_2}{R^2}$$

where F is the force of gravitational attraction between two masses separated by a distance R and having masses M_1 and M_2. The constant G is equal to 6.67×10^{-11} Newton•meter^2/kg^2 in the SI system. (See the appendix for an explanation of this system.) For a discussion of how Newton tested the law of gravitation, see "A Closer Look: Testing Newton's Law of Gravitation".

THE SOLAR SYSTEM

With his crude telescope Galileo could see only the five planets visible to the naked eye. As telescopes improved, however, the other planets of our solar system were discovered. The first of the telescope-discovered planets was Uranus, found by Sir William Herschel (1738-1822) in 1781. Herschel thought he had discovered a new comet, but measurement of the orbit showed it to be a planet. It was soon discovered that Uranus' orbit deviated slightly from the path calculated under the assumption that the only gravitational forces acting on Uranus were those of the Sun and the known planets. Using Newton's law of gravitation to explain the deviations, mathematicians predicted that an undiscovered planet was the cause and suggested where to look for the planet. On September 23, 1846, Johann Galle (1812-1910) in Berlin discovered the predicted planet, Neptune. The discovery was a great triumph for Newton's law.

Early in the twentieth century, evidence of a tiny perturbation in the orbit of Neptune suggested that there might be another undiscovered planet. Several searches were made, but it wasn't until February 18, 1930, that Pluto, the faintest and most distant planet, was finally discovered by Clyde Tombough, a 24-year-old American who had no formal training in astronomy and only a high school diploma. As far as we know, all the planets in the solar system have now been discovered.

The solar system consists of the Sun, nine planets, a vast number of small rocky bodies called asteroids, millions of comets, innumerable small fragments of rock and dust called meteoroids, and 61 known moons. All of the objects in the solar system move through space in smooth, regular orbits, held in place by gravitational forces. The planets, asteroids, comets, and meteoroids orbit the Sun, while the moons orbit the planets.

The Birth of the Solar System

The Sun is a star about 5 billion years old. The universe is at least twice and possibly three times as old as the Sun, and so the Sun is a relatively young star. The birth throes of the Sun and its planets were probably similar to those of billions of other stars, but some of the details remain uncertain. Scientists hypothesize that the solar system formed from a huge, rotating cloud of cosmic gas. One of the key questions that a hypothesis needs to answer is why the Sun and the planets have different compositions. Stars, including the Sun, consist largely of the two lightest chemical elements, hydrogen and helium. Rocky planets like the Earth, Mars, and Venus, on the other hand, consist largely of heavier elements such as carbon, oxygen, silicon, and iron.

One clue concerning the origin of the solar system is provided by the discovery that stars which formed during the earliest moments of the universe contained only the lightest chemical element, hydrogen. From that observation scientists conclude that, initially, hydrogen was the only chemical element in the universe. Stars generate light and heat through nuclear fusion, a process by which hydrogen atoms combine to form helium. As a star ages, hydrogen and helium atoms can combine through nuclear fusion to form still heavier elements. Indeed, the only way elements heavier than helium can form is by nuclear fusion inside stars, and the amounts so formed are tiny by comparison with the amounts of hydrogen and helium present in the universe.

In order for rocky planets to form, the heavy elements inside old stars must somehow be separated from the remaining hydrogen and helium. One hypothesis about the way separation occurs involves a massive star explosion called a supernova (Fig. 1.9). Astronomers have discovered and photographed the scattered remains of many exploded stars, and what they observe is that all of the hydrogen, helium, and heavier elements are scattered into space in a vast cosmic gas cloud. The next step in the process is the formation of a new star and a planetary system from the debris of the cosmic cloud.

Figure 1.9 A supernova.

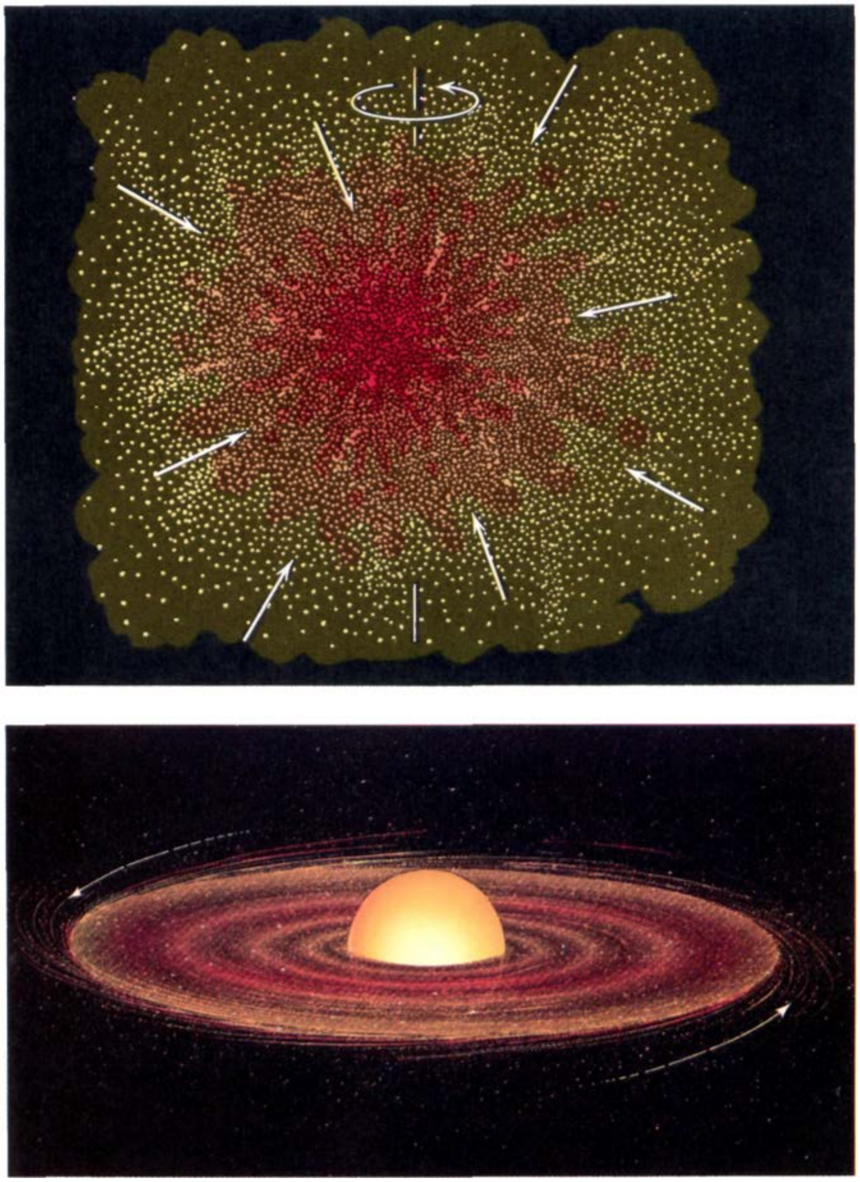


Figure 1.10 Formation of a planetary nebula. The gathering of atoms in space created a rotating cloud of dense gas. The center of the gas cloud eventually became the Sun; the planets formed by condensation of the outer portions of the gas cloud.

We don't know whether the hydrogen now in the Sun and the heavy atoms now in the planets were formed in one ancient star or in several, but scientists have estimated that the atoms now in the Sun and the Earth were part of a cosmic cloud about 6 billion years ago. Though thinly spread, the scattered atoms formed a tenuous, turbulent, swirling cloud of gas. Over a very long period of time, the gas thickened as a result of a slow re-gathering of the thinly spread atoms. The gathering force of the gas was gravity, and as the atoms moved closer together, the gas became hotter and denser as a result of compression. Near the center of the gathering cloud of gas, hydrogen atoms eventually became so tightly pressed and the temperature so high that nuclear burning started again and a new star was born. Surrounding the new sun was a flattened rotating disc of gas and dust, named a *solar nebula* (Fig. 1.10).

By the time the Sun started burning, about 5 billion years ago, the cooler outer portions of the solar nebula had become compacted enough to allow solid objects to condense in the same way that ice condenses from water vapor. The solid condensates eventually became the planets, moons, and all the other objects of the solar system. The planets and moons nearest the Sun, where temperatures are highest, consist mostly of compounds that can condense at high temperatures, mainly silicates, oxides, and iron-nickel alloys. Farther away from the Sun, where the temperatures are lower, only more volatile constituents like sulfur, water and methane were able to condense (Fig. 1.11).

A Closer Look

Time and the Calendar

The calendar is a way of dividing time. The modern calendar, a result of thousands of years of tinkering, is built around the day, month, and year, the lengths of which are determined by three primary astronomical motions. *Rotation* of the Earth, from one sunrise to the next, determines the length of the day; the *revolution* of the Moon around the Earth gives us the month; and *revolution* of the Earth around the Sun causes the cycle of the seasons and determines the length of the year. Unfortunately, the solar day, solar year, and lunar month are not commensurable units of time—the solar year is approximately 365.25 days, while the lunar month is approximately 29.5 days. As a result, the calendar is more complicated than it need be and sometimes needs adjusting.

The Length of the Year

The ancient Egyptians, from whom we derive a considerable portion of our calendar, observed that the Sun and the prominent star Sirius appear on the horizon together at daybreak at 365-day intervals. The Egyptian calendar was therefore based on a year of 365 days. Unfortunately, the calendar makers did not take into account the fact that a year measured by the Earth's revolution around the Sun is approximately 365.25 days, so that after four years the calendar was a day off from the year measured by Sirius and the Sun. The Greeks and the Romans adopted the Egyptian solar year of 365 days, and to clear up the problem of the extra quarter day Julius Caesar institutionalized an older Greek idea and decreed that every fourth year would have 366 days. (That is, the calendar year would *leap* ahead by a day and catch up with the solar year.) The length of a solar year, however, is not exactly 365.25 days (365 days and 6 hours); it is 365 days, 5 hours, 48 minutes, and 46 seconds, making the Julian calendar too long by 11 minutes and 14 seconds, or one day in 128 years. By the time of Pope Gregory XIII, the Julian Calendar was 10 days ahead of the seasons recorded by the solar year, and so the Pope ordered 10 days removed from the calendar: the day after Thursday, October 4, 1582 was decreed to be Friday, October 15, 1582. The Gregorian Calendar, the one used in much of the world today, corrects further misfits by not having a leap year on centennial years that are not divisible by 400. Thus, 1900 was not a leap year but 2000 will be. Even the Gregorian correction is not exact, however, and as a result the calendar year is still moving ahead of the solar year, but only by 26 seconds a year, or one day in 3323 years. Sometime about the year 4500 a further correction will be needed.

The Month

The Egyptians and many other early societies organized their first calendars around the phases of the Moon. A new moon rises approximately every 29.5 days, so 12 lunar months determined by new moons is only 354 days, considerably different from a solar year of 365.25 days. When the Egyptian calendar makers finally settled on a solar year for their calendar, they retained an aspect of their ancient lunar calendar and divided the year into 12 months of 30 days each. The remaining 5 days needed to bring the year up to 365 were simply added on at the end of each year. Such a scheme was quite unacceptable to the methodical Romans. Therefore, they devised the present scheme of some months having 31 days, some 30, and the second month 28 days (or 29 in leap years).

The names of the months are also Roman in origin. Before adopting the Egyptian calendar, the Romans had a calendar of 10 months: Martius (March), Aprilis (April), Maius (May), Junius (June), Quintilis (later changed to Julius, or July), Sextilis (later changed to Augustus, or August), September, October, November, and December. When the Romans adopted the Egyptians' 12-month year, they placed the two extra months at the beginning of the year and named them Januarius (January) and Februarys (February).

The Week

The names given to the days of the week suggest that the origin of a seven-day week might lie in seven celestial bodies and the way they are ranked for astrological purposes.

Viewed from a geocentric Earth, seven bodies move against the background of fixed stars: Mercury, Venus, Mars, Jupiter, Saturn, the Moon, and the Sun. These bodies were considered to be gods who ruled the heavens and controlled the days. Babylonian and Hindu astronomers apparently had metaphysical powers in mind when they ranked the celestial bodies according to the relative ruling powers of the gods as Sun, Moon, Mars, Mercury, Jupiter, Venus, and Saturn. The seven, in that order, gave their names to the day of the week of which each was thought to be in charge. In countries where a Romance language, such as Spanish, is spoken, most of the days are still named for the Roman gods. English names are a mixture, with most names being of Saxon origin.

ROMAN NAMES	*SPANISH NAMES*	*SAXON NAMES*	*ENGLISH NAMES*
Dies Solis	Domingo	Sun's day	Sunday
Dies Lunae	Lunes	Moon's day	Monday
Dies Martis	Martes	Tiw's day	Tuesday
Dies Mecurii	Miercoles	Woden's day	Wednesday
Dies jovis	Jueves	Thor's day	Thursday
Dies Veneris	Viernes	Frigg's day	Friday
Dies Saturnii	Sabado	Seterne's day	Saturday

Hours, Minutes, and Seconds

We inherit the 24-hour day from the ancient Egyptians, who divided the times of daylight and dark into 12 hours each. Because daylight lasts a longer time in summer and a shorter time in winter, the Egyptian hours varied in length through the year. An hour of variable length may be satisfactory for a farmer dealing with matters of the field, but it is a great disadvantage for calculations involving time. The Greeks finally cleared up the confusion about 2000 years ago when they divided the time of the Earth's rotation into 24 units of equal length. To make accurate calculations, the Creeks had to divide up the hour into still smaller units, and to do so they borrowed from the Babylonians. Today we use a counting system based on the number 10, but in ancient Mesopotamia, where the Babylonians lived, a counting system based on the number 60 was in use. The Greeks simply borrowed the Babylonian number system and divided each hour into 60 minutes and each minute into 60 seconds.

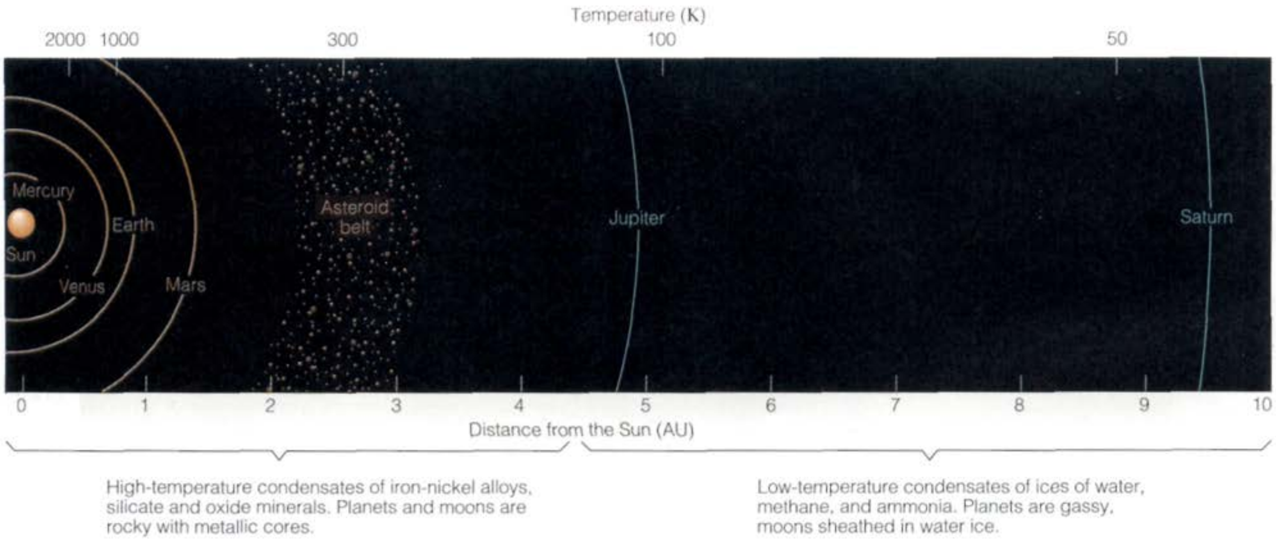


Figure 1.11 Temperature gradient in the planetary nebula. Close to the Sun, temperatures reached 2000 K and only oxides, silicates, and metallic iron and nickel condensed to form planets. Farther away, in the region of Jupiter and Saturn, temperatures were low enough for ices of water, ammonia, and methane to condense.

Condensation of a cosmic gas cloud is only one piece of the planetary birth puzzle. Condensation formed a cosmic snow of innumerable small rocky fragments, but the fragments still had to be joined together somehow in order to form the cosmic snowballs that we call planets. This apparently happened through impacts between fragments drawn together by gravitational attraction. The growth process—a gathering of more and more bits of solid matter from surrounding space—is called **planetary accretion.** Scientists estimate that condensation of the solar nebula and planetary accretion was complete about 4.6 billion years ago.

The revolutions and rotations of the Sun, planets, and moons are inherited from the rotation of the cosmic gas cloud. As the cloud thinned, the planets and moons all formed within the same disk, so that their orbits are all *coplanar,* or in the same plane. All the planets revolve around the Sun in the same direction. (Viewed from space, above the North Pole, the direction of revolution is counterclockwise.) The motions of the planets and moons are so regular and consistent that all societies have used them to keep track of the passage of time. (See "A Closer Look: Time and the Calendar.")

The planets can be separated into two groups based on density and closeness to the Sun (Fig. 1.12A). The innermost planets—Mercury, Venus, Earth, and Mars—are small, rocky, and dense (Fig. 1.12B). Because they are all similar in composition to our Earth, they are called the **terrestrial planets.** *(Terra* is the Latin for Earth.)

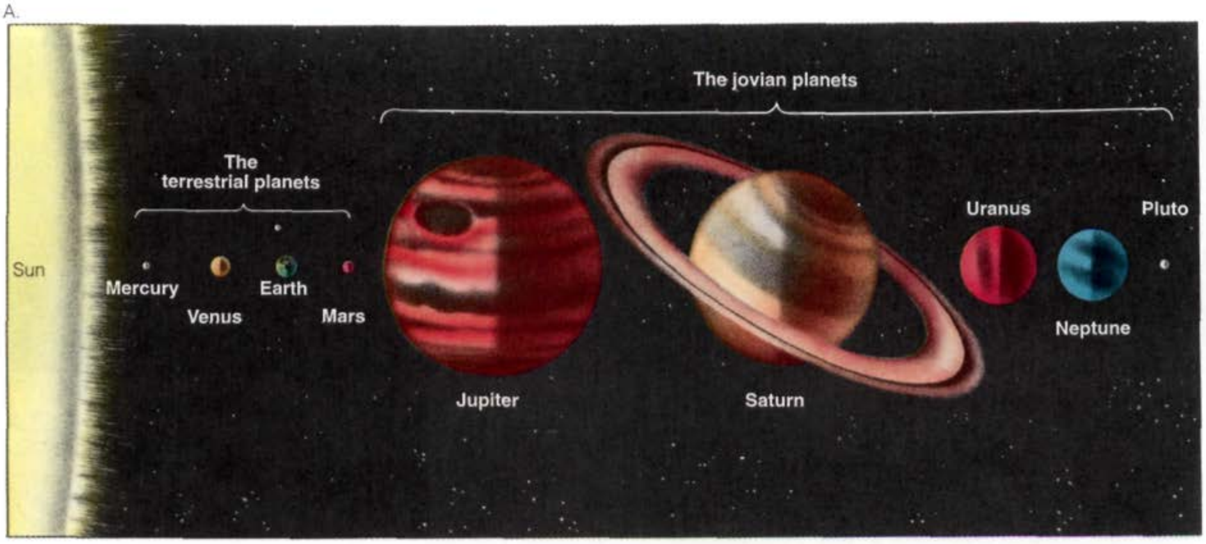
The asteroids are also rocky, dense bodies, but they are too small to be called planets. Refer back to Figure 1.7 showing the distances of the planets from the Sun and note the apparent gap in the sequence between Mars and Jupiter. The asteroids have orbits that fall in this gap, and astronomers hypothesize that they are rocky fragments that failed to accrete into a planet.

The planets farther from the Sun than Mars (with the exception of Pluto) are much larger than the terrestrial planets, yet much less dense. These **jovian planets**—Jupiter, Saturn, Uranus, Neptune, and Pluto—take their name *from Jove*, an alternative designation for the Roman god Jupiter.

Evolution of the Planets

Space missions have provided abundant evidence that all the objects in the solar system formed at the same time and from a single solar nebula. During the final phase of planetary accretion, the Moon and the four terrestrial planets became so hot that they all under-

A.



B.	Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune	Pluto
Diameter (km)	4880	12,104	12,756	6787	142,800	120,000	51,800	49,500	6000
Mass(Earth=1)	0.055	0.815	**1**	0.108	317.8	95.2	14.4	17.2	0.003
Density, g/cm^3 (water=1)	5.44	5.2	5.52	3.93	1.3	0.69	1.28	1.64	2.06
Number of moons	0	0	1	2	16	18	15	8	1
Length of day (in Earth hours)	1416	5832	24	24.6	9.8	10.2	17.2	16.1	154
Period of one revolution around Sun (in Earth years)	0.24	0.62	1.00	1.88	11.86	29.5	84.0	164.9	247.7
Average distance from Sun (millions of kilometers)	58	108	150	228	778	1427	2870	4497	5900
Average distance from sun (astronomical units)	0.39	0.72	1.00	1.52	5.20	9.54	19.18	30.06	39.44

Figure 1.12 The planets and their properties. A. The planets, shown in their correct relative sizes and in the correct order outward from the Sun. The Sun is 1.6 million km in diameter, 13 times larger than Jupiter, the largest planet. B. Numerical data concerning the orbits and properties of the planets.

went a period of partial melting. As a result, they separated into layers of different composition. The thick, cloud-encircling atmospheres of the jovian planets obscure details of the evolutionary history of those planets, so the following remarks refer only to the Moon and the terrestrial planets.

During and after melting and compositional separation, the Moon and the four terrestrial planets continued to be struck by rains of meteorites. Although meteorite impacts still do happen, the period of nearly continuous massive impacts ended more than 4 billion years ago. From about 4 billion years ago to the present, the terrestrial planets and the Moon seem to have evolved along somewhat different paths.

Three key factors played the determining roles in the evolution of the terrestrial planets. First, after partial melting, the planets remained hot inside because radioactive elements were and still are present. All the terrestrial planets are cooling down, but the rates of cooling are determined by the sizes of the planets and the rates vary greatly. The largest planets, Venus and the Earth, are cooling very slowly and therefore are still relatively hot today. One important indication of high internal temperature is volcanism, which continues on the Earth and possibly on Venus. Volcanic activity has occurred on Mars within the past billion years, but Mars is probably not active today. Both the Moon and Mercury, the two smallest bodies, have been volcanically dead for billions of years.

The second factor that controlled the way the terrestrial planets evolved is their distance from the Sun. The Sun-planet distance determines whether or not H_2O can exist as water and hence whether or not there can be oceans. The two planets closest to the Sun—Mercury and Venus—are too hot for liquid water to occur. Venus does have H_2O in its atmosphere, but the temperature at the surface of Venus is close to 500°C or (932°F). Mars, which is farther from the Sun than is the Earth, is too cold to have liquid water but does have ice.

The third factor is the presence or absence of a biosphere. The hydrosphere and the biosphere play essential roles in biogeochemical cycles that control the composition of the atmosphere. If life had evolved on Venus, that planet might have developed an atmosphere like the Earth's. On the Earth, plants and microorganisms have enabled carbon dioxide and water to combine, through photosynthesis, to make organic matter and oxygen. The burial of organic matter in sediment in effect removes carbon dioxide and at the same time adds oxygen to the atmosphere. Because life did not develop on Venus, all of the CO_2 is still in the atmosphere, and as a result Venus suffers from a horrendous greenhouse effect.

The Earth system and its many parts came into being a long time ago. What that system is today, and how the many parts interact, are very much a product of the Earth's long history and of the two great heat engines that drive it: the solar engine, which has warmed the Earth's surface for the last *4.6* billion years, and the internal heat engine, which drives all the activities of the solid Earth.

THE TERRESTRIAL PLANETS

Each of the terrestrial planets and the Moon have the same gross structure, consisting of three layers distinguished by differences in composition.

Layers of Different Composition

The structure common to all the planets is most clearly demonstrated in the Earth (Fig. 1.13). At the center is the densest of the three layers, the **core,** a spherical mass composed largely of metallic iron, with lesser amounts of nickel and other elements. The thick shell of dense, rocky matter that surrounds the core is called the **mantle.** The mantle is less dense than the core but denser than the outermost layer. Above the mantle lies the thinnest and outermost layer, the **crust,** which consists of rocky matter that is less dense than mantle rock.

Each of the terrestrial planets has a core, mantle, and crust, but there are considerable differences in detail, particularly in the crust. For example, Figure 1.13 shows that the core and the mantle of the Earth have nearly constant thicknesses, but the crust is far from uniform and differs in thickness from place to place by a factor of nine. The crust beneath the oceans, the **oceanic crust,** has an average thickness of about 8 km (5 mi), whereas the **continental crust** averages 45 km (28 mi) and ranges from 30 to 70 km (19 to 44 mi) in thickness. The two different kinds of crust are the result of the special internal processes that shape the Earth's surface, and in particular, plate tectonics. The crusts of the other terrestrial planets are thicker than the Earth's crust and approximately uniform in thickness. The uniformity of thickness is an indication that plate tectonics does not, and probably never has, been active on any of the other terrestrial planets.

Because we cannot see and sample either the core or the mantle of a planet, it is valid to ask how we know anything about their composition. The answer is that indirect measurements are used, and again the Earth is used as an example. One way to determine composition is to measure how the density of rock changes with depth below the Earth's surface. We can do this by measuring the speeds with which earthquake waves pass through the Earth because the speeds are influenced by rock density (see Chapter 3). At some depths, abrupt changes in the speed of earthquake waves indicate sudden changes in density. From the sudden changes, we infer that the solid Earth consists of distinct layers with different densities. Knowing these different densities, we can estimate what the composition of the different layers must be.

Slight compositional variations probably exist within the mantle, but we know little about them. We can see and sample the crust, however, and the sampling shows that, even though the crust is quite varied in composition, its overall composition and density are very different from those of the mantle, and the boundary between them is distinct.

The composition of the core presents the most difficulty. The temperatures and pressures in the core are so great that materials there probably have unusual properties. Some of the best evidence concerning core composition comes from iron meteorites. Such meteorites are believed to be fragments from the core of an asteroid, large enough to be a planet, that was shattered by a gigantic impact early in the history of the solar system. Scientists hypothesize that this now-shattered asteroid must have had compositional layers similar to those of the Earth and the other terrestrial planets.

Layers of Different Rock Strength

In addition to compositional layering, the sphere that is our Earth can be divided into three layers based on differences in the strength of the rock that makes up each layer: the mesosphere, asthenosphere, and lithosphere (Fig. 1.13).

The strength of a solid is controlled by both temperature and pressure. When a solid is heated, it loses strength; when it is compressed, it gains strength. Differences in temperature and pressure divide the mantle and crust into three distinct strength regions. In the lower part of the mantle, the rock is so highly compressed that it has considerable strength, even though the temperature is very high. Thus, a solid region of high temperature but also relatively high strength exists within the mantle from the core-mantle boundary (at 2883 km, or 1791 mi depth) to a depth of about 350 km (218 mi) and is called the **mesosphere** ("intermediate, or middle, sphere") (Fig. 1.13).

Within the upper mantle, from 350 to about 100 km (218 mi to 62 mi) below the Earth's surface, is a region called the **asthenosphere** ("weak sphere"), where the balance between temperature and pressure is such that rocks have little strength. Instead of being strong, like the rocks in the mesosphere, rocks in the asthenosphere are weak and easily deformed, like butter or warm tar. As far as geologists can tell, the compositions of the mesosphere and the asthenosphere are the same. The difference between them is one of physical properties; in this case, the property that changes is strength.

Above the asthenosphere, and corresponding approximately to the outmost 100 km (62 mi) of the Earth, is a region where rocks are cooler, stronger, and more rigid than those in the plastic asthenosphere. This hard outer region, which includes the uppermost mantle and all of the crust, is called the **lithosphere** ("rock sphere"). It is important to remember that, even though the crust and mantle differ in composition, it is rock strength, not rock composition, that differentiates the lithosphere from the asthenosphere.

The boundary between the lithosphere and the asthenosphere is caused by differences in the balance between temperature and pressure. Rocks in the lithosphere are strong and can be deformed or broken only with difficulty; rocks in the asthenosphere below can be easily deformed. One analogy is a sheet of ice floating on a lake. The ice is like the lithosphere, and the lake water is like the asthenosphere.

Layers of Different Physical State

Metallic iron in the Earth's core exists in two physical states. The solid center of the Earth is the **inner core.** Pressures are so great in this region that iron is solid despite its high temperature. Surrounding the inner core is a zone where temperature and pressure are so balanced that the iron is molten and exists as a liquid. This is the **outer core.** The difference between the inner and outer cores is not one of composition. (The composition of the two is believed to be the same.) Instead, the difference lies in the physical states of the two: one is a solid, and the other is a liquid.

Comparison of the Terrestrial Planets

The terrestrial planets, and possibly the Moon, seem to have had similar early histories. Where ancient surfaces exist, as on the Moon, Mercury, and the south-

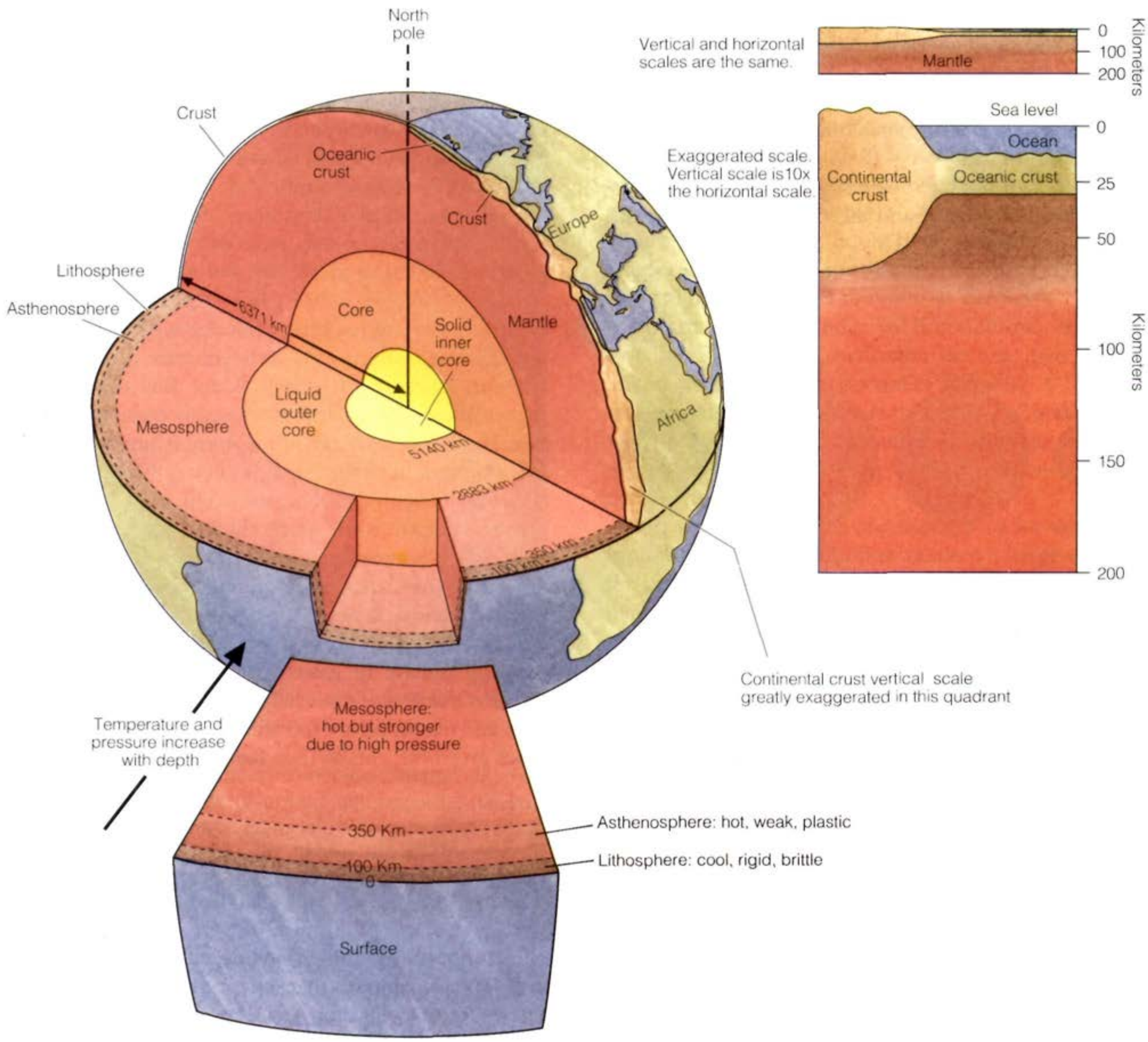


Figure 1.13 A sliced view of the Earth reveals layers of different composition and zones of different rock strength. The compositional layers, starting from the inside, are the core, the mantle, and the crust. Note that the crust is thicker under the continents than under the oceans. Note, too, that boundaries between zones of different physical properties—lithosphere (outermost), asthenosphere, mesosphere—do not coincide with compositional boundaries.

ern haff of Mars, evidence of a violent period of planetary accretion remains. Each body seems to have experienced a period of heating during which a core formed. The striking feature about the various cores, the sizes of which are calculated from the densities of the planets, is how greatly they differ in relative size (Fig. 1.14A). The most remarkable body is Mercury, for on this planet the core is 42 percent of the volume and an estimated 80 percent of the mass. At present, we cannot assert with any certainty whether any of the terrestrial planets besides the Earth have molten or partially molten cores. The molten outer core and the relatively rapid rotation of the Earth give rise to the Earth's strong magnetic field. Magnetic fields do exist on the other planets, but they are much weaker than the Earth's field.

Spacecraft have landed on the Moon, Mars, and Venus, and on those bodies we have been able to make direct measurements of the crust. Flyby missions to Mercury reveal that a crust is present there too. The existence of a core and a crust suggests a mantle, and the necessary measurements have been made on the Moon, Venus, and Mars to establish that indeed mantles are present. We can be reasonably sure, therefore, that the structures of all the terrestrial planets are similar.

Whether or not each terrestrial planet has a lithosphere, asthenosphere, and mesosphere is a more difficult question to answer. Simple observation reveals that rocks on the surface of each planet fracture and deform as they do on the Earth, and this indicates that a lithosphere is present. Astronauts left instruments on the Moon to measure the properties of moonquakes, and from those measurements the presence of an asthenosphere can be inferred (Fig. 1.14B), but the presence of a mesosphere seems unlikely. Measurements made on Mars have determined that an asthenosphere exists there, too, but for Venus and Mercury it is possible only to infer the existence of an asthenosphere. What little evidence we have suggests that asthenospheres and lithospheres probably are present in each terrestrial planet but that the asthenosphere of the Earth is unusually close to the surface and hence that the lithosphere is unusually thin. It is probable that the Earth is such a dynamic planet *because* its lithosphere is thin. The other terrestrial planets seem to have much thicker lithospheres and to be much less dynamic than the Earth.

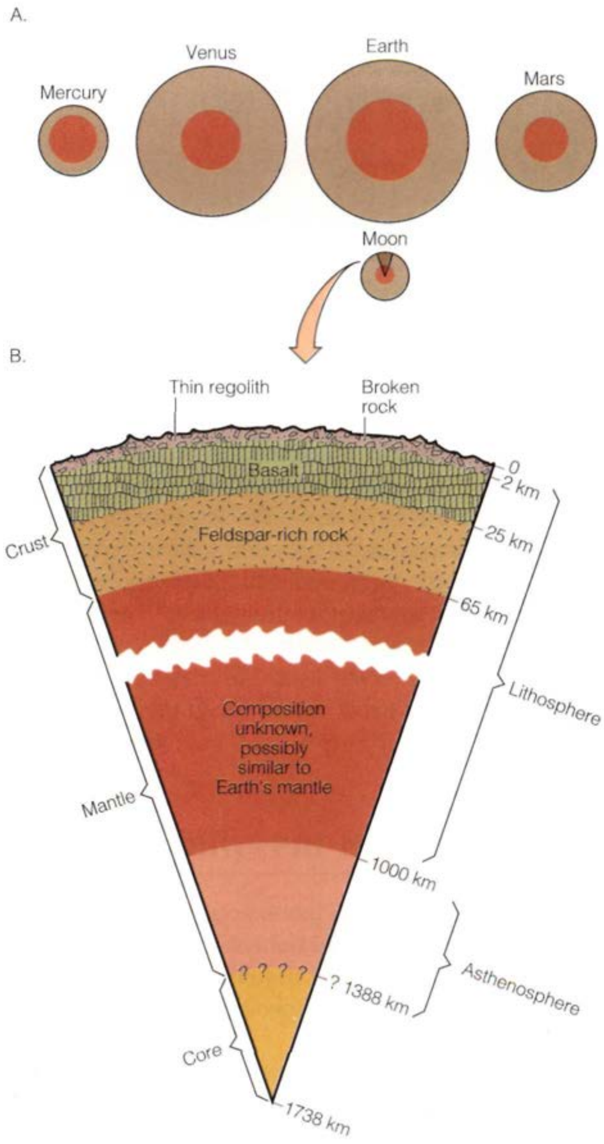


Figure 1.14 The internal structures of the Moon and the terrestrial planets. A. Comparative sizes of the cores. Mercury, nearest the Sun, where only the highest temperature materials could condense, has a huge core. Mars, farthest away from the Sun, has a small core. B. Structure of the Moon. Crusl composition is known with certainty-only in the vicinity of the astronauts' landing sites.

Venus, the Earth, and Mars are large enough that their gravitational fields have been able to retain the atmospheres formed as a result of melting and *outgassing:* release of gases from rocks or other nongaseuos materials, especially through volcanoes. Mercury and the Moon are too small to have held on to the gases given off, and so they lack atmospheres.

THE JOVIAN PLANETS

We cannot see anything that lies below the thick blankets of atmosphere that cover the jovian planets. Therefore, we can only hypothesize about their internal structure, based on remote-sensing measurements of various kinds. For example, we can calculate that the masses of Jupiter and Saturn are so great that none of their atmospheric gases has been able to escape their gravitational pulls. This is true even for the two lightest gases, hydrogen and helium, which made up the bulk of the planetary nebula. This means, therefore, that the bulk composition of the two largest jovian planets must be about the same as that of the solar nebula from which they formed. For example, the composition of Jupiter is estimated to be 74 percent hydrogen, 24 percent helium, and 2 percent heavy elements.

Because the moons of the jovian planets are rocky with thick sheaths of ice (Fig. 1.15), it is presumed that a rocky mass resides at the center of each planet. The rocky cores of Jupiter and Saturn may be as large as 20 or more earth masses. Surrounding the rocky

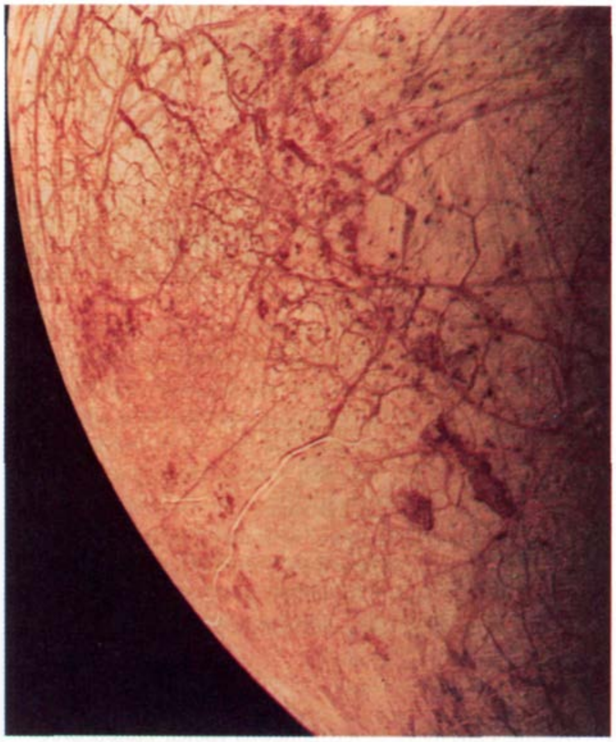


Figure 1.15 Europa, smallest of the four large moons of Jupiter. Europa has a low density, indicating it contains a substantial amount of ice. The surface is mantled by ice to a depth of 100 km. The fractures indicate that some internal process must be disturbing and renewing the surface of Europa. The dark material (here appearing red) in the fractures apparently rises up from below. The cause of the fracturing is not known. The image was taken by *Voyager 2* in July f979.

cores is possibly a layer of ice, analogous to the ice sheaths seen on the moons (Fig. 1.16).

Pressures inside the jovian planets must be enormous; we may therefore hypothesize that deep in the interiors hydrogen may be so tightly squeezed that it is condensed to a liquid. Proceeding inward from the outer atmosphere, which consists mostly of hydrogen gas, we hypothesize that a point is soon reached where a thick layer of liquid hydrogen is present. Still deeper inside Jupiter and Saturn, pressures equivalent to 3 million times the pressure at the surface of the Earth are reached. Under such conditions, the electrons and protons of hydrogen become less closely linked and hydrogen becomes metallic; a layer of molten metallic hydrogen is the result. In Jupiter pressures may even reach values high enough for solid metallic hydrogen to form a sheath around the ice core.

Neptune and Uranus are thought to be similar to Jupiter and Saturn, although neither is large enough for pressures to be sufficiently high to form metallic hydrogen. Pluto, the planet farthest from the Sun and the smallest of the jovian planets, is a little larger than the Moon. Pluto lacks the massive atmosphere of the larger jovian planets and has a density of 2.06 g/cm^3, intermediate between the high densities of the terrestrial planets and the low densities of the other jovian planets. The most probable explanation for the density of Pluto is that the planet has a structure like those of the moons of Jupiter and Saturn—a rocky center but a thick outer layer of ice.

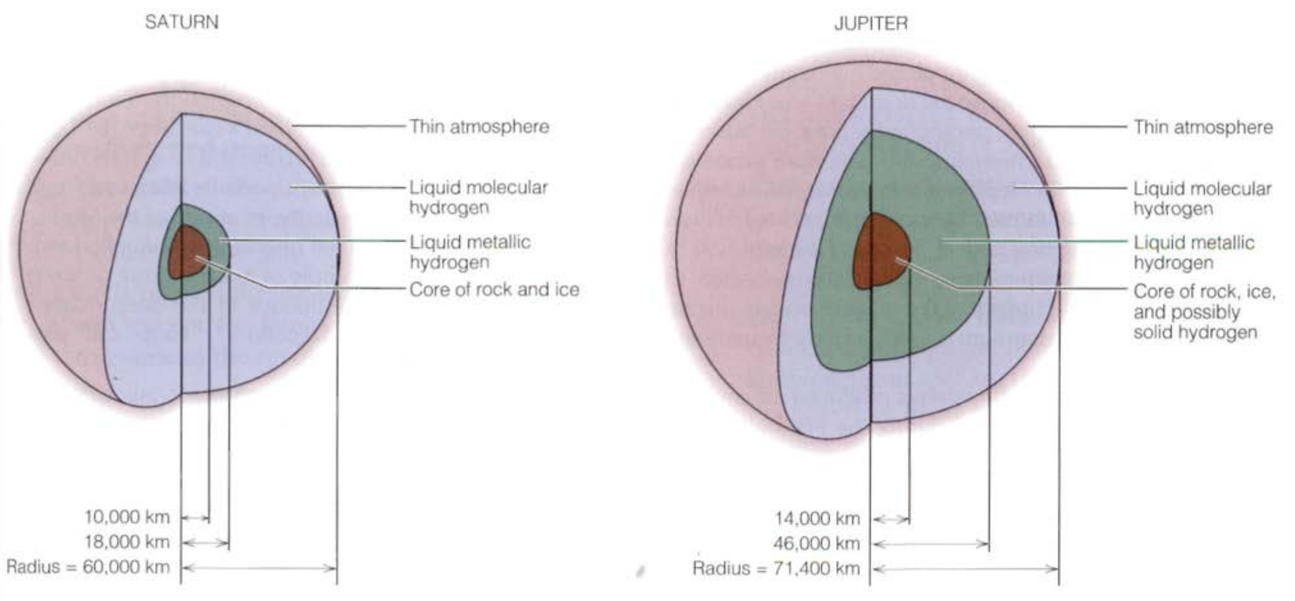


Figure 1.16 Comparison of the probable interior structures of Jupiter and Saturn.

Guest Essay

The First Stepping Stone in Space

Harrison H. Schmitt born on July 3, 1935, in Santa Rita, New Mexico, received a B.S. from Caltech and a Ph.D in geology from Harvard. He became a NASA scientist and jet pilot in 1965, landing on the moon with the Apollo 17 mission on December 11, 1972. After organizing and managing NASA's Energy Programs Office, he was elected to the U.S. Senate from his home state in 1976 and served one term. Schmitt married freelance writer Teresa Fitzgibbon in 1985, currently lives in Albuquerque, and consults and writes on lunar science, technology, nature, business, and public policy topics.

Beginning in the seventeenth century, scientific understanding of the Moon began to grow through more accurate observations, facilitated by the invention of the telescope and increasingly sophisticated scientific logic. Knowledge accelerated rapidly in the 1960s with direct measurements by space probes. Hands-on investigations of lunar features by humans first became possible when, on July 20, 1969, Apollo 11 with Neil Armstrong and Buzz Aldrin aboard landed on Mare Tranquillitatis. Three-and-a-half years later Eugene Cernan and I completed the six Apollo landings on the moon with the Apollo 1 7 mission to the Valley of Taurus-Littrow.

Explorations of the Moon during the Apollo Program provided a comprehensive identification of major events and processes through which the Moon evolved as a small terrestrial planet. Furthermore, these explorations led to an increased understanding of the early history of all the terrestrial planets, particularly the Earth. How the Moon began almost 4.6 billion years ago, however, remains a subject of intense speculation. In recent years, broad but not unanimous support has developed for a beginning after the collision between a young but geologically evolved Earth and a "Mars-sized asteroid." Others believe that the details of lunar geology require that the Earth captured a small, independently formed planet.

Whatever happened to create our only natural satellite, a hot, splattering, wave-tossed ocean of molten rock covered the Moon to a depth of at least 450 km during its final formative stages. We have learned about this magma ocean from the light-colored highlands and dark "mare" basins that comprise the dominant lunar features. With the aid of samples, photographs, personal observations by the Apollo astronauts, and data from geophysical instruments they deployed, we know, for example, that the highlands formed between 4.5 and 4.4 billion years ago. During that period, a silicate mineral rich in calcium and aluminum (anorthite feldspar) floated to form a crust on the surface of the magma ocean. Silicate minerals rich in magnesium (olivine and pyroxene) sank to form a mantle.

Continuously impacting debris produced an environment of incredible violence during the first half billion years of lunar history. Once the magma ocean solidified, a crust between 50 and 70 km thick had formed over the mantle. Between about 4.4 and 4.2 billion years ago, intense cratering continued to fracture and mix the crust, producing extremely heterogeneous rocks, called breccias, that form the outer 25 km of the Moon.

Then, about 4.2 billion years ago, highly energetic objects began to hit the Moon, forming at least 45 large circular basins more than 300 km in diameter. Extensive melting and redistribution of crustal materials accompanied the formation of large-impact craters, such as the 500-km-diameter Serenitatis basin. The 2300-m-deep, east-west valley of Taurus-Littrow, in which my Apollo 1 7 mission landed, crosses the high mountain rim of Serenitatis.

Our mission, in addition to three others, provided samples and observations on lunar volcanic lavas, called mare basalts. Magmas that formed the mare basalts resulted from the accumulation of radiogenic heat, which melted portions of the lunar mantle. Rising through the mantle and crust, these magmas erupted from innumerable vents for at least 900 million years after the large basins formed. Floods of the highly fluid lava, some flowing over a thousand kilometers, partially filled most low-lying regions, particularly on the near side of the Moon.

One of the least expected returns of any Apollo landing would have been a sample of material from portions of the Moon below the influence of the deep magma ocean. Our discovery of the Apollo 1 7 "orange soil" provided just such a return.

Finding the orange soil, a product of volcanic fire fountains, resulted from a serendipitous interaction of pre-mission planning, subtle visual observation, and the power of suggestion. "Shorty," an 80-m-diameter dark halo crater, was thought to be either an impact crater or a volcano that penetrated the light-colored dust and rock of a large avalanche. If a volcano, we speculated that hot gases might have altered the surrounding debris in some recognizable and possible colorful way.

Visual inspection of Shorty soon disclosed that it was an impact crater that had punched through the avalanche and ejected underlying dark basaltic material. The thought processes, however, that tested the various hypotheses on Shorty's possible volcanic origin played a key role in attracting my attention to the very light orange coloration in the gray dust covering the orange soil.

Analyses of the composition of glass beads that comprise the orange soil, and of the volatile elements and compounds adhering to those beads, indicate that the magma from which they had formed 3.5 billion years ago originated between 400 and 500 km depth near the base of the lunar mantle and that the volatiles came in part from original planetary (primordial) materials below the mantle. These conclusions may not be compatible with the hypothesis mentioned above, that the Moon formed by impact-induced fission of a nonprimordial Earth.

The vast majority of the huge volume of scientific data accumulated by the first phase of lunar exploration has come from analyses of samples, instrument readings, and photographs. The foundation for interpreting these data resulted from extensive training given the crews of the Apollo missions. Frequent simulations in a variety of real geological locales on Earth enabled the astronauts to respond quickly to the total situation encompassed by their explorations. They had become field geologists. Field geologists' personal involvement in future exploration and development of the first stepping stone in space probably will be essential to success. Robotic systems can make increasingly important contributions. But the simultaneous observation, integration, and interpretation of the total dynamic and often unpredictable situation inherent in space activities, and a calculated human response to that situation, will be as irreplaceable in the future as they have been throughout geology's past.

Summary

1. The solar system consists of the Sun, nine planets, 61 known moons, vast numbers of asteroids, millions of comets, and innumerable meteoroids.
2. The planets revolve around the Sun in elliptical orbits. The moons revolve around the planets, also in elliptical orbits. All revolve counterclockwise viewed from above.
3. The orbits of the other planets and all moons are approximately coplanar with the orbit of the Earth about the Sun.
4. Each planet and the Sun rotate around an axis, and except for Venus, the rotation direction is the same as the revolution direction.
5. The solar system formed through the condensation of a solar nebula followed by planetary accretion and was completed about 4.6 billion years ago.
6. The evolutionary history of a planet is controlled by its size, the distance from the Sun, and the presence or absence of life.
7. The planets can be divided into two groups: the terrestrial planets, the four nearest the Sun, each a small, rocky mass with a high density; and the jovian planets, the five outermost planets, each, with the exception of Pluto, large and gassy.
8. The terrestrial planets are compositionally zoned into a metallic core, a mantle, and a crust. Terrestrial planets also have layers that differ in strength. In the case of the Earth, the outermost layer, about 100 km thick, is hard and rigid and is called the lithosphere. Beneath the lithosphere is a region about 250 km thick called the asthenosphere where rocks are soft, weak, and easily deformed. Beneath the asthenosphere is the mesosphere, a region where high pressures keep rocks strong despite the high temperatures.

9. At the center of each of the terrestrial planets is a metallic core composed mostly of iron but with admixtures of nickel and other elements. The Earth's core has an inner component that is solid and an outer component that is molten.
10. The jovian planets, with the exception of Pluto, the smallest, are shrouded by thick atmospheres rich in hydrogen and helium. The cores of the jovian planets are inferred to be rocky, like a terrestrial planet, and to be surrounded by a thick layer of ice. Above the ice, liquid hydrogen grades outward to the hydrogen-rich atmosphere. Pluto has only a rocky core and an outer sheath of ice.

Important Terms to Remember

Asthenosphere (p. 36)
Continental crust (p. 35)
Core (p. 35)
Crust (p. 35)
Geocentric (p. 24)
Gravitation, law of (p. 28)
Heliocentric (p. 24)
Inner core (p. 36)
Jovian planet (p. 34)
Lithosphere (p. 36)
Mantle (p. 35)
Mesosphere (p. 36)
Oceanic crust (p. 35)
Outer core (p. 36)
Planetary accretion (p. 33)
Solar nebula (p. 31)
Terrestrial planet (p. 33)

Questions for Review

1. What are the differences between a geocentric and a heliocentric solar system?
2. What is retrograde motion of the planets? Ptolemy and Copernicus had different explanations for retrograde motion; how did the explanations differ?
3. What great conceptual advance did Kepler make concerning the orbits of planets?
4. Why do planets move in orbits around the Sun rather than through space along straight paths?
5. The existence of the two outermost planets, Neptune and Pluto, were predicted prior to their discoveries. On what were the predictions based?
6. Describe the two primary motions of the Earth with respect to the Sun.
7. What is a solar nebula and what role is a nebula thought to have played in the formation of the solar system?
8. The planets can be divided into two distinctly different groups; what is the basis of such a division? What are the names of the planets in each group?
9. Terrestrial planets are compositionally layered. Describe the layering and explain how it is thought to have happened.
10. What is a lithosphere? An asthenosphere? Sketch a section through the Earth showing the positions of the lithosphere, asthenosphere, and mesosphere.
11. Three major factors are thought to control the evolutionary history of a planet; what are the factors and what controls do they exert?

Questions for A Closer Look

1. What is Newton's law of gravitation? What clever guess made by Newton led to the law?
2. How did Newton test his law of gravitation?
3. What astronomical motions are used to measure the passage of time?
4. Why did Julius Caesar and Pope Gregory XIII adjust the calendar?

Questions for Discussion

1. If it were possible to send an unstaffed spaceship to a nearby star in order to inspect the planets in orbit around the star, what kind of measurements would you advise the space scientists to try to obtain as the spaceship flies past each planet?

2. Astronomers have observed a number of supernovas in recent years: research their findings and discuss the influence of the discoveries on hypotheses about the origins of planets.

Update: Flotsam in Space

All of the planets and moons in the solar system have been peppered through the geological ages by innumerable impacts with the meteorites, comets, and asteroids that make up bits of space flotsam—the scars are everywhere. The most recent near-collision between Earth and a comet fragment took place on June 30, 1908, when a massive explosion occurred in the atmosphere over the Tunguska River, in Siberia. The explosion flattened 3000 km^2 of forest and severely burned people 60 km away from the explosion. No solid object actually hit the Earth; a fragment about 300 m in diameter may have exploded in the atmosphere. No one is known to have witnessed the explosion. No one is known to have seen any other impact on the Earth or any other planet—until July, 16, 1 994, when fragments of a broken comet, named Comet Shoemaker-Levy 9, began to slam into the giant planet Jupiter.

When Comet Shoemaker-Levy 9 came close to Jupiter in 1 992, the gravitational pull of the planet broke the body into 19 large bits (the largest was about 3.2 km in diameter) and thousands of small bits. The close approach also changed the orbits of the fragments so that between July 16 and July 22, 1994, all of them slammed into Jupiter. Astronomers on the Earth were waiting with telescopes poised, and as this book went to press the first photographs of the impacts became available. The results are remarkable.

Most of the impacts occurred on the back side of Jupiter and the results could only be seen when the planet rotated the impact site into view (Jupiter rotates every 9.8 hours). The impact of fragment H on July 1 7 could be seen, however, and the event was photographed with a telescope set up at the South Pole (Fig. 1). The impact produced an intense flash of light as the 3 km-wide body penetrated an estimated 200 km into the atmosphere, broke apart, and was completely vaporized as a result of its high velocity—1 50,000 km/h. The plume of hot gas from the impact rose 2,200 km above the top of Jupiter's atmosphere and for a short time was as bright as Jupiter itself. The energy released by the impact is estimated to be equivalent to the explosion of 6 million megatons of TNT. (The largest nuclear device ever exploded on the Earth, a hydrogen bomb, produced a blast of about 5 megatons.)

The outermost layers of Jupiter are all gas and many people expected that the impact sites would quickly disappear. The reverse happened; huge boiling centers of disturbed gas, some as wide as the Earth, continued to scar the surface of Jupiter many days after the impact (Fig. 2). Much will be learned about the composition and structure of Jupiter's atmosphere as data are analyzed over the months and years ahead; there will be lessons for the Earth, too. One lesson is already apparent: space flotsam that slams into the atmosphere may cause devastating damage even though no solid object manages to reach the surface. Another lesson is that the recent paucity of big impacts on the Earth is no guarantee that we can rest easy. Millions of objects too small to be detected in space could hit the Earth and cause unimaginable damage.

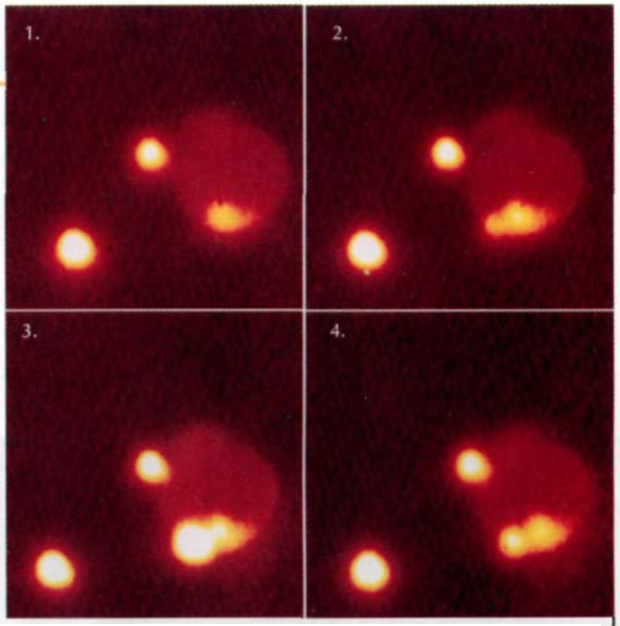
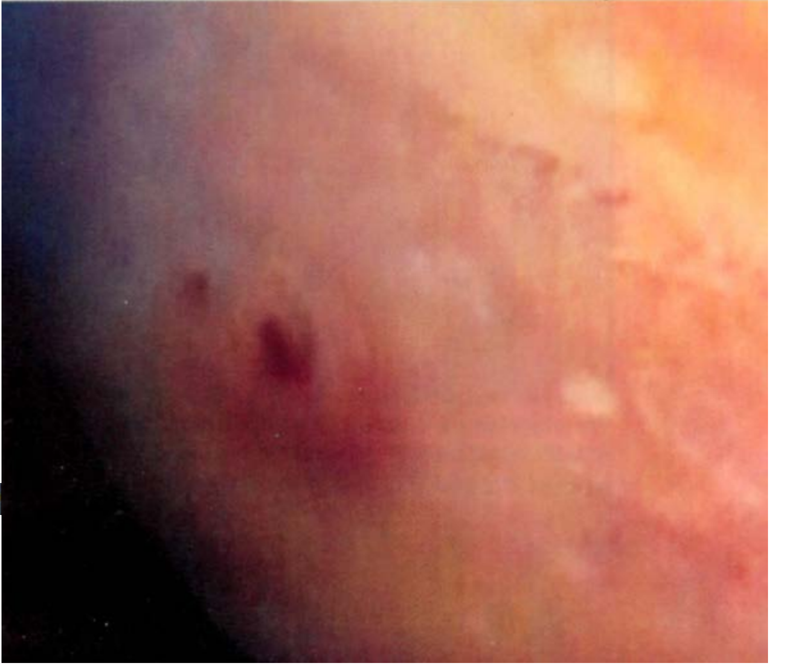


Figure I The impact of fragment It of Comet Shoemaker-Levy 9 on Jupiter. Jupiter is the diffuse red circle. The two bright spots (left) are the moons Io (closest to Jupiter) and Ganymede. In the first frame the scar from the earlier impact of fragment G is seen to the lower right. In frame 2, fragment II is iust starting to impact, Frame 3 shows the peak of the impact and frame 4 the scars of the G and II impacts side-by-side. These photos were taken within a 25 minute period by a camera mounted at the South Pole.

Figure 2 The impact area of fragment G, Comet Shoemaker-Levy 9, photographed through the Hubble Space Telescope. The image, which was made about 20 hours after the impact, shows a center of boiling gas that is the size of the Earth.



CHAPTER 2

The Sun, Giver of Life

The object in the constellation Andromeda known as M31 is a huge spiral galaxy consisting of millions of stars. The galaxy is 2.2 million light years from the Earth. Two smaller galaxies, N6C205 and M32, are also visible, one each side of M31.

How Many Suns in the Universe?

With the exception of the planets and their moons, each point of light in the night sky—each "star"—is a sun. On a clear, dark night, it's easy to count as many as 5000 stars. With an ordinary pair of binoculars, almost a million stars are visible, and with the aid of the most powerful telescopes, the number of stars that can be seen rises to the billions. In fact, the number of stars turns out to be so large that no one has ever tried to make an exact count. The best we can do is estimate.

Even a cursory look at the night sky reveals that the visible stars are not evenly distributed. When telescopes were invented several hundred years ago, one of the earliest discoveries scientists made was that stars occur in clusters. We now call each cluster of a billion or more stars a galaxy. A further discovery made with telescopes is that our Sun (and all the planets that circle it) are in a galaxy called the Milky Way.

The number of stars in the Milky Way is mind boggling enough, but modern telescopes reveal an estimated 100 billion other galaxies in the universe! Most of the other galaxies are so distant from us that they appear as single, tiny, fuzzy-looking stars through all but the largest telescopes. With those largest telescopes, however, astronomers can confirm that each fuzzy "star" is indeed a galaxy. Now, multiply the estimated number of galaxies (10^{n}) by the minimum number of stars in a galaxy (10^{9}). The result, a hundred billion billion (10^{20}), is an estimate of the minimum number of stars in the sky and therefore of the minimum number of suns in the universe. The reason no one has managed to make an exact count of all the suns is obvious—trying to do so would be like trying to count all the sand grains on all the beaches, all the river banks, and all the deserts in the world!

THE LIFE-GIVING PROPERTIES OF THE SUN

The Sun is a vast ball of gas, and at its center is a huge nuclear reactor. The light and heat generated by this fusion reactor control the Earth's climate and make the Earth a habitable planet. It is because three of the Earth's reservoirs—the biosphere, hydrosphere, and atmosphere—derive their energy from the Sun that the Earth system works the way it does. If the amount of light and heat were either more or less, our planet would be a very different place.

Because it is situated 150 million km (93 million mi) from the Sun, the Earth receives just the right amount of light and heat to support life. Venus and Mercury, the two planets closest to the Sun, are too hot and too dry for life to exist. Mars is so far from the

Sun that it is too cold for life and also too dry because H_2O does not exist there as water, only as ice and water vapor.

Most living species draw their energy from the Sun, so that the Sun is both giver and supporter of life. Plants take their energy directly from sunlight by the process called photosynthesis, and animals get their energy by eating plants or by eating other animals that eat plants.

If life is possible on the Earth because of the Sun's energy, could life exist on other planets in other solar systems? We don't know, but a lot of research has been carried out trying to derive an answer. As a result of space research, we are reasonably sure that, with the exception of the Earth, life does not exist on any of the planets or moons in our solar system. However, we can barely see the outer planets in our solar system, and so it is not surprising that astronomers are still trying to prove that planets exist around other stars. Although most scientists theorize that most suns have planets in orbit around them and hypothesize that life exists on at least some of these other planets, proving the point will be very difficult.

We do know a great deal about our Sun, however. We can even estimate when and how the Sun will die—and therefore when the Earth system will die. Our knowledge comes from analyzing the Sun's energy output, which takes approximately 8.3 minutes to reach the Earth. We can also analyze the energy output from other stars and then, as we will see in the rest of the chapter, use the Principle of Uniformitarianism to estimate how long the nuclear fusion reactor has been operating in the Sun, how steadily it sends out light and heat, and how long it will continue to do so in the future.

THE SUN'S VITAL STATISTICS

Our Sun is an ordinary, medium-sized, middle-aged, run-of-the-mill star with properties and characteristics identical to those of billions of other ordinary, medium-sized stars.

Size

The Sun is vastly larger than any other body in the solar system. At approximately 1.4 million km (0.87 million mi), the diameter of the Sun is 109 times the diameter of the Earth. We say approximately because it is hard to decide exactly where the edge of a ball of gas is. What is meant in the Sun's case is the edge of the glowing, visible sphere, even though there is a transparent blanket of gas several thousand kilometers deep outside of the glowing sphere.

The Moon is about 382,000 km (237,000 mi) from the Earth. If we were to draw a sphere the size of the Sun centered on a point at the center of the Earth, the edge of the sphere would be about 315,000 km (196,000 mi) beyond the Moon. Because of its vast size, the Sun's mass, 2×10^{30} kg, is 300,000 times greater than the Earth's mass. However, because the Sun is entirely gas, its density is only one-fourth that of the Earth.

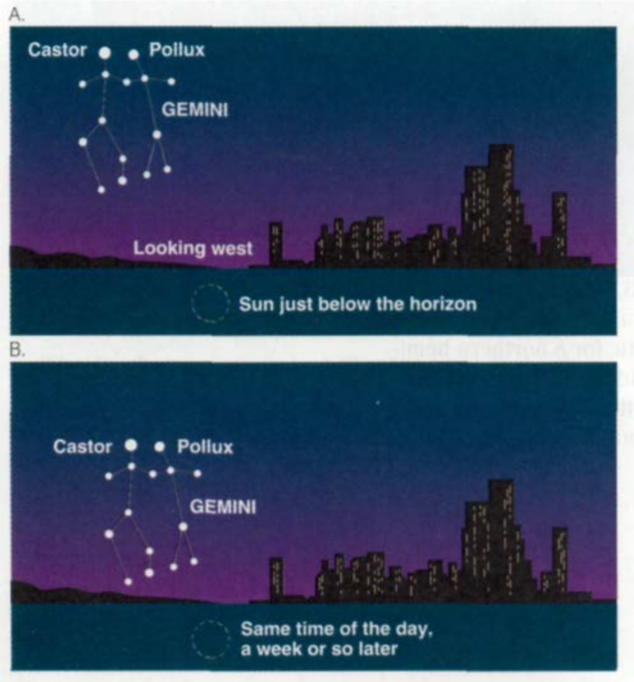
Apparent Motion

The Earth revolves around the Sun, but to an observer on the Earth it is the Sun that *appears* to revolve around the Earth. Because of the Earth's daily rotation about its axis, the Sun seems to arc across the sky from east to west every day. Thus, this first apparent motion of the Sun is a result of the Earth's rotation. If the stars were visible during the day, you would notice another apparent motion. Relative to the background of fixed stars, the daily arc traveled by the Sun is in a different place in the sky from one day to the next. This second apparent motion of the Sun is due to the Earth's revolution around the Sun. The way to see this relative movement for yourself is to pick out a group of bright stars—perhaps a recognizable constellation—located close to the western horizon soon after sunset. Note the exact time, your exact location, and the distance from the horizon to your star group. Wait a week or two and then look at the same stars, at exactly the same hour of the evening, and with you standing in exactly the same spot. You will notice that the stars have moved closer to the horizon and therefore closer to the position of the Sun (Fig. 2.1).

Each distinctive star pattern in the sky is called a constellation. The constellations are named mostly for animals and mythical characters, and most of the names we use today we inherited from the ancient Greeks and Babylonians. The sky is divided into 88 constellations (Appendix 000), which are a convenient way to divide the sky for purposes of location. For example, we can describe Castor and Pollux as the two bright stars in the constellation Gemini. (This is the constellation shown in Fig. 2.1.)

Relative to the stars, the Sun moves to the east, and in one year it moves completely around the sky and returns to its initial position. This means that the Sun makes a complete circuit of 360° in a year, or about 1° a day. The ancient Babylonians observed that the Sun's apparent eastward motion through the sky takes it through the same 12 constellations each year. You can make the same observations yourself if, at regular intervals through the year, you observe where the Sun sets relative to the stars. The 12 constellations through which the Sun passes are called collectively the **zodiac** (Fig. 2.2), a Greek term meaning circle of animals.

Remember that the Sun's motion against the background of fixed stars is an *apparent* motion. The body that actually moves is the Earth in its orbit around the Sun. The plane of the Earth's orbit is called the *ecliptic*, and so the path of the Sun through the constellations of the zodiac is the trace of the ecliptic on the background of fixed stars. If the Earth's axis of rotation were exactly perpendicular to the plane of its orbit, the trace of the ecliptic would be a straight line. In fact, the axis of rotation is tilted at 23.5° to the plane of the orbit, and so the trace of the ecliptic is a smooth curve (Fig. 2.3).



Energy Output

The total amount of energy radiated outward each second by the Sun or any other star is called the **luminosity.** It takes 8.3 minutes for energy (in the form of light and heat) radiated from the Sun's surface

Figure 2.1 An easy way to prove that the Sun moves relative to the more distant stars. A. Just after sunset, noting the exact time and your exact location, pick a bright constellation close to the western horizon near the point where the Sun sets. The constellation shown here is Gemini, close to the horizon in July. In other months choose other constellations. B. Observe your constellation a week later, at the same time of the evening as the initial observation. Your stars will be closer to the horizon and therefore closer to the Sun. After a few weeks, your constellation will set before the Sun, so that it is absent when you look at the sky just after sunset.

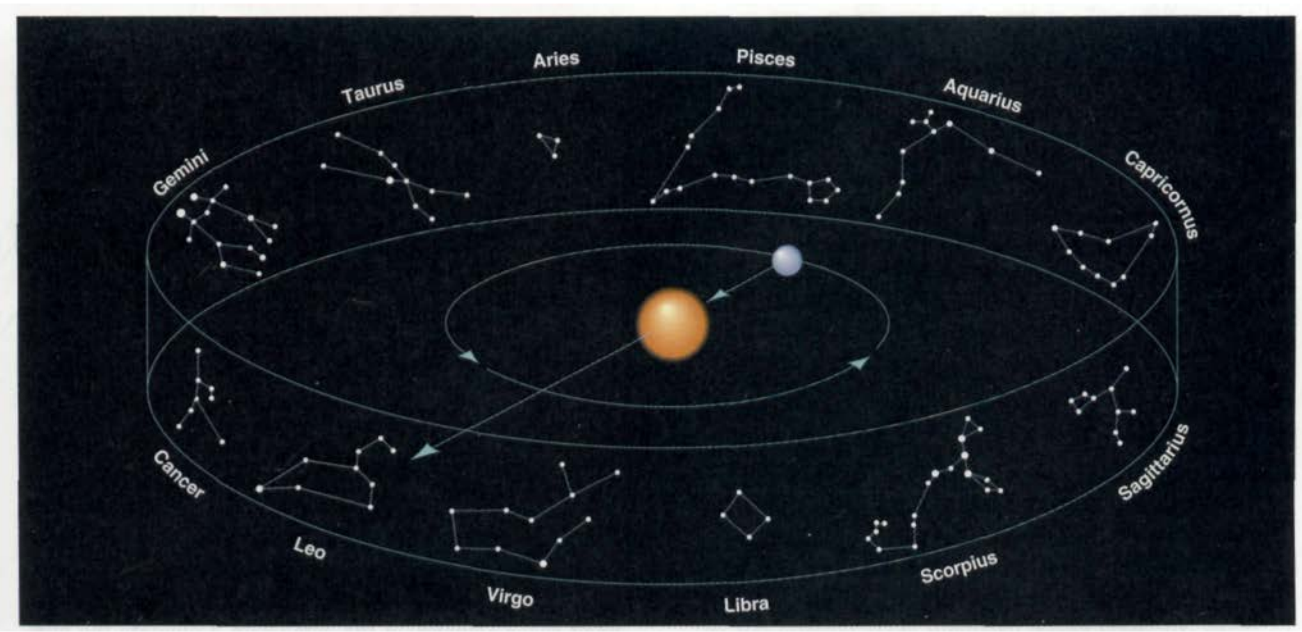


Figure 2.2 The apparent motion of the Sun as seen by a viewer on the Earth. The Sun appears to move when viewed against the background of fixed stars. The constellations through which the Sun's apparent motion takes it make up the zodiac.

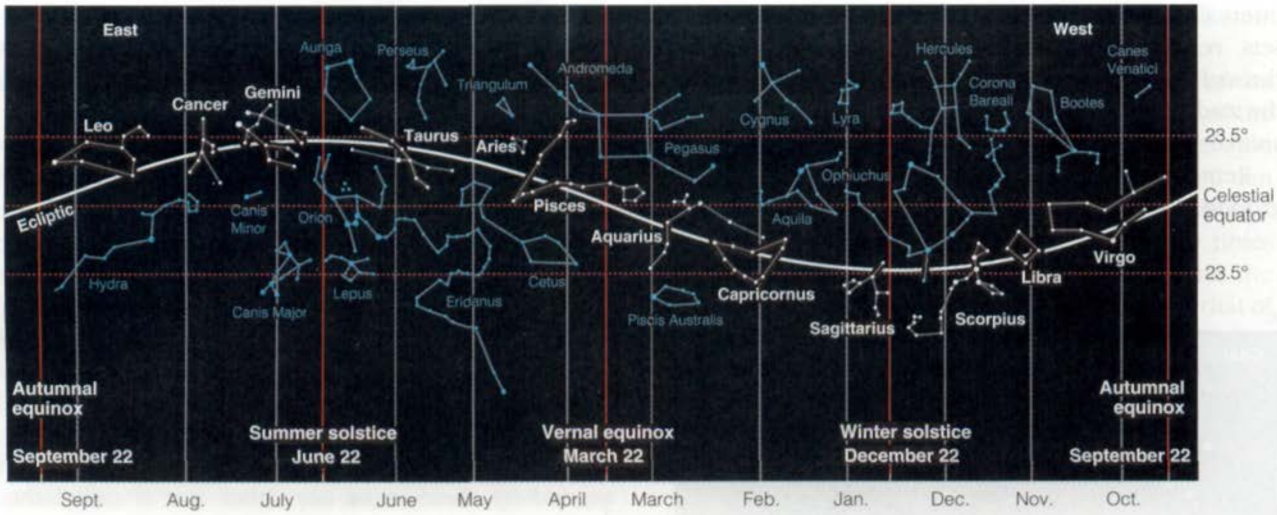


Figure 2.3 The ecliptic and the constellations of the zodiac. The trace of the ecliptic is curved because the Earth's axis of rotation is tilted at 23.5° to the plane of the Earth's orbit. A viewer on the Earth is therefore sometimes above the plane of the ecliptic, sometimes below. The maximum point above the ecliptic for a northern hemisphere viewer is the summer solstice on June 22, when the midday Sun is at its most northerly point. The winter solstice is December 22, when the Sun is at its most southerly point. The times when the Sun is directly over the equator at midday are the vernal (spring) and autumnal equinoxes.

to reach the Earth. Because the Sun radiates energy equally in all directions, only a tiny fraction of the total energy it emits reaches the Earth. Therefore, an Earth-bound scientist wanting to determine the Sun's luminosity must do some calculating.

Artificial satellites that orbit the Earth get their energy, via solar panels, from the Sun. From such satellites we know that solar energy reaches the Earth at a rate of 1370 watts per square meter (1.2 sq yd) of surface. That's enough energy to light over 13 100-watt bulbs. When energy continuously passes through or continuously falls on a unit area, we say there is an energy **flux** through or on that area. The energy flux reaching the Earth from the Sun is therefore 1370 watts/m^2. The energy flux can be used to calculate the Sun's luminosity in the following manner. Picture an imaginary sphere with a radius equal to the Earth-Sun distance (1.5×10^{11} m) and centered on the Sun (Fig. 2.4). Such a sphere has a surface area of $2.8 \times 10^{23} m^2$. The energy flux through every square meter on the surface of the sphere is 1370 watts. The total energy output of the Sun must therefore be the number of square meters multiplied by the flux:

$$2.8 \times 10^{23} m^2 \times 1370 \text{ watts/m}^2 = 3.8 \times 10^{26} \text{ watts}$$

and this is the Sun's luminosity.

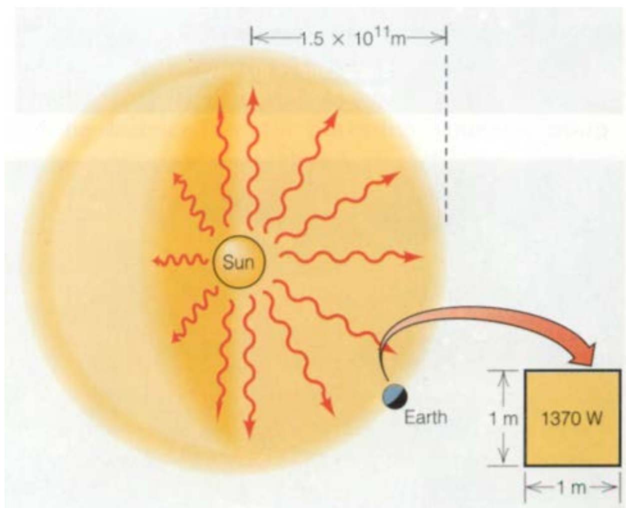


Figure 2.4 To measure the Sun's luminosity, create an imaginary sphere that is centered on the Sun and has a radius equal to the average Earth-Sun distance, 1.5×10^{11}m. The inside surface of this imaginary sphere would capture all of the energy radiated by the Sun. Energy from the Sun reaches the Earth at a rate of 1370 w/m^2, so every square meter of the inside surface of the imaginary sphere must receive energy at the same rate. Calculate the number of square meters in the surface from the formula, surface area = $4pr^2$, multiply by 1370 W/m^2, and you have the Sun's luminosity, 3.8×10^{26}W.

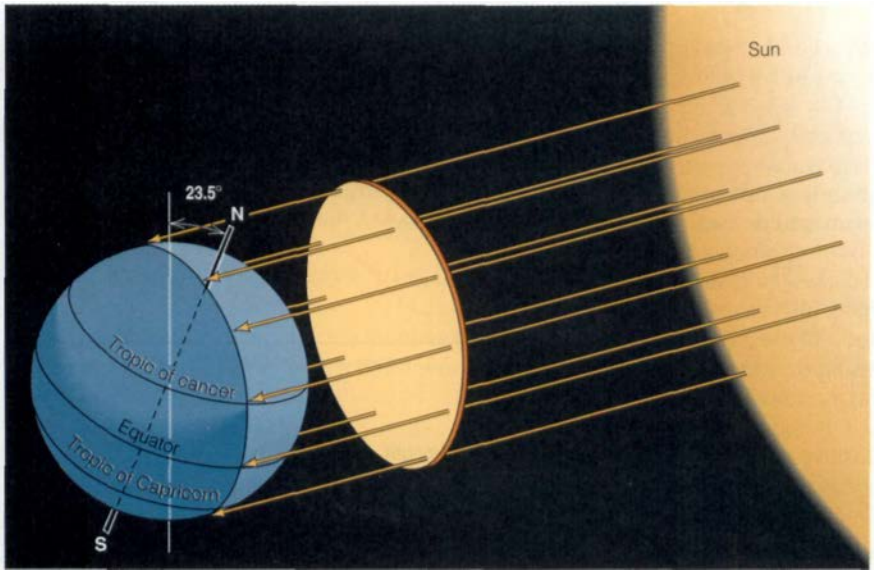


Figure 2.5 Energy from the Sun falls on an imaginary disc that has a diameter equal to the Earth's diameter. Each square meter of the disc receives energy at a rate of 1370 watts. The amount of energy that hits a square meter on the Earth's surface is maximum at the point where the incoming radiation is perpendicular to the Earth's surface (that is, where the Sun is directly overhead at midday). This point changes daily because the Earth's axis is tilted at 23.5° to the ecliptic. The most northerly point is reached on June 22 (the summer solstice), and the most southerly point is reached on December 22 (the winter solstice).

The Earth receives only a tiny fraction of this luminosity. Viewed from the Sun, the Earth is a disc with a radius of 6.4 X 10^6 m (Fig. 2.5). The surface area of this disc is 1.3 X 10^{14} m^2, and so the luminosity that hits the whole Earth is 1.3 x 10^{14} m^2 X 1370 watts/m^2 = 1.8 X 10^{17} watts. Thus, the Earth receives only one 2 billionth of the total solar output of energy! Although this fraction is tiny, it is sufficient to supply all the energy needed to drive the Earth's external processes and to keep the biosphere growing healthily.

SOURCE OF THE SUN'S ENERGY

Nuclear reactions inside stars involve the fusion of lightweight chemical elements, particularly hydrogen, to form heavier elements such as helium and carbon. The fusion process, which happens only at exceedingly high temperatures, converts matter to energy. The first person to have an inkling about how energy is stored in atoms was Albert Einstein (1879-1955), who, in 1905, showed that matter and energy are connected through the now famous equation $E = mc^2$, where E is energy, m is mass, and c is the speed of light in a vacuum. Nuclear fusion converts some of the mass of an atom into energy. Nuclear fusion has been achieved on the Earth, but only in an uncontrolled manner in hydrogen bombs. If it could be done so that energy is released in a controlled manner, society would have a nearly limitless energy supply because there is so much hydrogen available on Earth. Much research is therefore being done on nuclear fusion.

There are many possible fusion reactions, but the Sun and most other stars produce their energy by two of them: the proton-proton (PP) chain and the carbon-nitrogen-oxygen (CNO) chain. In both pro cesses the net result is the same—four hydrogen nuclei (protons) fuse while absorbing two electrons; the electrons combine with two of the protons to form two neutrons, and the result is a single helium nucleus containing two protons and two neutrons[1] as well as an enormous amount of energy. The difference between the PP and CNO chains is that in the PP chain the protons fuse directly to helium, whereas in the CNO chain the process has intermediate steps that involve carbon, nitrogen, and oxygen, as well as protons. In the Sun the PP chain accounts for about 88 percent of the energy produced, and the CNO chain the remaining 12 percent.

The chemical symbol for a hydrogen nucleus is ^1_1H; the superscript indicates the sum of the protons plus neutrons in the nucleus (in this case a sole proton), while the subscript indicates the number of protons, in this case also one. Helium is written ^4_2He, which indicates a total of four particles in the nucleus: two of them are protons, as we can read from the subscript, and so the other two must be neutrons.

The mass of an atom is expressed in terms of atomic mass units (AMU), where 1 AMU is one-twelfth

[1]As you probably recall from high school chemistry, atoms are made up of protons and neutrons, bunched together in the nucleus, and electrons that move in orbits around the nucleus. The stucture and properties of atoms are discussed more fully in Chapter 4.

A Closer Look

Electromagnetic Radiation

Whenever an electrically charged particle is accelerated, it radiates energy in the form of **electromagnetic radiation.** Light is the most familiar form of electromagnetic radiation, but X rays, y rays, infrared rays, and radio waves are also electromagnetic radiation; all these many forms differ only in wavelength. A group of electromagnetic rays arranged in order of increasing or decreasing wavelength is called a **spectrum.** The most familiar example is the *visible spectrum,* which is the range of wavelengths to which our eyes are sensitive.

Waves have three essential properties: the *wavelength X,* which is the distance between two successive crests (Fig. C2.1); the *speed v,* which is the distance traveled by a crest in one second; and the *frequency f,* which is the number of crests that pass a given point each second. The relation between the three wave properties is *fk = v.* All wavelengths of electromagnetic radiation travel with the speed of light, which in vacuum is 299,793 km/s (186,291 mi/s) and is usually designated c. For electromagnetic radiation, therefore, *fk = c.* Because all electromagnetic radiation has exactly the same speed, *c,* it is possible to refer to electromagnetic radiation either in terms of wavelength or in terms of frequency.

Wavelengths of electromagnetic waves are usually measured in meters (Fig. C2.2). The unit of frequency is the *hertz* (Hz). A frequency of 1 Hz is one wave crest passing a given point each second. Scientists prefer to use the term *cycle* rather than wave crest when referring to frequency. One cycle per second is just another way of saying one wave crest per second.

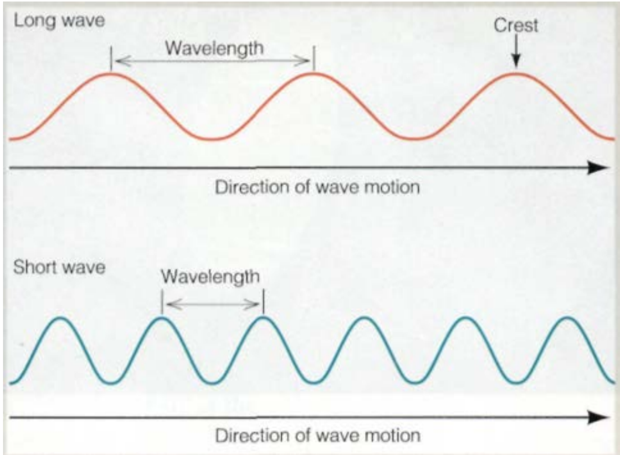


Figure C2.1 The properties of waves. The wavelength is the distance from one crest to the next. The wave frequency is the number of crests that pass a given point each second. If both a short wave and a long wave move at the same speed, the short wave has the higher frequency.

Electromagnetic radiation can be described equally well in terms of waves or in terms of packets of radiant energy called *quanta* or *photons.* Sometimes it is more convenient to deal with the wave properties of the radiation; at other times it is more convenient to deal with the packets of energy properties. The relationship between these two "forms" of electromagnetic radiation is *E = hf,*

of the mass of a carbon-12 atom ($^{12}_{6}C$). A proton has a mass of 1.00758 AMU and a neutron 1.00893 AMU. (A neutron is slightly more massive than a proton because it is formed when a proton combines with an electron.) The mass of one $^{4}_{2}He$ should therefore be (2 X 1.00758) + (2 X 1.00893) = 4.03302 AMU. When the mass of a helium nucleus is determined, however, it is found to be only 4.00260 AMU. Some of the mass has been lost, and it is this lost mass that is converted to energy according to Einstein's $E = mc^2$. The amount of energy released by the fusion of four $^{1}_{1}H$ to produce one $^{4}_{2}He$ is small, about 4.2 X 10^{-12}J, but 4.5 X 10^6 metric tons of hydrogen is converted to helium every second in the Sun. Thus, the total amount of energy released is enormous.

Because a huge amount of its hydrogen is continuously being converted to helium, the Sun will eventually run out of hydrogen fuel. Fortunately, the Sun's supplies of hydrogen are enormous, and scientists calculate that there is enough in the Sun's interior to keep the nuclear fusion reactor operating for another 4 to 5 billion years.

Proton-proton fusion requires a temperature of at least 8 X 10^6 K (1.4 X 10^7°F), and the CNO fusion chain requires a temperature of 15 X 10^6 K (2.7 X 10^7°F). Although we cannot see into the interior of the Sun where fusion is occurring, scientists are reasonably certain that, because the CNO chain is operating, the temperature at the center of the Sun is at least 15 X 10^6K.

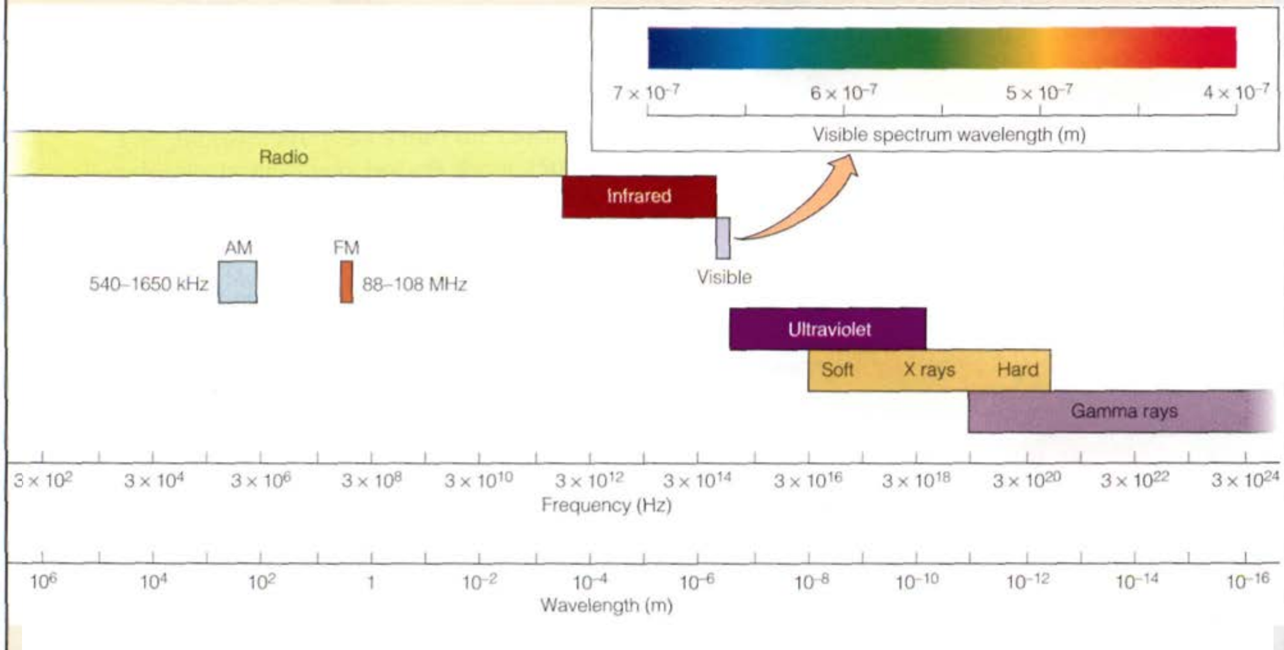
The energy produced by fusion reactions in the Sun appears as *gamma rays* (g rays), which are extremely short electromagnetic waves (see "A Closer Look: Electromagnetic Radiation"), and *neutrinos,* which are electrically neutral, essentially massless particles that move at the speed of light. Of the total where E is the energy of a photon, f is the frequency of the corresponding electromagnetic wave, and h is a constant known as Planck's constant. When f is measured in hertz, h is equal to 6.63×10^{-34} J·s, and E is in joules.

The energy of a photon corresponding to a 1,000-kHz wave, in the middle of the AM radio band, is

$$E=(6.63 \times 10^{-34}\ \text{J·s})\ (1.000 \times 10^{6}/\text{s})=6.63 \times 10^{-28}\ \text{J}$$

By contrast, the energy of a photon that has a frequency of 6×10^{14} Hz, which has a wavelength in the middle of the visible light range, is 3.98×10^{-19} J, and a photon in the g-ray range, which has a frequency of 10^{23} Hz, is 6.63×10^{-11} J. Obviously, the higher the frequency (and therefore the shorter the wavelength of the corresponding electromagnetic wave), the greater the amount of energy carried by a photon.

Figure C2.2 The electromagnetic spectrum. Because all electromagnetic waves travel with the same speed (the speed of light, 3.0×10^{8} m/s), they can be discussed either in terms of frequency or in terms of wavelength.



energy, 2 percent is in the form of neutrinos and 98 percent is g rays. Neutrinos escape from the Sun's core so easily that they escape about 2 seconds after they are formed. Neutrinos can pass, unchanged, through the Earth and do not play any part in bringing solar energy to the Earth. Gamma rays, however, cannot easily get free from the Sun, but as we will see later in the chapter, they are responsible for the energy that reaches the Earth.

STRUCTURE OF THE SUN

Figure 2.6 shows that the Sun consists of concentric layers, four inner regions that make up the sphere we see, as well as two gaseous outer layers that we cannot see.

The Sun's *core*, the site of all the nuclear fusion reactions, is about 170,000 km (106,000 mi) in radius. The temperature of the core ranges from 8×10^{6} K at the margin to 15×10^{6} K at the center. The composition of the core is estimated to be about 62 percent helium and 38 percent hydrogen by mass.

Surrounding the core is a region that is very hot but not hot enough for fusion to occur. Stretching from 170,000 to 590,000 km (106,000 to 367,000 mi), measuring out from the center, this region is called the *radiative layer*. The energy released in the core moves across the radiative layer by radiation, and it is this layer that makes the escape of energy from the Sun such a slow process. It is electrons in the radiative layer that absorb the *y* radiation and make the layer opaque.

Above the radiative layer is the *convective layer*, from 590,000 to 695,500 km (367,000 to 432,000 mi),

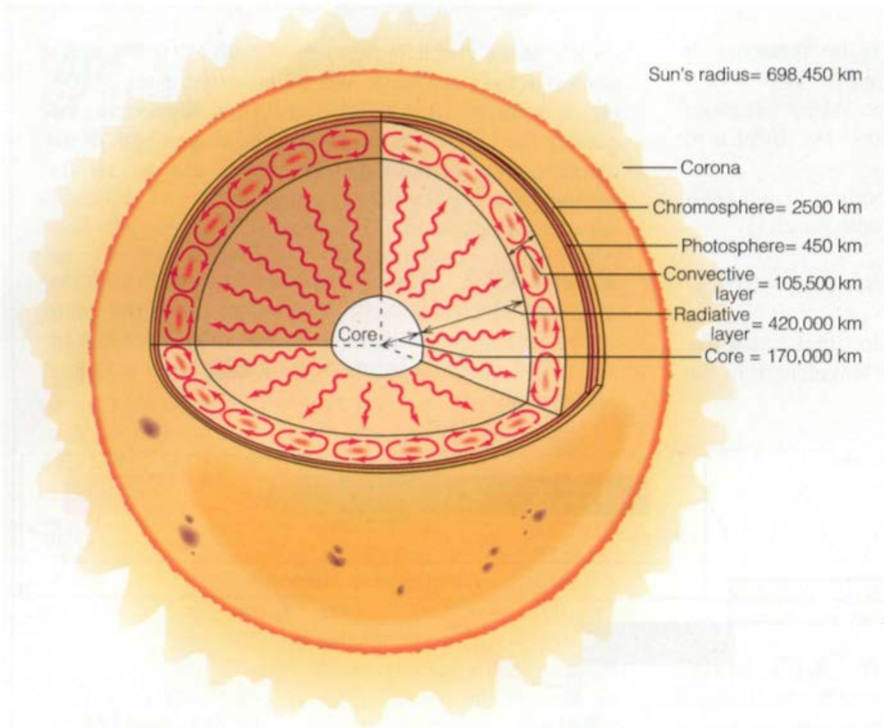


Figure 2.6 A model of the Sun's interior. Energy is created in the core when hydrogen is fused to helium. This energy flows out from the core by radiation through the radiative layer, by convection through the convective layer, and by radiation from the surface of the photosphere, which is the portion of the Sun we see.

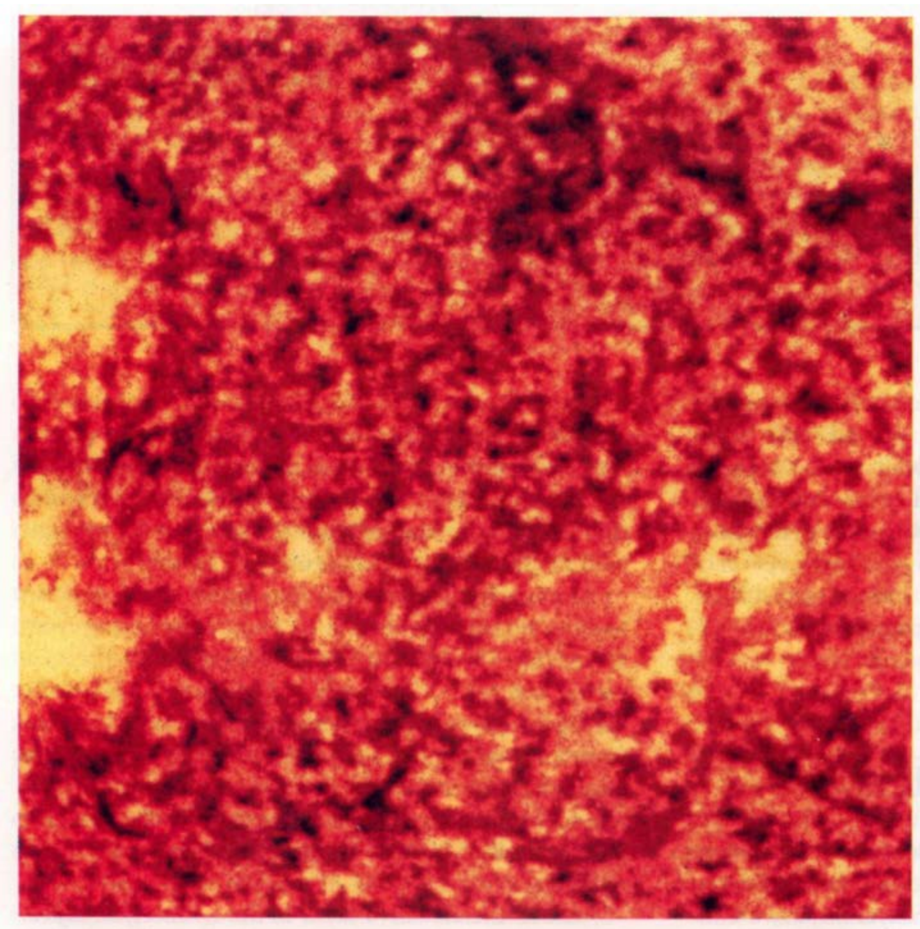


Figure 2.7 Granules up to 1500 km across on the surface of the Sun's photosphere. Granules are the tops of huge, upward-rising bubbles and columns of intensely hot gas. The darker regions between the bright granules are places where cooler gas flows back into the photosphere.

across which energy moves by convection. Both the radiative and the convective layer have compositions that are little changed from the composition of the original solar nebula and about the same as that of Jupiter, which is 72 percent hydrogen, 26 percent helium, and 2 percent heavier elements, by mass. In a sense, the radiative and convective layers are kept gassy and are prevented from collapsing into the core by the intense pressure created by *y* radiation attempting to move upward.

Above the convective layer is the surface layer, the portion of the Sun we see. Called the *photosphere*, it is an intensely turbulent zone that emits the light that reaches the Earth (Fig. 2.7). The photosphere is about 450 km (280 mi) thick and has an average temperature of about 5800 K (9980°F), ranging from 8000 K at its boundary with the convective layer, to 4000 K at its outer edge.

The photosphere passes into the *chromosphere*, a low-density layer of very hot gas about 2500 km (1554 mi) thick. The chromosphere has such a low density that it is transparent to light passing through and therefore very difficult to see.

The chromosphere merges into the outermost layer of the Sun, the *corona*, a zone of even lower density gas than the chromosphere. The corona, which grades off into space, is quite hot, but, like the chromosphere, it is not visible because it has such a low density. Because both the chromosphere and the corona are transparent, we are able to observe and measure them only during a solar eclipse, when light from the body of the Sun is obscured by the Moon (Fig. 2.8).

THE SOLAR SPECTRUM

The radiation energy released in the Sun's core by the PP and CNO fusion chains has a frequency of about 10^{23} Hz, in the g-ray range; such radiation has a very short wavelength and is extremely energetic. As the *y* rays move out through the radiative layer, they are repeatedly absorbed and re-emitted by electrons and in the process converted to longer wavelength, lower energy radiation. No energy is lost in the process; it is just parceled out into a greater number of less energetic rays. By the time the radiation reaches the photosphere, it has a frequency in the range from 10^{14} to 1.5×10^{15} Hz or, as more commonly designated, wavelengths in the range 3×10^{-6} m (3.3×10^{-6} yds) to 2×10^{-7} m (2.2×10^{-7} yds).

Note in Figure 2.9 that the energy flux from the Sun varies with the wavelength, the peak being close to the wavelength of yellow light. The shape of the

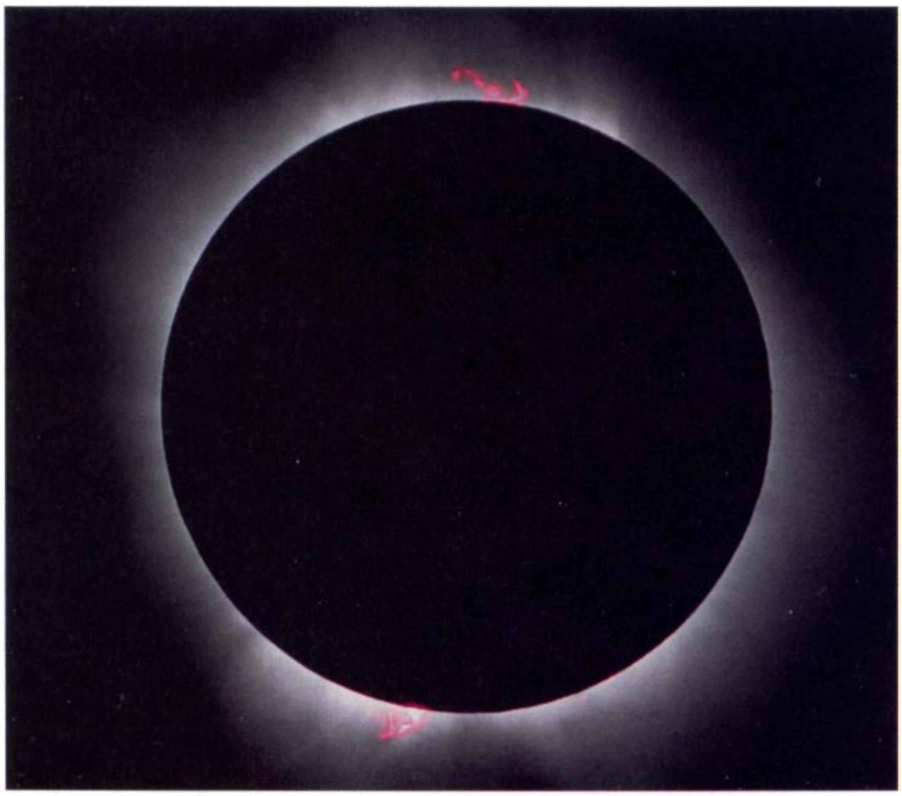


Figure 2.8 The Sun's chromosphere and corona can be clearly seen only during a solar eclipse, when the Moon blocks the light coming from the photosphere. This photo, taken with a special camera during an eclipse in 1988, shows faintly glowing gas streaming hundreds of thousands of kilometers out from the corona.

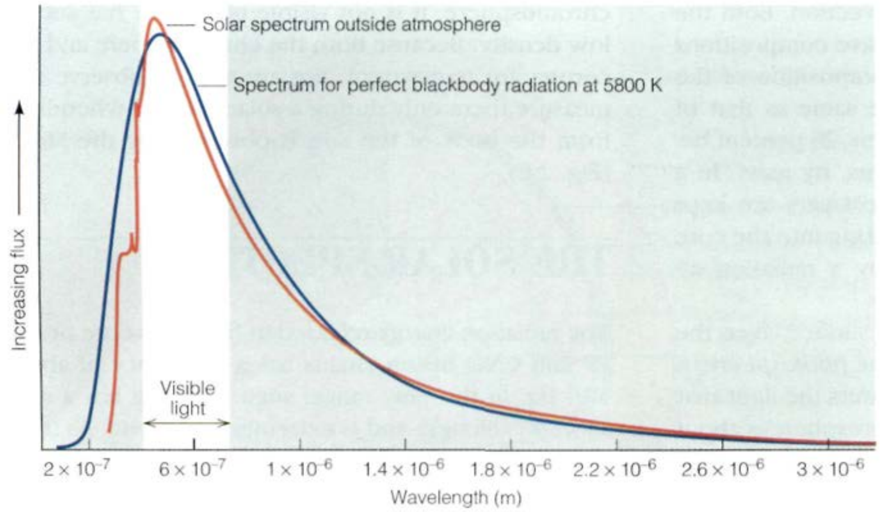


Figure 2.9 The Sun's spectrum is nearly identical to that of a perfect blackbody radiator. The minor differences occur because gases in the chromosphere and corona selectively absorb some wavelengths of the electromagnetic radiation emitted by the Sun.

Sun's spectral curve (which is a graph of wavelength versus flux) is interesting because it matches almost exactly the spectrum of radiation given off by a perfect blackbody heated up to 5800 K (9980°F) the average temperature of the photosphere.

When a piece of metal is heated in an intensely hot flame, the metal first starts to glow a dull red. Then, as it gets hotter and hotter, it becomes more brightly red, then orange, yellow, white, and finally bluish white. The color of the metal at any given time while it is being heated is a measure of the temperature, and this relationship has many practical uses. Blacksmiths, for example, use color to estimate when the temperature of a piece of iron is high enough for the task in hand.

If we measured the electromagnetic radiation emitted by the hot metal, we would find that the spectral curve has the same shape as the curves in Figure 2.9. Any body, no matter what its composition, that has a spectral curve similar to Figure 2.9 is called a **blackbody radiator.** The term *blackbody* seems confusing when applied to something that is glowing brightly and emitting electromagnetic radiation. In fact, the term refers to the *radiation-absorbing* properties of a body, and a perfect blackbody is one that absorbs all light that strikes it and reflects none. If you shone a very powerful beam of light at the Sun, almost none would be reflected back, making the Sun a nearly perfect blackbody. The Earth is also very close to being a perfect blackbody radiator. (Note that the two terms *blackbody* and *blackbody radiator* mean the same thing and are interchangeable.)

There are several important points to remember about blackbody radiators:

1. The hotter the radiating body, the shorter the wavelength of the radiation peak (Fig. 2.10). The peak of radiation for the Sun is at a wavelength of about 5×10^{-7} m (5.5×10^{-7} yds), corresponding to a temperature of 5800 K.
2. All objects, no matter what their temperature, emit electromagnetic radiation and have a spectral curve approximating, at least to some degree, the curve for a blackbody radiation. The Earth, for instance, has an average surface temperature of about 290 K (63° F), and the radiation it emits has a peak at a wavelength of 1×10^{-5} m (1.1×10^{-5} yds), much longer than the Sun's peak wavelength of 5×10^{7} m (5.5×10^{-7} yds). The Earth's peak wavelength is in the infrared region of the electromagnetic spectrum, and so the Earth's radiation cannot be seen by the human eye.
3. The hotter an object is, the more energy it radiates. We know this is true from such simple observations as the amount of energy given off by a hot stove coil versus a cold coil. The same relationship is apparent in Figure 2.10: the higher the temperature of the blackbody, the greater the energy flux at any given wavelength.

The solar spectrum in Figure 2.9 is the spectrum measured in space, far above the Earth's atmosphere. The spectrum measured at sea level is a lot different because the electromagnetic radiation has had to pass through the atmosphere, which selectively absorbs certain wavelengths. Gases in the atmosphere—principally oxygen, water vapor, and carbon dioxide—absorb radiation selectively over narrow wavelength ranges. The curve in Figure 2.11 shows how absorption by the atmosphere modifies the Sun's blackbody spectrum.

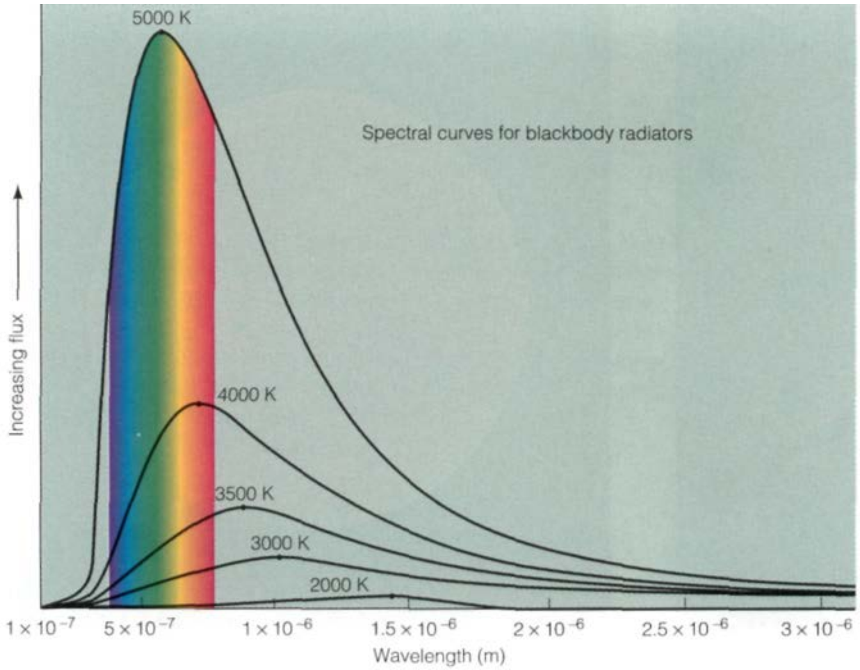


Figure 2.10 The energy flux from blackbody radiators at different temperatures. Note how the radiation peak moves to shorter wavelengths as the temperature increases. The area under any one curve is the total flux of energy emitted by a radiator at a given temperature. Note that the higher the temperature, the greater the flux.

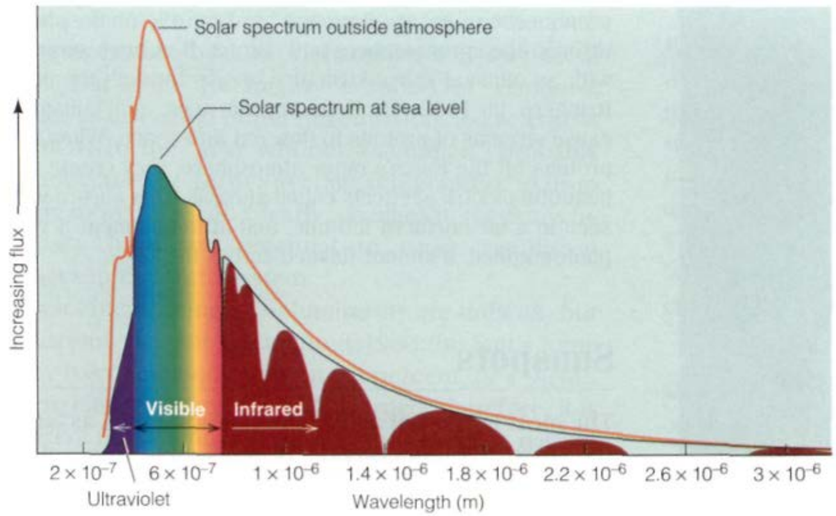


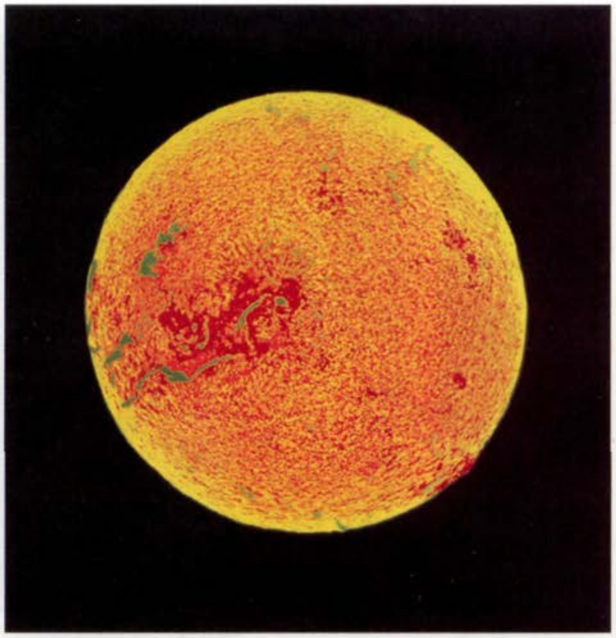
Figure 2.11 The outer-space and sea-level spectra of solar radiation. The two curves are different because gases in the atmosphere selectively absorb some of the wavelengths of emitted radiation.

THE ACTIVE SUN

So far, we have been discussing the Sun as if it were a reliable, smoothly operating body. Most of the time the Sun does work smoothly, and astronomers tend to refer to it under such conditions as a quiet Sun. At other times, however, and generally for short periods, the Sun becomes intensely turbulent and erupts with huge fiery prominences and vast sunspots (Fig. 2.12). Such times are referred to as periods of the active Sun, and the places where the events occur are called *active regions.*

The development of active regions apparently has two principal causes. Differential rotation, the first cause, arises because the Sun is a rotating ball of gas and rotates faster at the equator (once every 25 days) than at its poles (once every 31 days). Differential rotation is ever-present in the Sun, but much of the time the turbulence that results is unseen because it occurs below the surface of the photosphere. The second, and by far the more important cause, is magnetism.

A.



B.

C.

Figure 2.12 A period of the active Sun. A. A vast, fiery prominence of hot gas bursting out from the photosphere throughthe chromosphere and corona. B. A huge sunspot with an unusual spiral structure breaks through the photosphere in 1982. C. Both prominences and sunspots cause streams of protons to flow out into space. When the protons hit the Earth's outer atmosphere, they create the beautiful electrical effects called auroras. This aurora was seen in a far northern latitude, and at the moment it was photographed, a meteor flashed across the sky.

The gases in the Sun consist of electrically charged particles. The vast flows of gas inside the Sun are, in effect, electric currents, and electric currents create magnetic fields. The Sun is therefore a magnet. Its magnetic field can be distorted by the differential rotation or by turbulence inside the Sun, and this distortion can disrupt the flowing gases. When the internal turbulence breaks through to the surface, a period of an active Sun follows.

Sunspots

The most important active Sun phenomenon, as far as the Earth is concerned, is sunspots—huge dark blotches on the solar surface. Sunspots are relatively cool regions on the surface of the photosphere, and so they appear dark by comparison with the rest of the photosphere. (Even so, they are intensely hot.) When a sunspot starts to form, the granular surface of the photosphere separates, and a tiny dark spot that is intensely magnetic appears and starts to grow. Exactly how and why sunspots form is not clearly understood, but the fact that they seem to occur in cycles of about 11 years (Fig. 2.13), which is the period calculated by astronomers for turbulent interactions between the solar magnetic field and differential rotation, suggests they may be a normal part of the Sun's activities. Even more important, because the Sun's magnetic field influences the Earth's outer atmosphere, many experts believe that sunspots influence the climate on the Earth.

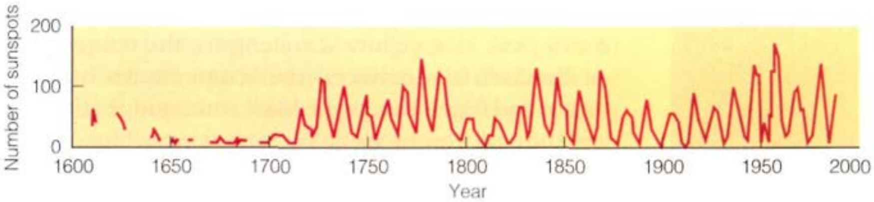


Figure 2.13 The sunspot cycle over the past 400 years. Note the period before 1715, when, for reasons that are not understood, very few sunspots were observed. Sunspots have reached a maximum about every 11 years since 1715, and there is also a suggestion of some sort of cycle on a 55- to 57-year time scale. Because the pre-1715 period of low sunspot activity coincides with a prolonged cool period that is sometimes called the Little Ice Age, some scientists have speculated that sunspot activity and climate are somehow connected.

Changes in Luminosity

Astronomers monitor the Sun's luminosity very carefully. A 1 percent decrease in the flux of electromagnetic radiation, from 1370 W/m^2 to 1356 W/m^2, is estimated to reduce the Earth's average temperature by 1 K (1.8°F). Similarly, a 1 percent increase in luminosity to 1384 W/m^2 would probably increase the average temperature by 1 K. A change of 1 K may seem small, but even the mighty eruption of Tamboro, which, as discussed in the Introduction, led scientists to call 1816 the "year without a summer," probably did not cause the Earth's average temperature to drop by any more than 1 K. Clearly, changes in the Sun's luminosity have the potential to cause significant changes in the Earth system.

Exact measurements of luminosity are difficult, but indications are that since about 1980 the Sun's luminosity has decreased by about 0.3 percent, or 4 W/m^2. Where climate changes are concerned, therefore, it is apparent that changes in luminosity, as well as natural and human-engendered changes to the atmosphere, must be considered.

The Sun's luminosity changes have been monitored only during the twentieth century; thus, questions of long-term changes remain unanswered. As we will see later in this chapter, a star the size of the Sun commences life with a luminosity about 10 percent lower than the present solar luminosity. This means one of two things: either the early Earth was much colder than today's Earth, or else the Earth system of that time adjusted in some way in order to keep the surface warm, perhaps by allowing the carbon dioxide level of the atmosphere to be higher in order to provide a better thermal blanket. In the carbon dioxide model, it is hypothesized that, as the sun's luminosity slowly increased, the biosphere removed carbon dioxide from the atmosphere, thus keeping a balance between incoming radiation and a comfortable climate.

OTHER SUNS

Because the stars are so far away from the Earth[2], it is not possible to measure how big they are or to see all the detail we can see on the Sun. Almost everything we know about stars comes by way of the electromagnetic radiation they emit. The way we decode and interpret the messages carried by starlight depends to a large degree on our understanding of how the Sun works. Fortunately, a lot of information can be gathered from some straightforward measurements.

Star Color and Luminosity

When you look at the stars on a clear night, two things are quickly apparent. The first observation is that the colors are not all the same; they range from red through yellow to bluish-white. For example, in Orion, one of the most familiar and easily recognized constellations, we see a bright but distinctly red star called Betelgeuse and an equally bright star called Rigel that is a striking bluish-white (Fig. 2.14).

The second observation is that stars vary greatly in their brightness. Some, like Sirius and Rigel, blaze out

[2]Astronomical distances are measured by the time it takes for light to travel between the two points being measured; 1 light-minute is 18 X 10^9m, and the Sun-Earth distance is only 8.3 light-minutes. The star nearest to the Earth, Alpha Centauri, is 4 light-years away (1 ly = 9.5 X 10^{15}m). When we look through a telescope at stars a billion light-years away, we are seeing light that left those stars a billion years ago; we are looking back in time.

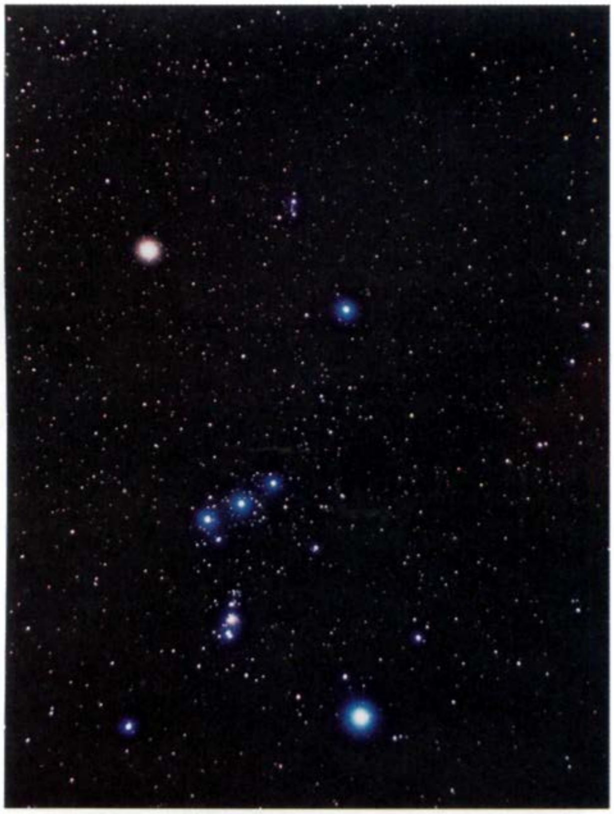


Figure 2.14 The constellation Orion (the hunter). The reddish star at the upper left is Betelgeuse; the bright, bluish-white star at the lower right is Rigel. Because this photo is a time exposure, many faint stars not visible to the eye are shown.

quickly and draw attention because they are so bright. If you look closely, however, you can find stars so faint you can hardly be sure they are there.

Like the Sun, all other stars are blackbody radiators. As shown in Figure 2.10, the peak wavelength of a blackbody radiator is determined by temperature. The higher the temperature, the shorter the wavelength and the bluer the star. Conversely, the lower the temperature, the longer the wavelength and the redder the star. Rigel, with its peak at the blue end of the spectrum, is hotter than red Betelgeuse. Because the Sun's peak is in yellow wavelengths, the temperature of the Sun falls between the temperatures of Betelgeuse and Rigel. Just as the blacksmith judges the temperature of iron by its color, so it is possible to judge the temperature of a star by its color.

Astronomers classify stars based on color and hence temperature (Table 2.1 and Fig. 2.15). Each color—or as it is more commonly called, spectral class—is further split into 10 subdivisions ranging

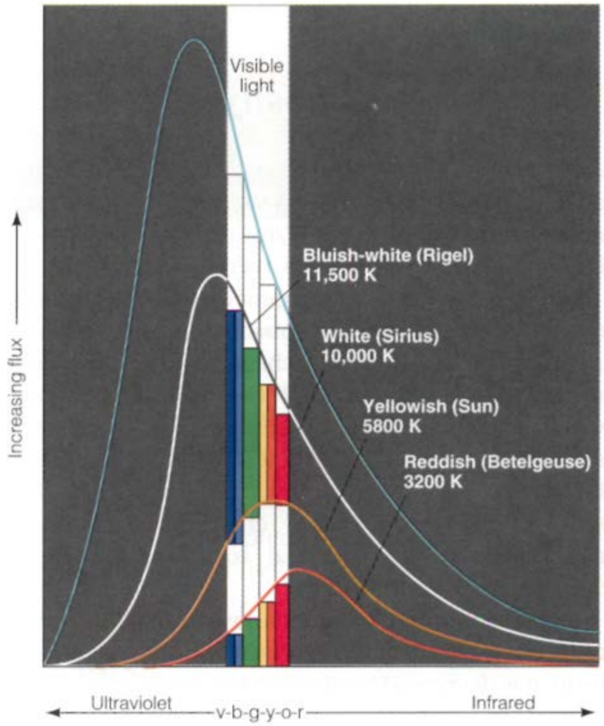


Figure 2.15 Stars as blackbody radiators. Note that only for yellowish stars is the radiation peak in the visible range. For reddish, white, and bluish-white stars, the radiation peaks lie outside the visible range. We can see such stars because they do emit some radiation in the visible range.

Table 2.1
The Spectral Classes of Stars

Spectral Class	Color	Surface Temperature	Example
O	Bluish-white	Greater than 30,000K	Naos
B	Bluish-white	11,000-30,000K	Rigel
A	Bluish-white	7,500-11,000K	Sirius
F	White to bluish-white	6,000-7,500K	Canopus
G	White to yellowish-white	5,000-6,000K	Sun
K	Yellowish-orange	3,500-5,000K	Aldebaran
M	Reddish	Less than 3,500K	Betelgeuse

from 0 (hottest) to 9 (coolest). The Sun is a G2 star.

In order to measure a star's luminosity, we need to know the Earth-star distance. Earth-star distances are difficult to measure, but they can be determined for stars out to a distance of 300 light-years from the Earth using a system called *parallax.* You can demonstrate parallax very easily by holding a pencil perpendicular to the floor and at arm's length, and alternately opening and closing each eye. The pencil, viewed against a fixed background, appears to move from side to side (Fig. 2.16A). A parallax measurement of the distance to a nearby star uses the Earth at two opposite points on its orbit for the two "eyes" and very distant stars as the fixed background. The method is illustrated in Figure 2.16B. The diameter of the Earth's orbit is known; the angle subtended by the star to the two points of observation is measured, and from the data it is possible to calculate the distance to the star. For very distant stars, where the angles are tiny, measurements are imprecise. That is why, at present, only stars out to 300 light-years can be measured with any degree of accuracy.

Once the distance to a star is known, its luminosity can be calculated. Remember that luminosity is the total amount of energy emitted by a star each second, and to get an accurate measurement we must know how far away a star is. The process is the same as measuring the luminosity of the Sun as discussed earlier. First, the energy flux of the star is measured through a telescope (see "A Closer Look: Telescopes"). Then the surface area of a sphere with a radius equal to the Earth-star distance is calculated just as we did for the Sun in Figure 2.4. Since flux is energy/second/unit area, the total energy emitted by a star each second is easily calculated.

Once the temperature and luminosity of a star are known, they can be compared with the values for other stars. One convenient way to make a comparison is through the **Hertzsprung-Russell diagram (H-R)**, a plot of luminosity versus temperature.

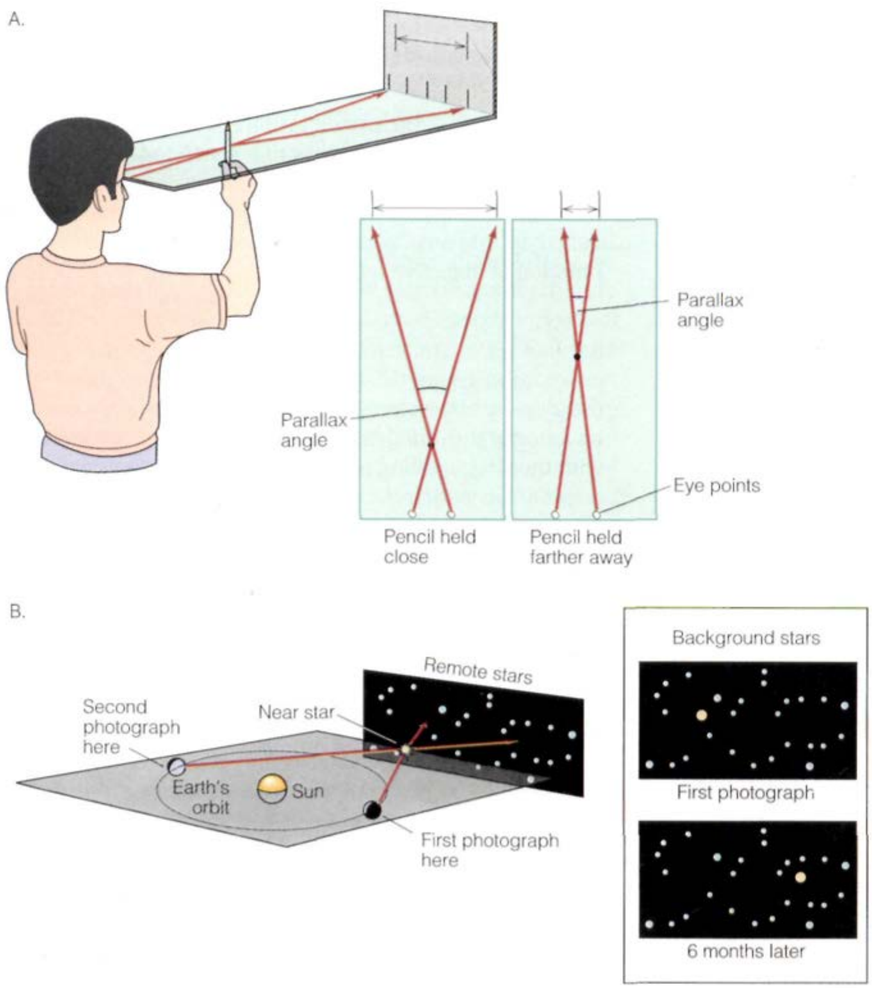


Figure 2.16 Parallax used to measure star distances. A. An example of parallax. Alternately shut your left and right eyes, and the pencil will appear to move relative to a fixed background. B. As the Earth goes around the Sun, a near star appears to move against the background of more distant stars. By observing the near star at six month intervals, knowing that the base of the green triangle is the diameter of the Earth's orbit, and measuring the angle of the shift, the distance between the near star and the Earth can be calculated.

A Closer Look

Telescopes

Astronomy is an observational science. The only way hypotheses concerning stars can be tested is through observation, and because stars are so faint and so distant, the observations have to be made with telescopes.

All telescopes have one function: they gather and concentrate electromagnetic radiation. Those that gather visible light are called *optical telescopes,* and those that gather radio waves are *radio telescopes; infrared* and *ultraviolet telescopes* gather infrared and ultraviolet waves, respectively.

Optical Telescopes

Optical telescopes use either of two properties to gather and concentrate light. The first property is **refraction,** which means the path of a beam of light is bent when the beam crosses from one transparent material to another (Fig. C2.3A). Refraction occurs because the speed of light is different in different media. Remember that *all* electromagnetic radiation travels with the same speed in a vacuum. The speed of light is less in water, glass, or any other transparent medium than in a vacuum. Consider what happens when a beam of light is traveling through air at some initial speed and then hits a glass prism, as in Figure C2.3B. The part of the beam that hits the glass first slows down, but the rest of the beam is still traveling at the initial fast speed and therefore catches up with the slow-moving part. The net effect is to rotate the direction in which the entire beam moves through the glass.

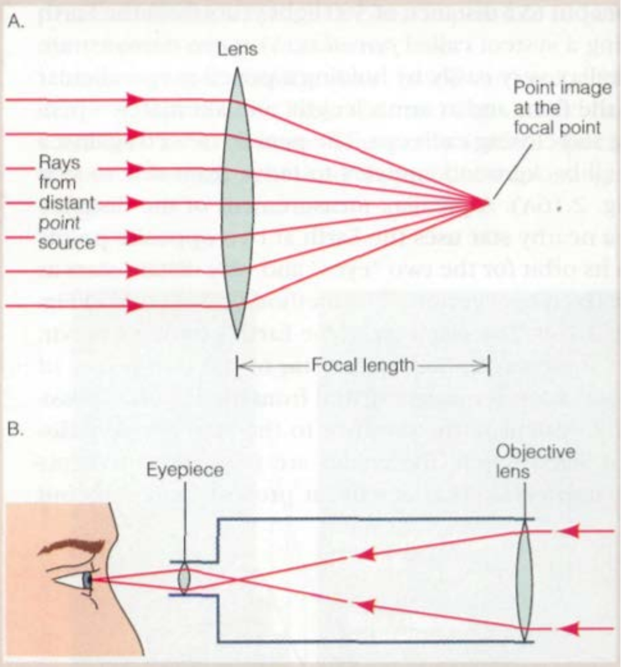


Figure C2.4 The principle of a refracting telescope. A. A simple lens refracts rays to a sharp focal point. B. A telescope gathers light through the *objective* lens and focuses the image at the focal point. An additional lens, called an *eyepiece,* is used to view the image. The combination of objective lens and eyepiece is the telescope.

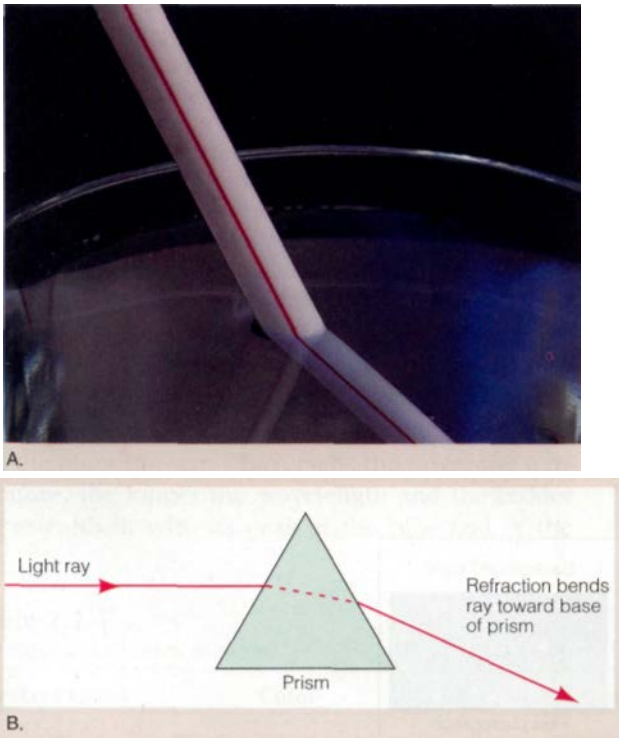


Figure C2.3 Examples of refraction. A. Refraction of light causes a drinking straw to appear to be bent at the air-water boundary. B. Refraction of light through a prism.

The way refraction is used to construct a *refracting telescope,* or *refractor,* is shown in. Figure C2.4. The first telescopes made were refractors, and it was with a simple refractor that Galileo made his epochal observations of the moons around Jupiter.

The second property of light used in telescopes is **reflection,** which means light bounces off a surface. When the reflecting surface is curved, all the reflected light beams can be brought to a focus, as shown in Figure C2.5A. By using either of two geometries, the focused beams can be viewed with an eyepiece, thus creating a *reflecting telescope* (Fig. C2.5B). The telescope Newton made to observe the planets was a reflector.

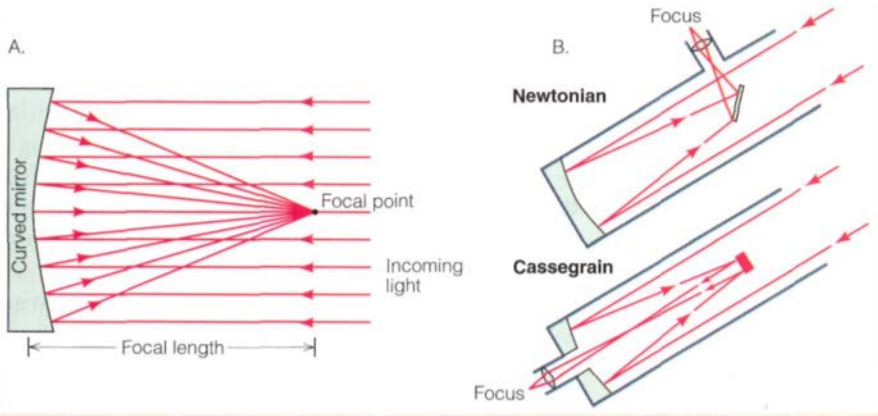


Figure C2.5 The principle of a reflecting telescope. A. A curved mirror reflects incoming light rays to a focal point. B. A reflecting telescope requires an eyepiece to examine the image at the focal point. Two geometries, Newtonian and Cassegrain, are used to prevent the head of the observer from blocking incoming light rays.

Nonoptical Telescopes

Galileo, Newton, and many generations of later astronomers had only their eyes to look through telescopes and see the stars and planets. Today's astronomers rarely look through telescopes. Instead they substitute other light-sensitive devices for their eyes. Such devices, which can be photographic films, solid-state photoelectric chips like those in solar-powered calculators, or photomultiplier tubes, are more sensitive than eyes and have an additional advantage: they can measure the *amount* of incoming electromagnetic radiation. Such measurement is essential for the determination of luminosity.

Unique information about stars can be obtained at almost every wavelength in the electromagnetic spectrum. Astronomers therefore seek to "look at" the sky at many different wavelengths, not only those in the visible region. Electromagnetic radiation that penetrates the atmosphere can be studied with *ground-based,* nonoptical telescopes. In the case of *radiotelescopes,* which operate in a wavelength range of about 10^{-6} to 10^{2} m, the "mirror" is a large metal dish and the detector is a radio receiver placed at the focal point (Fig. C2.6). Near-infrared wavelengths, from about 10^{-6} to 10^{-3} m, are gathered and focused in an *infrared* telescope using a mirror, just as in an optical telescope, but the detector is a chip of metallic germanium, which is sensitive to infrared rays.

As is clear from Figure 2.11, some wavelengths are absorbed by the atmosphere. To "see" the sky at those wavelengths, it is necessary to put *space* telescopes outside the atmosphere. Wavelengths in the ultraviolet, from 10^{-10} to 10^{-7} m, and in the far infrared, beyond about 4 X 10^{-5} m (4.4 X 10^{-5} yds), are partly absorbed by the atmosphere. Telescopes operating in these wavelength ranges are either lifted aloft by huge balloons or placed in orbit by rockets.

The best possible place to view the sky is completely outside the atmosphere. Recently, a large optical telescope, called the Hubble Telescope, with a mirror 2.4 m (2.6 yds) in diameter, was placed in orbit 400 km (249 mi) above the Earth. Despite some initial mechanical problems which astronauts were able to correct, Hubble has been an extraordinary success. Indeed, it has been such a success that plans are now being drawn up to place a telescope on the far side of the Moon, where it will be free not only from the atmosphere but also from all the electromagnetic radiation from the Earth. Space astronomy will probably become a standard way of operating in the twenty-first century.

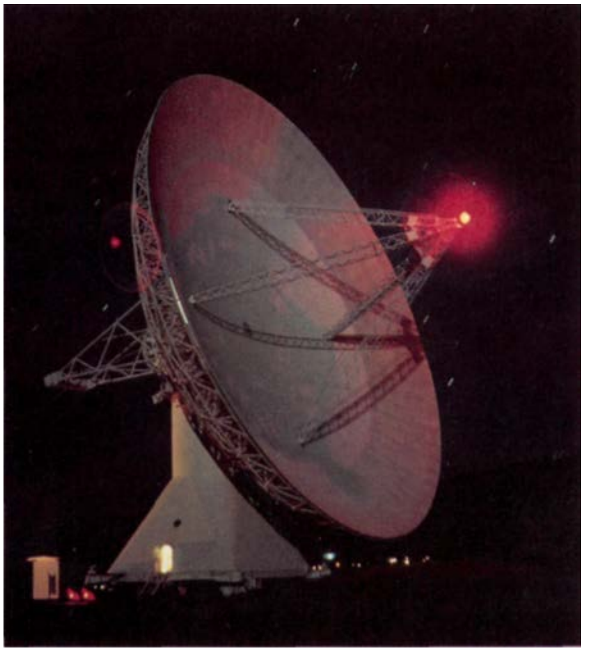


Figure C2.6 A radiotelescope in Owens Valley, Cailfornia, operated by astronomers of the California Institute of Technology. Radio waves from space hit the curved dish and are focused on the reciever mounted on the top of the four legs. The reciever, which works like a radio, can be turned to recieve different wavelengths.

Hertzsprung-Russell Diagrams

The H-R diagram was devised early in the twentieth century by two astronomers, Ejnar Hertzsprung (1873-1967), a Dane, and Henry N. Russell (1877-1957), an American, with the two men working independently of each other. Although the spectral class, and hence the temperature, of a star can be measured no matter how far away the star is, it is possible to measure luminosity only at known distances. The H-R plot was therefore developed from stars in our galaxy that are near enough for the distance to be measured by parallax.

About 85 percent of the stars plotted on an H-R diagram fall on or close to a smooth curve (Fig. 2.17) that astronomers refer to as the **main sequence.** The rest of the stars fall in two groups—one above and the other below the main sequence.

Luminosity is a function of both star temperature and star size. At a given temperature, the larger the star, the greater the luminosity. Stars that plot above the main seqvience are all very luminous. However, based on color, astronomers know that these stars are cooler than main-sequence stars of equal luminosity. For example, the red star Antares in Figure 2.17 has the same luminosity as the blue star Spica but is cooler. We can draw only one conclusion: in order to have such high luminosities at lower temperatures, the stars above the main sequence must be very large; they are giants. About 2 to 3 percent of all stars are giants. Depending on their size, they may be referred to as **red giants** or, in the case of the very largest stars, such as Betelgeuse, super giants. Betelgeuse is so large that if it were the Sun, its edge would be beyond the edge of Mars and the Earth would be inside a huge star.

Below the main sequence is a small group of stars that are much less luminous and therefore much smaller than the main sequence stars. Although not all the stars in this group fall in the white color class (class A), they have come to be called **white dwarfs.**

Astronomers discoverd the importance of the H-R diagram a long time ago: it can be used to explain the history of a star. All stars have lives; they are born, they age, and they die. Their life cycle has much to do with their size at any given time.

Stellar Evolution

Stars have long lives, and the smaller a star, the longer it can live. Very massive stars live for only tens of millions of years, intermediate-mass stars like the Sun live up to 10 billion years, and the smallest stars can live 20 billion years or longer. The reasons for the differences lie in the balance of forces inside a star.

As we have previously pointed out, all stars are huge, hot balls of gas. The hot gas does not drift off into space because the force of gravity continuously pulls it *inward.* The greater the mass of gas in a star,

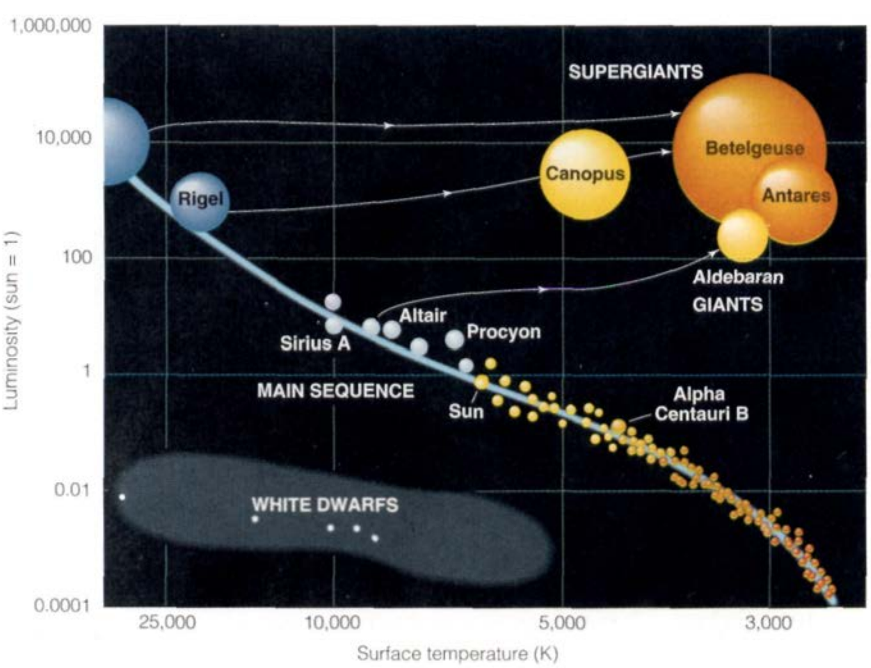


Figure 2.17 A Hertzsprung-Russell diagram of star luminosity versus surface temperatures. The vertical axis is a comparative one based on the Sun having a luminosity of 1. The horizontal axis is reversed from the normal order, with values of surface temperature increasing to the left. Note that the Sun is a middle-range, main-sequence star.

the stronger the inward pull. In the absence of a counterbalancing *outward* force, gravity would cause a star to collapse into a small, dense mass. The principal force that counterbalances gravity is electromagnetic radiation. The outward flux of electromagnetic radiation generated by nuclear fusion keeps the interior of a star very hot, and that prevents gravitational collapse; the balance between the inward gravitational force and the outward radiation force determines the size of the star. A fusion reactor requires fuel, and just as an automobile no longer runs when its gasoline supply is used up, so a fusion reactor can no longer operate when its nuclear fuel is depleted. When its fuel supply runs low, the balance of forces in a star changes, and as a result the size, temperature, and luminosity all change too. When and how the changes occur are a function of how much fuel was present to start with and how fast it was used up.

Because star lifetimes are vastly longer than human lifetimes, it is reasonable to ask how we humans can ever decipher star histories. The answer is provided by the Principle of Uniformitarianism. Among the billions of stars in the sky, astronomers can find examples of every mass, every luminosity, and every stage of star life. Understanding how a star ages is like deciphering the stages of a human life by studying the population of a large town for a few weeks. The result is a composite, an *average* life, rather than the life of a single individual. In the case of stars, the problem is made a little more complex because star masses differ by a factor of a thousand and the life of a small star differs from the life of a large star.

The mass of the Sun (one solar mass, 1 S) is used as a measure of star masses. Masses smaller than about 0.1 S are too small for the temperature in the core to get hot enough for nuclear fusion to start, and so below 0.1 S there are what we call "almost stars." The planet Jupiter is an example of an almost star. At the other end of the size range, stars more massive than 100 S generate such intense radiation that gravity is overcome by the outward radiation forces and the star simply blows itself apart.

The Life Cycle of a 1-S Star

The life of a star the size of the Sun begins with the gravitational compression of a mass of gas. Just as the air in a bicycle pump becomes heated as a result of compression, so does star gas heat up as a result of gravitational compression. When the temperature at the center of a compressed protostar reaches 8×10^6 K (1.4×10^{-7} °F), PP fusion commences.

Initially, gravitational compression exceeds the radiation counterforce, and a newly burning star continues to shrink. Within about 50 million years, a balance is reached between the gravitational and radiation forces, and the star attains a stable luminosity and temperature that place it on the main sequence of the H-R plot. The evidence that stars spend most of their lives on the main sequence is straightforward—most of the stars in the sky plot on the main sequence. Stars appear to be born and to die at about the same rate, yet 85 percent of all stars plot on the main sequence. That can happen only if stars have, through most of their lives, a temperature and luminosity that plot on the main sequence.

The nuclear fusion reactor of a 1-S star converts hydrogen to helium. When the hydrogen in the core is used up, fusion in the core ceases, gravity asserts control, and the now helium-rich core contracts. However, there is still abundant hydrogen surrounding the core in the radiative layer. As the core collapses and becomes even hotter, therefore, a shell of hydrogen in the inner part of the radiative layer starts the nuclear fusion process in what is called **shell fusion.** Core collapse heats the star interior by compression; increasing temperature speeds up the rate of nuclear fusion in the radiative-layer shell, and as a consequence such an immense amount of heat is generated that the star expands. Such a shell-fusion star moves off the main sequence and becomes a red giant.

Despite the fact that a red giant is very large, its helium-rich core continues to contract even after shell fusion commences. Eventually, the core temperature is hot enough for helium fusion to start by a process called the triple-alpha reaction, in which three helium nuclei fuse to form carbon. When all the helium fuel in the core is used up, shell fusion in the radiative layer starts again, but this time it is the fusion of helium. Inside the shell, the now carbon-rich core continues to contract gravitationally. Shell fusion of helium is a violent process accompanied by great explosions, and each explosion blasts star gas out into space. Finally, all that remains of a 1-S star is the carbon-rich core surrounded by a slowly diminishing shell of helium; when shell fusion of helium ceases, the star starts to die.

The mass of the carbon-rich core of a 1-S star is too small for contraction to raise the temperature to the point at which carbon can fuse to heavier elements. The carbon-rich core becomes a white dwarf, a small, very dense star that is slowly cooling. On the H-R plot a white dwarf plots below the main sequence. As cooling proceeds, a white dwarf loses its luminosity, moves toward the lower right-hand corner of the H-R plot, and eventually becomes a dead star called a *black dwarf.* The star's life cycle is now complete.

The scenario just described is approximately that which our Sun will follow. There is nothing for hu-

Profile

MARIA MITCHELL (1818-1889)

Off the coast of Massachusetts, on the island of Nantucket, is a modest brick observatory founded in memory of the United States' first woman astronomer, Maria Mitchell. (Maria is pronounced Ma-rT-ah.) As an undergraduate, I joined a group of students who spent a summer at Nantucket observing, photographing, and analyzing variable stars in the nucleus of the Milky Way. To explore our galaxy's magnificent star fields, we pointed the telescope toward the constellation of Sagittarius on clear, dark, moonless nights. We used Maria Mitchell's own 5-inch Clark refractor to locate a region rich in variable stars. Then we swung open the shutter of the photographic telescope to capture the stellar images on glass plates. Toward dawn, we processed the astronomical photographs in a tiny closet that had been outfitted as a darkroom. The cubicle adjoined the historic gray-shingled house where Maria Mitchell was born. During the day we often ventured into her home to learn about her childhood and career. And at night, as we photographed the stars, we kept each other awake by recounting the many stories we had heard about this famous woman astronomer.

From a window in the front parlor of the Mitchell House, Maria helped her father observe an annular eclipse of the sun in 1831. She was only 12 years old at the time but already skilled in making accurate observations. To escape the pandemonium of her younger srblings, she converted a small closet with a window into a study where she could read, write, compute, and analyze.

It may have been George Bond who influenced Maria to set up her father's little 21/2-inch refractor and scan the sky for faint new comets. The two young people were very much attracted to each other and competed to be the first American to make such a discovery. Although they each independently sighted several new comets, European astronomers invariably received recognition for these discoveries. Finally, the king of Denmark presented Maria with a gold medal for discovering a fuzzy patch of light just above Polaris on October 1, 1847, and being the first to report it as a new telescopic comet.

At a time when most women married and had families, Maria also aspired to wed a bright young man. She was devastated when George Bond married a young

Barbra C. Welther is a historian of astronomy at the Harvard-Smithsonian Center for Astrophysics in Cambridge Massachusetts. Currently, she is writing a biography of Annie Jump Cannon, a pioneer woman astronomer whose career she celebrated in an educational documentary video. Welther received the Annie Jump Cannon Medal from the Wesley College in Dover, Delaware, and had an asteroid, Minor Planet (3682) Welther, named in her honor.

mans to fear, however, because the hydrogen fuel in the Sun's core is sufficient to keep the Sun on the main sequence for at least 4 billion more years.

The Life-Cycle of a 0.25-S Star

It is not possible to know the full life cycle of a small-mass star. Such stars are so long-lived, 20 billion years or more, that there hasn't been enough time for them to evolve off the main sequence. The universe itself is only 15 billion years old!

A small-mass star is born in the same way as a 1-S star—by gravitational contraction and the onset of the PP cycle. However, gravity in a small-mass star is so weak that the core where nuclear fusion commences is small. As a consequence, the small amount of electromagnetic radiation released counteracts the gravitational contraction force. Such stars have low luminosities and low temperatures and plot on the lower right-hand end of the main sequence. The hydrogen fuel in a small-mass star is used up so slowly that it lasts an incredibly long time. The fuel will eventually be depleted, however, and shell fusion, core contraction, and helium fusion will follow. Beyond that stage, however, observation tells us nothing about what will happen in the future.

The Life-Cycle of a 5-8 Star

Details in the life of a very massive star differ from those of a 1-S star. The first difference is that the initial gravitational contraction of a massive star is so intense

Bostonian. Mitchell's talent and fame, however, allowed her to pursue a career in astronomy. In 1848 the fellows of the American Academy of Arts and Sciences elected her an honorary member; she had recognition but no voting privileges. The next year she accepted a stipend to compute daily positions of Venus for the newly created *American Ephemeris and Nautical Almanac.* In the 1850s Mitchell was elected to membership in the American Association for the Advancement of Society and was also honored by a group of American women, who presented her with enough money to buy the fine Alvan Clark telescope still in use today on Nantucket today. With that instrument, she observed the colors of three dozen pairs of double stars and published her results in the July 1863 issue of the *American Journal of Science and Arts.* Meanwhile, Vassar College, a women's school opened in 1865, invited Mitchell to become its first professor of astronomy. She had a gift for writing and rhetoric as well as for science, and her lectures at Vassar on astronomy were inspiring and memorable. She used her observations to motivate her students to learn about astronomy and develop a philosophy of life. In 1 869 she led a group of young college women westward by train from Poughkeepsie, New York, to Burlington, Iowa, to observe a total solar eclipse. They saw a splendid display of rosy prominences and brilliant streamers. They also recorded scientific data that were published in an official eclipse report.

Although Mitchell continued her astronomical observations and calculations at Vassar, she excelled as an educator, role model, advocate for women's rights, and proponent of women's scientific abilities. Maria did not hesitate to speak up for women's rights and equal salaries, and she felt compelled to question and challenge authority on many college policies, such as requiring attendance in chapel and in the classroom. Her influence began to extend beyond the college campus, and in 1873 she became a founder of the Association for the Advancement of Women (AAW). With her numerous honors in astronomy, she presided over the AAW Science Committee and continued to promote public awareness of women's distinguished contributions to science. In turn, some of her students like Mary Whitney and Antonia Maury followed directly in her footsteps. Whitney succeeded Mitchell in 1888 as director of Vassar Observatory, and Maury worked at Harvard where she made important contributions to the spectral classification of stars. Other students like Ellen Swallow Richards and Christine Ladd-Franklin pursued graduate degrees in related sciences. These women also broke tradition and managed to combine marriage with a career in science.

To perpetuate recognition for the work and ideas of "America's first woman astronomer," Mitchell's family suggested in the late 1890's that Vassar alumnae use the homestead on Nantucket to establish the Maria Mitchell Association. By 1908 the organization had built a small brick observatory in the yard beside the Mitchell House. Somewhat later it added a study. The complex as I knew it several summers ago has recently expanded both architecturally and astronomically. Today, students sit at modern computer workstations to reduce and analyze observations that were made at distant observatories using the latest instrumentation. The old homestead, however, remains much the same as it was when Maria lived there. Students can still step into the past and wonder how Maria ever discovered a comet with a 21/2-inch telescope and how she ever computed hundreds of sums by hand in her little study some 150 years ago.

that the temperature in the star's core is soon high enough for the main hydrogen fusion reaction to be the CNO cycle rather than the PP cycle. Both cycles work, of course, but with two cycles active, massive stars burn fuel very rapidly, have very high luminosities, plot on the upper left hand of the main sequence, and quickly deplete their fuel. As a result, massive stars have short lives on the main sequence.

Once off the main sequence, massive stars go though the same steps of hydrogen burnout, core contraction, shell fusion of hydrogen, core fusion of helium, helium burnout, and contraction of a carbon-rich core. Next, however, a very different thing happens. As the carbon core contracts and shell fusion of helium proceeds, a temperature is reached where carbon starts to fuse to heavier elements. This happens because the core is much larger than the core in a 1-S star, and this is the environment in which all the heavy elements now in the Earth are believed to have formed. Such heavy-element formation is thought to happen in a flash and to release so much energy that the star blows up in a **supernova.** After a supernova, what remains of the core is crushed into an immensely dense mass, a *black hole.* Scientists speculate that the matter now in the Sun and the planets was blasted into space in one or more supernovas about 10 billion years ago. In a very real sense, all the atoms in our bodies and in everything around us are Stardust from an ancient supernova.

Summary

1. The Sun, which is so hot it is gaseous throughout, has a diameter that is 109 times the diameter of the Earth but a density that is only a quarter that of the Earth.
2. Viewed from the Earth, the Sun *appears* to make a complete circuit of 360° against the background of fixed stars. The apparent motion of the Sun is actually due to the Earth's orbit around the Sun. The constellations through which the Sun's apparent motion carries it are the constellations of the zodiac.
3. The rate at which energy leaves the Sun in the form of electromagnetic radiation (that is, the luminosity) is 3.8×10^{26} watts. The rate at which energy reaches the Earth is only 1.8×10^{17} watts, or 1370 w/m^2.
4. The source of the Sun's energy is nuclear fusion, which occurs in the core and involves the fusion of four hydrogen nuclei to produce one helium nucleus plus energy. Eighty-eight percent of the Sun's energy arises from the PP chain and 12 percent from the CNO chain.
5. The Sun has an internal structure. At the center is the core, the site of nuclear fusion. Surrounding the core is a radiative layer, which is a gas containing atomic particles through which electromagnetic radiation moves very slowly. Beyond the radiative layer is the convective layer, and then successively outward are the photosphere, chromosphere, and the corona.
6. The Sun is a blackbody radiator with a temperature of 5800 K.
7. During periods of an active Sun, sunspots—areas of great turbulence—appear on the surface of the photosphere and gigantic prominences burst out into the chromosphere. The causes of the sunspots and prominences are the differential speed of rotation of the Sun's gas—faster at the equator, slow at the poles—and the Sun's magnetic field.
8. A plot of luminosity versus blackbody temperature of stars is called a Hertzsprung-Russell (H-R) plot. An H-R plot is used to follow the life cycle of a star.
9. Stars spend most of their lifetimes on the main sequence of an H-R plot. As fuel burns out, they move off the main sequence and become red giants. In Sun-sized stars, the red giant phase is followed by a white dwarf phase. In a star much larger than our Sun, the red giant phase is followed by a supernova, which is a tremendous explosion that destroys the star.

Important Terms to Remember

Terms in italic are defined in A Closer Look

black body radiator (p. 54)
constellation (p. 46)
electromagnetic radiation (p. 50)
flux (p. 48)
galaxy (p. 45)
Hertzsprung-Russell diagram (H-R) (p. 59)
luminosity (p. 47)
main sequence (p. 62)
red giant (p. 62)
reflection (p. 60)
refraction (p. 60)
shell fusion (p. 63)
spectrum (p. 50)
supernova (p. 65)
white dwarf (p. 62)
zodiac (p. 47)

Questions for Review

1. Describe how you can demonstrate that the Sun follows a regular path against the background of fixed stars. Why do we refer to the Sun's motion as an apparent motion?
2. What are constellations, and what is special about the constellations of the zodiac?
3. Explain the source of the Sun's energy. What is luminosity and how do astronomers measure it?
4. The temperature in the core of the Sun reaches 15×10^6 K, yet the temperature of the surface of the Sun is only 5800 K. Explain.

5. Sketch a cross section through the Sun and label the various layers.
6. What is meant by a blackbody radiator? What is the difference in the radiation spectrum of a blackbody radiator at 6000 K and one at 2000 K?
7. Why and how does the Sun's spectrum of electromagnetic radiation in space differ from that measured at the surface of the Earth?
8. What is a Hertzsprung-Russell diagram? Where does the Sun plot on such a diagram?
9. Briefly describe how a Hertzsprung-Russell diagram is used to follow the life cycle of a star.

Questions for A Closer Look

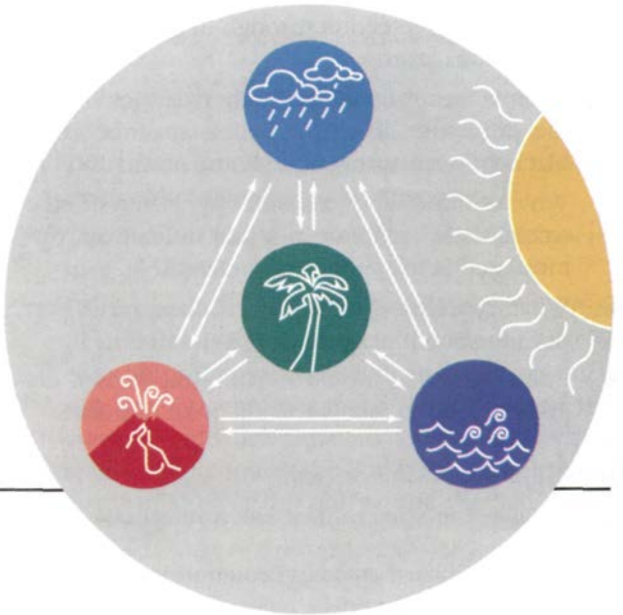
1. What is the spectrum of electromagnetic radiation?
2. What are the three essential properties of waves?
3. What is the relationship between the frequency and the wavelength of electromagnetic radiation?
4. Describe how the two kinds of optical telescopes work.
5. Name some types of nonoptical telescopes. Do they work on the principle of refraction or reflection?
6. Explain why astromomers believe it is important to have telescopes in orbit around the Earth or even mounted on the Moon.

Questions for Discussion

1. Do you consider the launching of more sophisticated space telescopes worthwhile? Why or why not? If yes, what kind of information would you hope to gain from such exploration?
2. When was the most recent burst of sunspot activity? (You will need to do some research to find out.) Was any change in climate associated with the activity?
3. Do you think it may someday be possible to generate energy on the Earth in the same way the Sun generates energy? What advantages or drawbacks do you foresee if fusion energy is attained?

PART TWO

The Earth Beneath Our Feet



The Way the Earth Works

For centuries, people have been asking such questions as, why do continents have such irregular shapes and why do ocean basins, mountain ranges, earthquakes, volcanoes, and many other features occur where they do? In the sixteenth century, when the coastlines on either side of the Atlantic Ocean were first mapped, it was apparent that the coastlines run parallel, like the two banks of a gigantic river. People started to speculate why. Some thought that a flood—perhaps the biblical flood—had cut an immense canyon. Such speculations get people thinking about why the Earth is the way it is. Initially, individual hypotheses were offered for almost every feature on the Earth, but eventually scientists began to think that there might be a single, underlying cause for the whole array. What could that cause possibly be?

Scientists first thought about the consequences of the Earth having once been molten. When the Earth cooled sufficiently for a solid, rocky crust to form, they reasoned, continued cooling caused the crust to contract, be compressed, and become wrinkled like an old apple. They pointed to mountain ranges full of bent and deformed rocks as the places where past contractions occurred and to regions of intense earthquake activity as places where present contractions could be happening. Although this contraction model does explain some features, it does not help with questions about the parallelism of the Atlantic coast, the distribution of continents and ocean basins, or why volcanoes are where they are. Nor does this model explain features like the great African Rift Valley or the Rio Grande Valley in New Mexico, two elongate zones where numerous fractures cut through the crust because it is being stretched.

At the beginning of the twentieth century, when scientists discovered that the Earth's interior is kept hot by naturally occurring radioactive elements, some of them offered the hypothesis that the Earth might be heating up (and therefore expanding). A much smaller Earth, they hypothesized, could once have been covered entirely by continents. Heating caused the Earth to expand, and as a result, the crust stretched, leading the continents to crack into irregularly shaped fragments. As expansion continued, the cracks grew into ocean basins, lava oozed up through the cracks, and the sea floor became covered with volcanic rock. The hypothesis of expansion does offer a plausible explanation for the parallel coastlines of the Atlantic Ocean continents, for the African Rift Valley, and for the fact that the sea floor is underlain by lava, but it does not easily account for mountain ranges formed by compression.

Both the expansion hypothesis and the contraction hypothesis postulate an Earth on which the continents remain fixed in the same relative positions. The size of the Earth may change, but the positions of the continents are fixed just as the markings on a balloon are unchanging whether the balloon is inflated or deflated. To get around the flaws in both hypotheses, scientists eventually began to hypothesize about the possibility of a "mobilist" Earth on which the continents move around. As to what forces might be large enough to cause such movement, scientists were not

Villarica in Chile and Lanin in Argentina (rear), active volcanoes in the southern Ancles mountains of South America. The volcanoes are members of a chain of volcanoes situated above the western edge of the South American plate.

at all clear, however, and some of their ideas were very imaginative. One suggestion was that a huge chunk of the spinning Earth had broken free and become the Moon; the tremendous shock of the breakup caused the irregular shapes of the continents and moved them to their present locations. Another suggestion was that the gravitational attractions of the Sun and Moon cause tides in the solid Earth just as they do in the ocean. At some time in the past, these "Earth tides" had caused very large continents to break up and the fragments, today's continents, to move to their present positions.

All the hypotheses have attractive features, but none answers all the questions and none is fully supported by evidence. By the middle of the twentieth century, all reasonable hypotheses concerning the shapes and positions of continents, mountain ranges, and volcanoes seemed to have been exhausted. The time was ripe for a totally new approach.

That new approach turned out to be plate tectonics. This approach says that the outer 100 km of the Earth's surface moves not as one piece but in several fragments. The driving force is not external; what makes these fragments move is the earth's internal heat energy, through the process of convection. The next five chapters discuss the way the Earth works and the role of plate tectonics in shaping the Earth.

CHAPTER 3

Earthquakes and the Earth's Interior

Damage in Cairo, Egypt, caused by an earthquake, magnitude 5.9, on October 12, 1992. Several factors accounted for the severity of damage: the earthquake focus was at a depth of 10 km immediately below Cairo. The city is densely populated, many dwellings are not built to withstand earthquakes, and much of the city is built on weak, unconsolidated sediment deposited by the River Nile.

Bad News and Good about Earthquakes

There is a saying among geologists and engineers that earthquakes don't kill but buildings do. Shaking ground may make people fall down, and falls may break legs and arms, but they don't kill. However, shaking ground can make buildings collapse, and collapsing buildings can definitely kill.

The worst earthquake of the twentieth century (so far) happened on July 28, 1976. At 3:45 A.M., while 1 million inhabitants of T'ang Shan, China, slept, a 7.8 magnitude quake leveled the city. Hardly a building was left standing, and the few that did withstand the first quake were destroyed by a second, magnitude 7.1, which struck at 6:45 P.M. the same day. When the wreckage of T'ang Shan was cleared, 240,000 people were dead. Losses were large because most of the buildings had not been constructed to withstand an earthquake. They had unreinforced brick walls. When the ground started to shake, the walls collapsed, the roofs caved in, and the sleeping inhabitants were crushed.

Earthquakes are caused by sudden releases of stored elastic energy in the Earth, and the source of most of that energy is plate tectonics. One of the great challenges scientists face today is to understand the dynamics of the Earth system sufficiently well that they can accurately predict the timing and probable size of an earthquake and thereby assure human safety. Although Chinese scientists have been more successful than most in predicting quakes, the T'ang Shan quake gave no recognizable warning signs and was completely unexpected. We still have a long way to go.

Not everything about earthquakes is bad; in fact, some things about them are so important to our knowledge of how the Earth works that they might even be called good. Earthquakes can be used to study the Earth's interior, for example. The way the Earth vibrates after a large quake is controlled in large degree by the properties of the rocks inside. Used in this way, earthquake vibrations are like the X rays a doctor uses to study the inside of a human body—they are the probes we use to sense and measure the world beneath our feet.

In this chapter, we discuss how earthquakes occur and how the vibrations they cause can be used to assemble a picture of the Earth's interior. We also discuss how earthquakes are measured and how they cause damage.

EARTHQUAKES

Place a thin piece of wood across your knees, press both ends downward, and the wood bends. Stop pressing and the wood springs back to its original shape. Any change of shape or size that disappears when the deforming forces are removed is called **elastic deformation.** The muscle energy you used to bend the wood doesn't disappear—it is stored as elastic energy in the wood. When the bending force is removed, it is this elastic energy that restores the wood to its original shape. Consider what happens, however, when the pressure is so great that the elastic limit is exceeded and the wood breaks with a sudden snap. The elastic energy is now converted in part to heat at the breakage point, in part to the sound waves that make the snapping noise, and in part to vibrations in the wood.

Earthquake vibrations are the same kind of vibrations that you feel when the wood breaks. When the Earth quakes, it is as if a huge log has broken somewhere inside the Earth. The more energy released, the stronger the quake. Just how elastic energy is stored and built up in the Earth, however, and how and why it is suddenly released continue to be the subjects of intensive research.

Of course, there are no logs inside the Earth, and so breaking logs is not the cause of earthquakes. However, the sudden breaking of a mass of rock or the sudden slippage of two rock masses past each other will serve just as well. The most widely accepted theory concerning the origin of earthquakes involves slipping rock masses and the elastic rebound theory.

Origin of Earthquakes

When slippage of rock occurs along a fracture in a rock, the fracture is called a **fault.** The cause of most earthquakes is thought to be sudden movement along faults, but it cannot be that simple. Some earthquakes are millions of times stronger than others. The same amount of energy that in one case is released by thousands of tiny slips and tiny earthquakes is in another case stored and released in a single immense earthquake. The **elastic rebound theory** suggests that, if fault surfaces lock rather than slip easily past one another, the rocks on either side of the fault will bend and in bending will store elastic energy, just as in a bent piece of wood. Then, when the fault finally does slip and the bent rocks rebound to their original shapes, an enormous amount of energy is released in a huge earthquake.

The first evidence supporting the elastic rebound theory came from studies of the San Andreas Fault: the fault is a vertical fracture reaching deep into the crust, and rock bending along the San Andreas Fault is due to the two sides of the fault moving in opposite directions. During long-term field observations in central California, beginning in 1874, scientists from the U.S. Coast and Geodetic Survey determined the precise position of many points both adjacent to and distant from the fault (Fig. 3.1). As time passed, changes in the relative positions of these points revealed that the Earth's crust on each side of the fault was slowly being bent. For some reason, in the area of the measurement near San Francisco, the fault was locked and did not slip. On April 18, 1906, the two sides of the locked fault shifted abruptly. The elastically stored energy was released as the rock masses moved and the bent crust rebounded back, thereby creating a violent earthquake. Repetition of the survey then revealed that the rocks on each side of the fault were no longer bent.

The 1906 San Francisco earthquake caused an enormous amount of damage, but it also started a

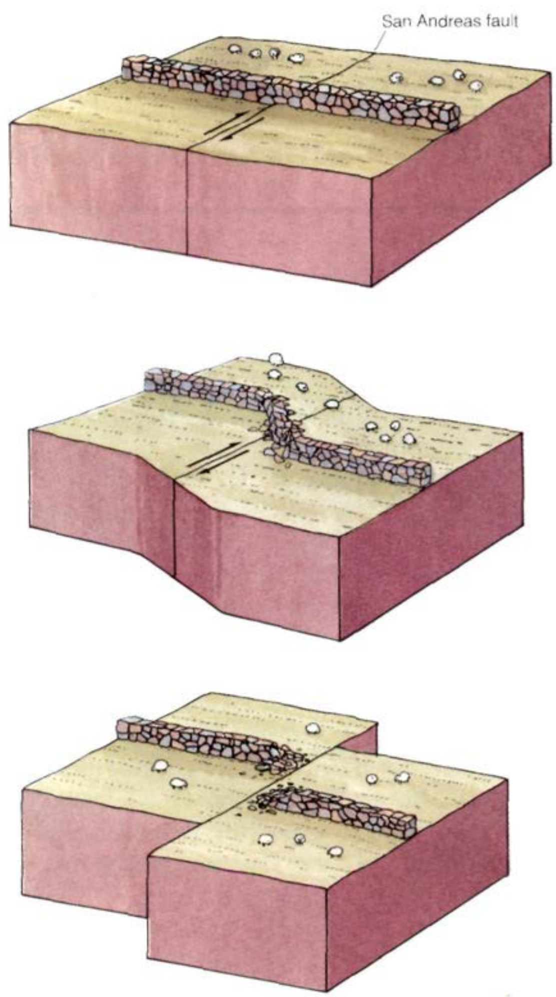


Figure 3.1 An earthquake is caused by the sudden release of elastic energy stored in rocks. This sketch is based on surveys near the San Andreas Fault, California, before and after the earthquake of 1906. A stone wall crosses the fault and is slowly distorted as the rock beneath it is bent and elastically strained. After the earthquake, two segments of the wall are offset 7 m.

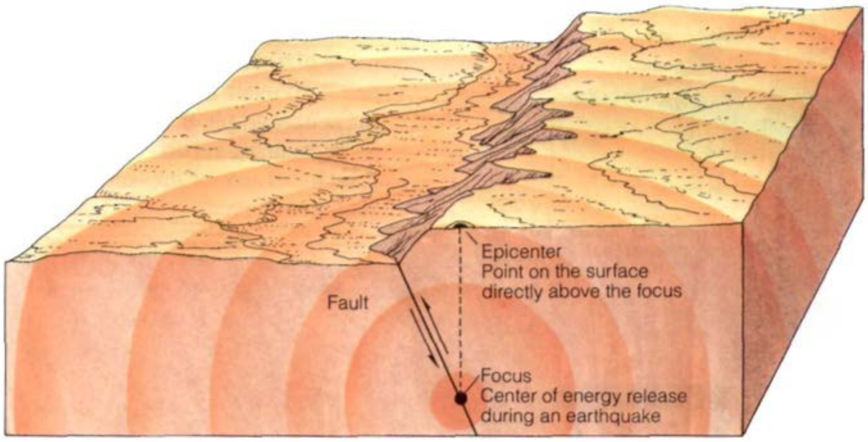


Figure 3.2 The focus of an earthquake is the site of first movement on a fault and the center of energy release. The epicenter of an earthquake is the point on the Earth's surface that lies vertically above the focus.

great deal of research. That research has shaped our present-day understanding of how earthquake damage can be minimized and how earthquakes can be used to study the internal structure of the Earth. Before proceeding with a discussion of these points, let's discuss in a little more detail how earthquake vibrations travel through rock.

Seismic Waves

The point where energy is first released is called the **earthquake focus.** In reality, because most earthquakes are caused by rock movement along a fault, the focus is not a point but rather a region that may extend for several kilometers (Fig. 3.2). Because the focus generally lies at some depth below the Earth's surface, it is more convenient to identify an earthquake site from the **epicenter,** which is the point on the Earth's surface that lies vertically above the focus (Fig. 3-2). The usual way to describe the location of an earthquake focus is to state its depth and the location of its epicenter.

When an earthquake occurs, the elastically stored energy is carried outward from the focus to other parts of the Earth by vibrations. The vibrations, which are also called **seismic waves,**[1] spread out spherically in all directions, just as sound waves spread out spherically in all directions from a sound source.

Seismic waves are elastic disturbances, and so unless their elastic limit is exceeded, the rocks through which the waves pass return to their original shapes and there is no permanent record of the vibrations. Seismic waves must therefore be recorded while the rock is still vibrating. For this reason, many continuously recording devices that can detect seismic waves, called *seismographs* (Fig. 3 -3), have been installed around the world.

There are several kinds of seismic waves, and they belong to two families. **Body waves** travel outward in all directions from the focus and have the capacity to travel through the Earth's interior. **Surface waves,** on the other hand, travel around but not through the Earth; they are guided by the Earth's surface. Body waves are analogous to light and sound waves, both of which travel outward in all directions from a source. Surface waves, analogous to ocean waves, travel only on the Earth's *solid* surface, both where it meets the atmosphere and where it meets the ocean.

Body Waves

Rocks can be elastically deformed by body waves in two ways: by a change in shape (like bending or twisting a piece of wood) or by a change in volume (like squeezing a tennis ball or blowing up a balloon).

Body waves that cause volume changes consist of alternating pulses of compression (squeezing) and expansion (stretching) acting in the direction of wave travel (Fig. 3-4A). Sound waves are a familiar example of compressional/expansional waves. A sound wave passes through air by alternating compressions and expansions of the air. Our ears sense the pulses of compression and expansion, and our brains transform these pulse signals into the sound. Compressional/expansional waves can pass through gases, liquids, and solids. That is why we can hear sounds not only in the air but also through the walls of houses and when we swim under water. Compressional/expansional waves can pass easily through rocks. They have the greatest velocity of all seismic waves—6 km/s (3-7 mi/s) is a typical value near the Earth's surface—and they are the first to be recorded by a seismograph after an

[1] "Seismic" means caused by an earthquake and comes from the Greek verb seiein, to shake.

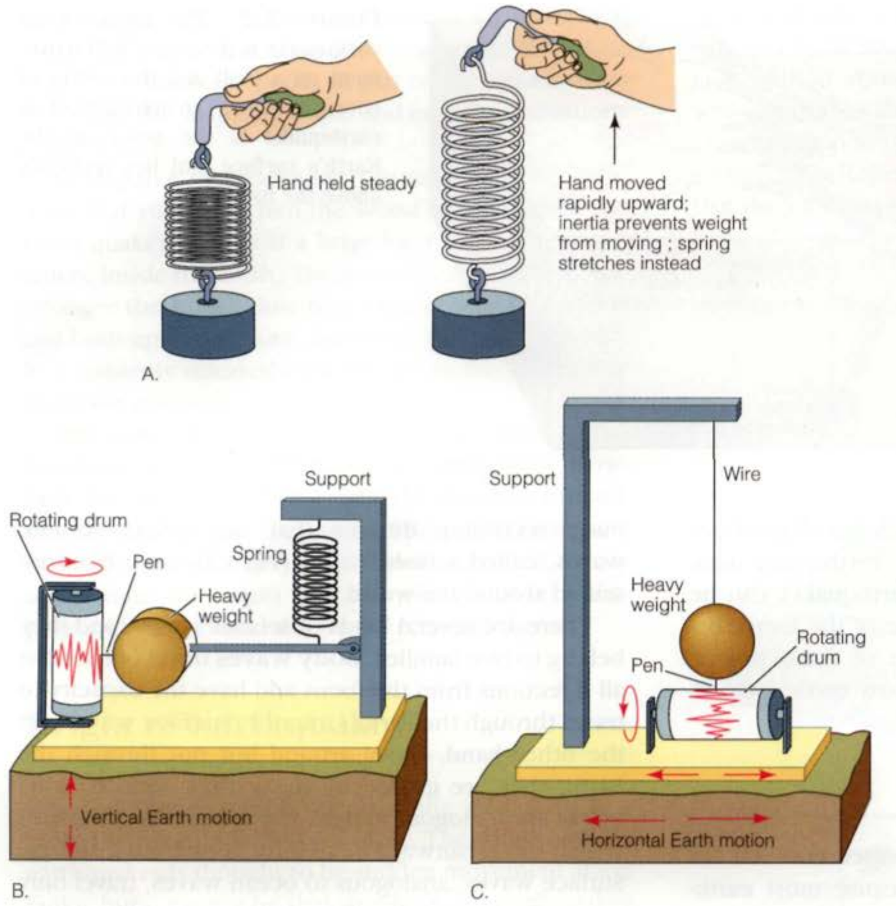


Figure 3.3 Seismographs make use of inertia, which is the resistance of a stationary weight to sudden movement. A. The principle of the seismograph. B. A seismograph for measuring vertical motions. C. A seismograph for measuring horizontal motions.

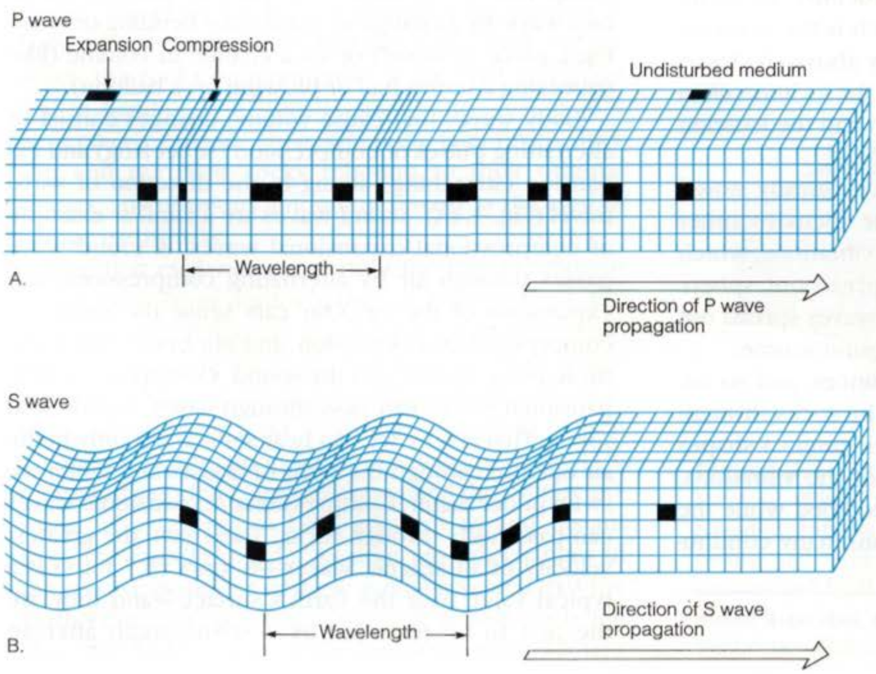


Figure 3.4 Body waves of the P (compressional/expansional) and S (shear) types. A. P waves cause volume changes in the rock through which the wave is passing by alternate compressions and expansions. An individual point in a rock moves back and forth parallel to the direction of P-wave propagation. As wave after wave passes through, a square repeatedly expands to a rectangle, returns to a square, contracts to a smaller rectangle, returns to a square, and so on. B. S waves cause a change in rock shape because they result in a shearing motion. An individual point in the rock moves up and down, perpendicular to the direction of S-wave propagation. A square repeatedly changes to a parallelogram, then back to a square.

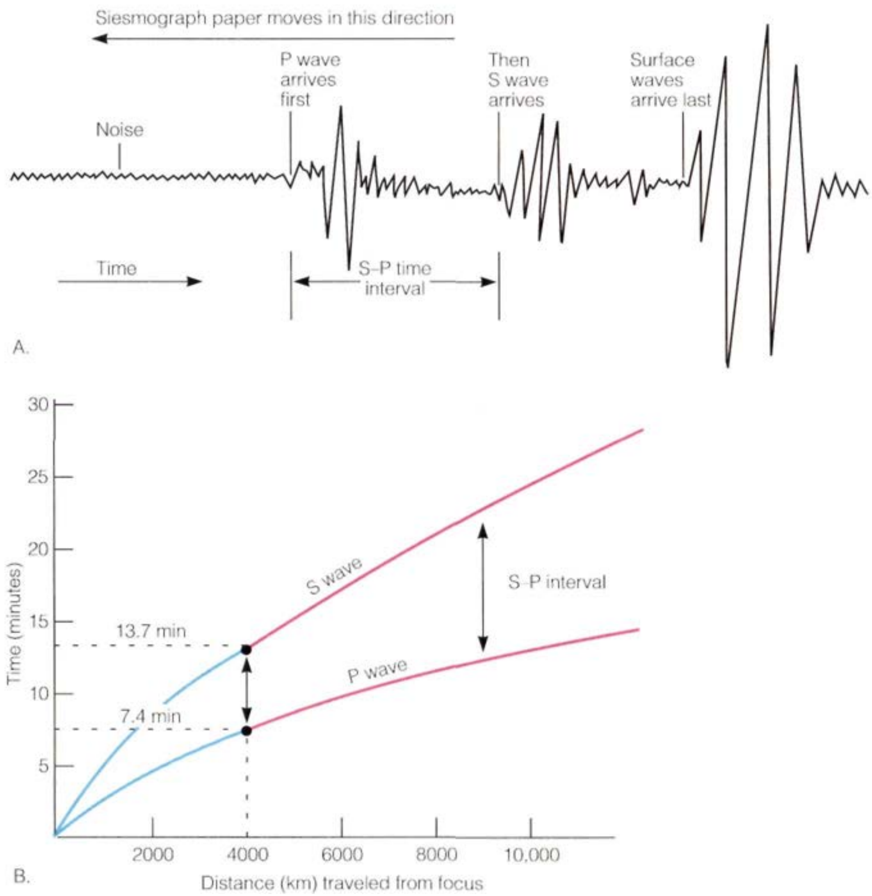


Figure 3.5 Travel times of P body waves, S body waves, and surface waves. A. Typical record made by a seismograph. All three types of waves leave the earthquake focus at the same instant. The fast-moving P waves reach the seismograph first, and some time later the slower moving S waves arrive; the delay in arrival times is proportional to the distance traveled by these two waves. The surface waves travel more slowly than either P or S waves. B. Seismologists use a travel-time graph for P and S waves to locate an epicenter. For example, when a station measures the S-P time interval to be 13.7 min - 7.4 min = 6.3 min, they know the epicenter is 4000 km away.

earthquake. They are therefore called **P (for primary) waves**.

Body waves that deform materials by change of shape are called shear waves. Liquids and gases don't have shapes; they simply flow freely to fill any container we put them in. Therefore, liquids and gases cannot transmit waves that depend on a change in shape, and so shear waves can be transmitted only by solids. As a shear wave travels through a material, each particle in the material is displaced perpendicular to the direction of wave travel (Fig. 3-4B). A typical velocity for shear waves in rocks near the Earth's surface is 3-5 km/s (2.2 mi/s). Because shear waves are slower than P waves and reach a seismograph later, they are called **S (for secondary) waves**.

Seismic body waves behave like light waves, which is to say that, in addition to being able to pass through a medium, they can also be reflected and refracted. *Reflection* is the familiar phenomenon of light bouncing off a mirror or other shiny surface, and body waves are reflected by numerous boundaries in the Earth. The less familiar process, *refraction*, as discussed in Chapter 2, occurs when the speed of a wave changes as it passes from one medium to another, and this change causes the wave to bend (see Fig. C2.3).

Surface Waves

To an observer recording seismic waves at the surface of the Earth, surface waves appear very similar to body waves because both are recorded simply as vibrations. However, surface waves travel more slowly than P and S waves, and in addition they pass around the Earth rather than through it. Surface waves are the last to be detected by a seismograph (Fig. 3.5).

Surface waves can have very long wavelengths—up to hundreds of kilometers—and the longer the wavelength the greater the amplitude (wave height). Just as an ocean wave disturbs the water to some distance beneath the ocean surface, so do surface waves cause rock below the Earth's surface to be disturbed. The greater the amplitude, the deeper the wave motion reaches.

Layers of Different Composition

Reflection and refraction of seismic body waves are the way information is gained about the different compositional layers in the Earth.

The speed of body waves is determined, in part, by the density of the rocks they are passing through. The higher the density, the greater the speed. If the Earth had a homogeneous composition, rock density would increase steadily with depth as a result of increasing pressure: body wave velocities would also increase steadily. Measurements reveal, however, that body waves are also abruptly refracted and reflected at several depths inside the Earth. This means that density does not increase smoothly, and within the Earth there must be some boundaries separating materials having distinctly different densities. The most pronounced of these boundaries occurs at a depth of 2883 km. When P waves reach the 2883-km (1791 mi) boundary, they are refracted so strongly that the boundary is said to cast a P wave shadow, which is an area of the Earth's surface opposite the epicenter where no P waves are observed (Fig. 3.6). Because this 2883-km boundary is so pronounced, geologists infer that it is the compositional boundary between the mantle and the core. The same boundary casts an even larger S-wave shadow. Here, however, the reason is not refraction, but the fact that S waves cannot travel through liquids. Therefore, the huge S-wave shadow lets us conclude that at least the outer portion of the core must be liquid.

Seismic waves cannot tell us the composition of the core, but they help us to deduce what it might be. Seismic-wave speeds indicate that rock density increases slowly from about 2.7 g/cm^3 at the top of the crust to about 5.5 g/cm^3 at the base of the mantle. The average density of the whole Earth is 5.5 g/cm^3. Therefore, to balance the lower density of the crust and upper mantle, the core must be composed of material having a density of at least 10 g/cm^3. The only common substance that comes close to fitting this requirement is iron with a little admixture of nickel.

Early in the twentieth century, the existence of the compositional boundary between the Earth's crust and mantle was demonstrated by a Croatian scientist named Andrija Mohorovicic. For earthquakes 'whose focus lay within 40 km (25 mi) of the surface, Mohorovicic noticed that seismographs about 800 km (500 mi) from the epicenter recorded two sets of body waves that arrived at the seismograph at different times. He concluded that the set that arrived second must have traveled from the focus to the station by a direct path through the crust, whereas the set that arrived first must have been refracted into rock that was denser than crustal rock. These refracted

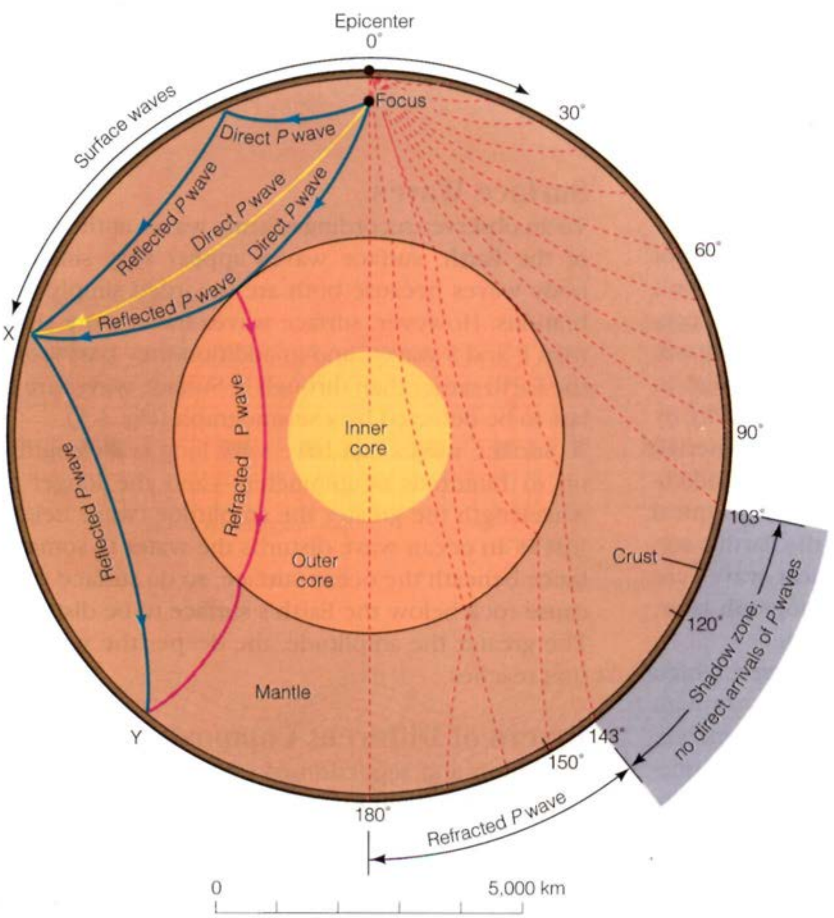


Figure 3.6 Refraction and reflection of body waves. On the left-hand side of the figure are shown various paths of P waves in the Earth, which is made up of concentric spheres of different compositions. Seismographs at some places (locations X and Y, for example) receive both direct P waves and reflected and refracted P waves. The right-hand side shows paths of P waves moving out from an epicenter at 0°. Reflection and refraction of P waves at the core-mantle boundary create a P-wave shadow from 103° to 143°.

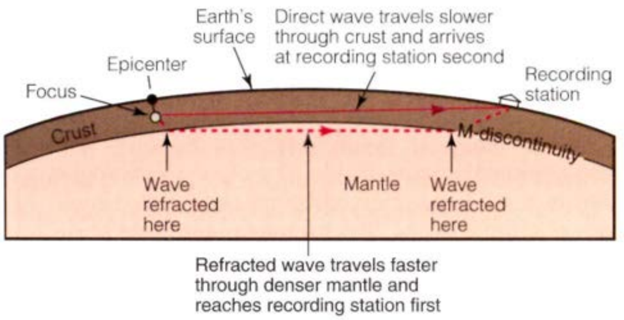


Figure 3.7 Travel paths of direct and refracted body waves from shallow focus earthquake to nearby seismograph station.

waves, moving through the denser zone, traveled faster within that zone and so reached the surface first (Fig. 3.7). Mohorovicic hypothesized that a distinct compositional boundary separates the crust from this underlying zone of denser composition. Scientists now refer to this boundary as the **Mohorovičić discontinuity** and recognize it as the boundary that marks the base of the crust (or, put another way, the top of the mantle). Crust thickness ranges from 30 to 70 km (19 to 43 mi) in continental regions but is only 8 km (5mi) beneath oceans. The feature is commonly called the **M discontinuity** and in conversation is shortened still further to **moho.**

Layers of Different Strength

Density is not the only rock property that affects the speed of seismic waves. Rock strength also plays a role. Rock strength is an expression of the elasticity, and this in turn can be equated to the tendency of a rock to fracture (called brittleness; higher brittleness means higher elasticity), as opposed to a tendency to deform and flow like putty (called ductility; higher ductility = lower brittleness = lower elasticity). Rock strength, which is strongly affected by temperature and pressure, has a marked effect on the speed of both body and surface waves. The more ductile a rock, the lower the speed.

Studies of seismic wave speeds at various depths have given us the following picture of our Earth's interior. In addition to the boundary between the crust and the mantle (the moho), and between the mantle and core, there are three strength boundaries. The first boundary is 100 km below the Earth's surface and separates brittle rocks above from ductile rocks below. This is the lithosphere asthenosphere boundary. At about 350 km (220 mi) there is a diffuse boundary between very ductile rock above and less ductile rock below: this is the asthenosphere mesosphere boundary. Finally, there is a boundary between molten iron above and solid iron below: this is the outer core inner core boundary.

Location of the Epicenter

The location of an earthquake's epicenter can be determined from the arrival times of the P and S waves at a seismograph. The farther a seismograph is from an epicenter, the greater the time difference between the arrival of the P and S waves (Fig. 35B). After using a graph like the one shown in Figure 3.5B to determine how far an epicenter lies from a seismograph, the seismologist draws a circle on a map around the station with a radius equal to the calculated distance to the epicenter. The exact position of the epicenter can be determined when data from the three or more seismographs are available—the center lies where the circles intersect (Fig. 3 8).

If a local earthquake is recorded by several nearby seismographs, the focal depth can be determined in the same way that the epicenter is determined, by

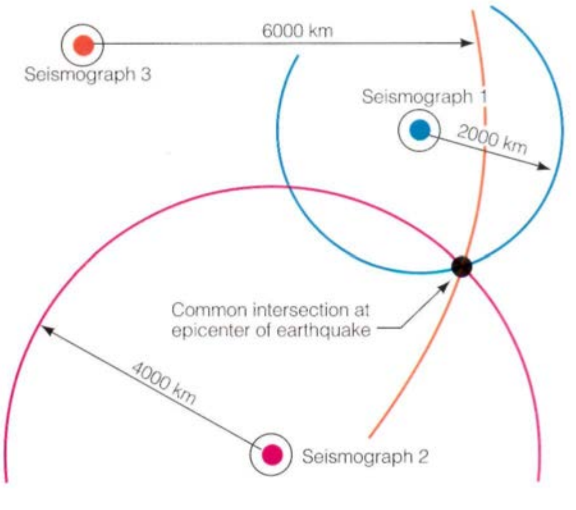


Figure 3.8 Locating an epicenter. The effects of an earthquake are felt at three seismograph stations. The time interval between the arrival of the first P and the first S waves depends on the station epicenter distance. The following distances are calculated by using the curves in Figure 3.5B.

		Time Interval	Calculated Distance
Seismograph 1	8.8 min	4.7 min = 4.1 min	2000 km
Seismograph 2	137 min	7.4 min = 6.3 min	4000 km
Seismograph 3	17.5 min	98 min = 7.7 min	6000 km

On a map, a circle of appropriate radius is drawn around each station. The epicenter is where the three circles intersect.

A Closer Look

Earthquake Magnitudes

Very large earthquakes (of the kind that destroyed San Francisco in 1906; T'ang Shan, China, in 1976; parts of Mexico City in 1985; parts of San Francisco again in 1989; and parts of Los Angeles in 1994) are, fortunately, relatively infrequent. In earthquake-prone regions, massive earthquakes occur about once a century. They occur more frequently in some areas and less frequently in others, but a century is an approximate average.

This frequency rate means that the time needed to build up elastic energy to a point where the frictional locking of a fault is overcome is about 100 years. Small earthquakes may occur along a fault during this time as a result of local slippage, but even so, elastic energy is accumulating because most of the fault remains locked. When the lock is broken and an earthquake occurs, the elastic energy is released during a few terrible minutes. By careful measurement of elastically strained rocks along the San Andreas Fault, seismologists have found that about 100 joules 0) of elastic energy can be accumulated in 1 m^3 (35 ft^3) of deformed rock. This is not very much—100 J is equivalent to only about 25 calories of heat energy—but when billions or trillions of cubic meters of rock are strained, the total amount of stored energy can be enormous. The amount of elastically stored energy released during the Loma Prieta earthquake of 1989 was about 10^{15} J, and the 1906 San Francisco earthquake released at least 10^{17})!

The Richter Magnitude Scale

Measurements of the bending of elastically deformed rocks before an earthquake, and of those same rocks after an earthquake has released the deforming force, can provide an accurate measure of the amount of the energy released. The task is very time consuming, however, and all too frequently the pre-earthquake measurements are simply not available. Therefore, seismologists estimate the energy released by measuring the amplitudes of seismic waves. The **Richter magnitude scale,** named after Charles F. Richter, the seismologist from the California Institute of Technology who developed it, is defined by the maximum amplitudes of seismic waves (that is, the heights of the waves on a seismogram) 100 km (62 mi) from an epicenter. Because wave signals vary in strength by factors of a hundred million or more, the Richter scale is logarithmic, which means it is divided into steps called magnitudes, starting with magnitude 1 and increasing upward. Each unit increase in magnitude corresponds to a tenfold increase in the amplitude of the wave signal. Thus, a magnitude 2 signal has an amplitude that is ten times larger than a magnitude 1 signal, and a magnitude 3 is a hundred times larger than a magnitude 1 signal.

We can see from Figure C3.1 how a Richter magnitude is calculated. The energy of a seismic wave is a function of both its amplitude and the duration of a single wave oscillation, *T.* Divide the maximum amplitude, *X,* measured in steps of 10^{-4} cm on a suitably adjusted seismograph, by *T,* measured in seconds. Then add a correction factor, *Y,* determined from the S-P wave interval. The ratio *X/T* is a measure of the maximum energy reaching the seismograph. The formula is

$$M = \log X/T + Y$$

where *M* is the Richter magnitude.

One Richter magnitude scale unit corresponds to a tenfold increase in *X.* However, the energy increase is proportional to X^2, which is to say, a hundredfold. The duration of a single oscillation differs greatly from one earthquake to another. In particular, the most energetic earthquakes have a higher proportion of long-duration waves. As a result, the energy increase corresponding to one Richter scale unit increase, when summed over the whole range of waves in a wave record, is only a thirtyfold increase. Thus, the difference in energy released between an earthquake of magnitude 4 and one of magnitude 7 is 30 X 30 X 30 = 27,000 times!

How big can earthquakes get? The largest recorded to date have Richter magnitudes of about 8.6, which means they release about as much energy as 10,000 atom bombs of the kind that destroyed Hiroshima at the end of World War II. It is possible that earthquakes do not get any larger than this because rocks cannot store more elastic energy. Before they are deformed further, they fracture and so release the energy.

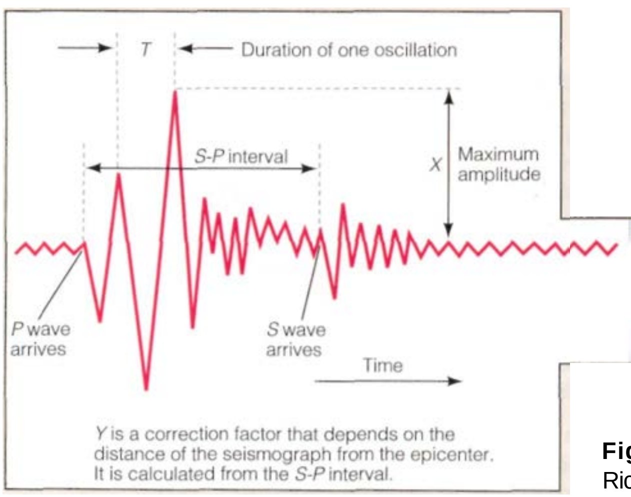


Figure C3.1 Measurements used for determining the Richter magnitude (M) from a seismograph record.

using the P-S time intervals. For distant earthquakes, a different method is employed. Note in Figure 3.6 that a seismograph at position X would record both a direct P wave and a P wave reflected from the Earth's surface (labeled PP). The direct P wave, having a shorter path length, would arrive before the PP wave. The travel-time difference between them is a measure of the focal depth of the earthquake.

Location of an epicenter and focal depth are only part of the information that can be read from the seismograph records. Of equal importance is the calculation of the amount of energy released during an earthquake or, as it is commonly stated, the magnitude of the earthquake. To see how magnitudes are calculated, see "A Closer Look: Earthquake Magnitudes."

EARTHQUAKE RISK

Most people in the United States think immediately of California when earthquakes are mentioned. However, the most intense earthquakes to jolt North America in the past 200 years were centered near New Madrid, Missouri. Three earthquakes of great size occurred on December 16, 1811, and January 23 and February 7, 1812. The exact sizes of these earthquakes are unknown because instruments to record them did not exist at the time. However, judging from the local damage and from the fact that tremors were felt and minor damage occurred as far away as New York and South Carolina, it is estimated that the largest of these quakes was larger than the one that leveled San Francisco in 1906.

Based on known geological structures (mainly faults) and on the location and intensity of past earthquakes, the National Oceanographic and Atmospheric Administration prepared the seismic-risk map shown in Figure 3.9.

Earthquake Disasters

Every year the Earth experiences hundreds of thousands of earthquakes. Fortunately, only one or two either are large enough or close enough to major population centers to cause loss of life. Certain areas are known to be earthquake-prone, and special building codes in such places require structures to be as resistant as possible to earthquake damage. All too often, however, an unexpected earthquake will devastate an area where buildings are not adequately constructed, as in the T'ang Shan earthquake discussed at the beginning of this chapter. Other examples are the earthquake that destroyed parts of the center of Mexico City and killed 9500 people in 1985 (Fig. 3.10) and one that struck in Armenia in 1988, killing an estimated 25,000 people (Fig. 3.11).

Historically, seventeen earthquakes are known to have caused 50,000 or more deaths apiece (Table 3.1). The most disastrous one on record occurred in 1556 in Shaanxi Province, China, where an estimated 830,000 people died. Many of those people lived in cave dwellings excavated in a soft, wind-deposited sediment called loess, which collapsed as a result of the quake. Since 1900, there have been 42 earthquakes worldwide in which 500 or more people have died.

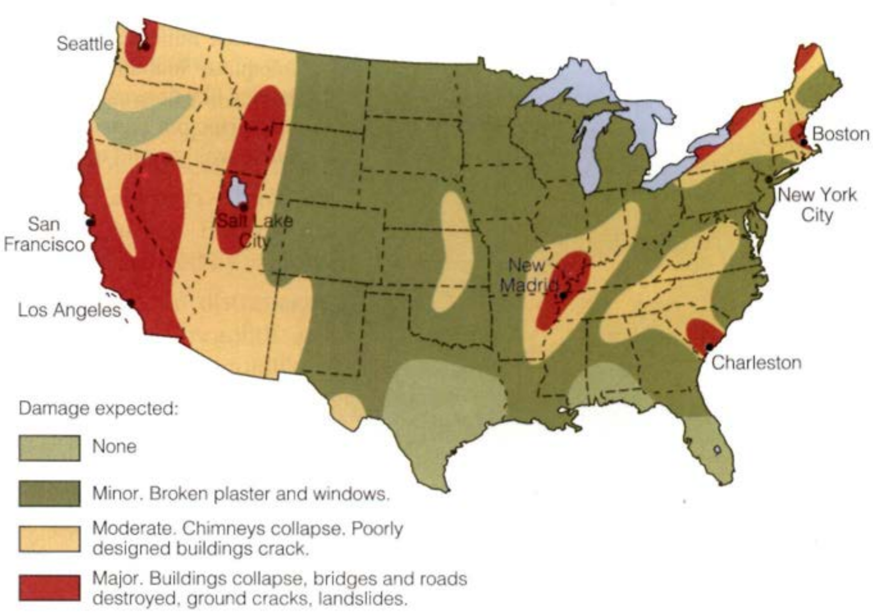


Figure 3.9 Seismic-risk map of the United States based on quake intensity. The map does not indicate earthquake frequency. For example, frequency is high in southern California but low in eastern Massachusetts. Nevertheless, when earthquakes do occur in eastern Massachusetts, they can be as severe as the more frequent quakes in southern California.

Figure 3.10 A building that was not constructed to withstand expected earthquakes, the Hotel DeCarlo, was one of the buildings that collapsed during the earthquake that struck Mexico City in 1985. Proper building design can minimize damage. Nearby buildings of sturdier construction withstood the shaking.

Table 3.1
Earthquakes During the Past 800 Years That Have Caused 50,000 or More Deaths

Place	Year	Estimated Number of Deaths
Silicia, Turkey	1268	60,000
Chihli, China	1290	100,000
Shaanxi, China	1556	830,000
Shemaka, Azerbaijan	1667	80,000
Naples, Italy	1693	93,000
Catania, Italy	1693	60,000
Beijing, China	1731	100,000
Calcutta, India	1737	300,000
Lisbon, Portugal	1755	60,000
Calabria, Italy	1783	50,000
Messina, Italy	1908	160,000
Gansu, China	1920	180,000
Tokyo and Yokohama, Japan	1923	143,000
Gansu, China	1932	70,000
Quetta, Pakistan	1935	60,000
T'ang Shan, China	1976	240,000
Iran	1990	52,000


Figure 3.11 When a magnitude 6.8 earthquake struck Armenia on December 7, 1988, poorly constructed buildings with inadequate foundations collapsed like houses of cards. The principal cause of collapse was ground motion.

Earthquake Damage

The dangers of earthquakes are profound, and the havoc they can cause is often catastrophic. Their effects are of six principal kinds. The first two, ground motion and faulting, are primary effects, and they cause damage directly. The other four effects are secondary and cause damage indirectly as a result of processes set in motion by the earthquake.

1. Ground motion results from the movement of seismic waves, especially surface waves, through surface-rock layers and regolith. The motion can damage and sometimes completely destroy buildings and roads. Proper design (including such features as steel framework and a foundation tied to bedrock) can do much to prevent such damage, but in a very strong earthquake even the best buildings and roads may suffer some damage.
2. Where a fault breaks the ground surface, buildings can be split, roads disrupted, and any feature that crosses or sits on the fault broken apart.
3. A secondary effect, but one that is sometimes a greater hazard than moving ground, is fire. Ground movement displaces stoves, breaks gas lines, and loosens electrical wires, thereby starting fires. Ground motion also breaks water mains, and so there is no water available to put out fires. In the earthquakes that struck San Francisco in 1906 and Tokyo and Yokohama in 1923, more than 90 percent of the building damage was caused by fire.
4. In regions of steep slopes, earthquake vibrations may cause regolith to slip and cliffs to collapse. This is particularly true in Alaska, parts of southern California, China, and hilly places such as Iran and Turkey. Houses, roads, and other structures are destroyed by rapidly moving regolith.
5. The sudden disturbance of water-saturated sediment and regolith can turn seemingly solid ground to a liquidlike mass of quicksand. This process is called liquefaction, and it was one of the major causes of damage during the earthquake that destroyed much of Anchorage, Alaska, on March 27, 1964 (Fig. 3.12), and that caused apartment houses to sink and collapse in Niigata, Japan, that same year.

Figure 3.12 Gaping fissures in a residential area of Anchorage, Alaska, formed during the 1964 earthquake. The fissures result from liquefaction and failure of weak subsurface rocks.

6. Finally, there are **seismic sea waves**, called **tsunami,** which have been particularly destructive in the Pacific Ocean. *Tsunami* is a Japanese term meaning harbor wave. About 5 hours after a severe submarine earthquake near Unimak Island, Alaska, in 1946, for instance, a tsunami struck Hawaii. The wave traveled at a speed of 800 km/h (500 mi/h). Although the amplitude of the wave in the open ocean was less than 1 m (1.1 yd), the amplitude increased dramatically as the wave approached land. When it hit Hawaii, the wave had a crest 18 m (20 yd) higher than normal high tide. This destructive wave demolished nearly 500 houses, damaged a thousand more, and killed 159 people.

Modified Mercalli Scale

Because damage to the land surface and to property is so important, the scale of earthquake-damage intensity (called the **modified Mercalli scale)** is based on the amount of vibration people feel during low-magnitude quakes and the extent of building damage during high-magnitude quakes. The correspondence between Mercalli intensity and Richter magnitude is listed in Table 3-2.

Table 3.2
Earthquake Magnitudes, Frequencies for the Entire Earth, and Damaging Effects

Richter Magnitude	Number per Year	Modified Mercalli Intensity Scale[a]	Characteristic Effects of Shocks in Populated Areas
<3.4	800,000	I	Recorded only by seismographs
3.5-4.2	30,000	II and III	Felt by some people who are indoors
4.3-4.8	4,800	IV	Felt by many people; windows rattle
4.9-5.4	1,400	V	Felt by everyone; dishes break, doors swing
5.5-6.1	500	VI and VII	Slight building damage; plaster cracks, bricks fall
6.2-6.9	100	VII and IX	Much building damage; chimneys fall; houses move on foundations
7.0-7.3	15	X	Serious damage, bridges twisted, walls fractured; many masonry buildings collapse
7.4-7.9	4	XI	Great damage; most buildings collapse
>8.0	One every 5-10 yr	XII	Total damage; waves seen on ground surface, objects thrown in the air

[a]Mercalli numbers are determined by the amount of damage to structures and the degree to which ground motions are felt. These depend on the magnitude of the earthquake, the distance of the observer from the epicenter, and whether an observer is in or out of doors.

Source: After B. Gutenberg, 1950.

EARTHQUAKE PREDICTION

Some of the most dreadful natural disasters have been caused by earthquakes. It is hardly surprising, therefore, that a great deal of research around the world focuses on earthquakes. The hope is that through research we will be able to improve our forecasting ability.

Because China has suffered so many terrible earthquakes, Chinese scientists have tried everything they can think of to predict quakes. They have even observed animal behavior, and on one occasion animals did successfully foretell a quake. On July 18, 1969, zookeepers at the People's Park in Tianjin observed highly unusual animal behavior. Normally quiet pandas screamed, swans refused to go near water, yaks did not eat, and snakes would not go into their holes. The keepers reported their observations to the earthquake prediction office, and at about noon on the same day a 7.4 magnitude earthquake struck.

There have been many informal reports of strange animal behavior before earthquakes, but the Tianjin quake is the only well-documented case. Unfortunately, most quakes do not seem to be preceded by anything odd. While scientists haven't given up on animals, they measure many other factors besides animal behavior.

Most research on earthquake prediction today is based on the properties of elastically strained rocks—properties such as rock magnetism, electrical conductivity, and porosity. Even simple observations, such as the level of water in a well, might indicate a porosity change. Tilting of the ground or slow rises and falls in elevation may also indicate that strain is building up. Most significant are the small cracks and fractures that can develop in severely bent rock. These openings can cause swarms of tiny earthquakes—foreshocks—that may be a clue that a big quake is coming. One of the most successful cases of earthquake prediction, made by Chinese scientists in 1975, was based on slow tilting of the land surface, on fluctuations in the magnetism, and on the swarms of small foreshocks that preceded the 7.3 Richter magnitude quake that struck the town of Haicheng. Half the city was destroyed, but authorities had evacuated more than a million people before the quake. As a result, only a few hundred were killed.

In places where earthquakes are known to occur repeatedly, such as around the margins of the Pacific Ocean, geologists can sometimes discern recurrence patterns. If such a pattern suggests a recurrence interval of, say, a century, it may be possible to predict where and when a large quake may happen. Certainly, it is possible to monitor such areas closely when a big quake is thought to be due. Studies of recurrence patterns have identified a number of *seismic gaps* around the Pacific rim (Fig. 3.13). These are places where, for one reason or another, earthquakes have not occurred for a long time and where elastic strain is steadily increasing. Seismic gaps receive a lot of research attention because they are considered the places most likely to experience large earthquakes. (See the "Guest Essay" at the end of this chapter for a discussion of earthquake prediction efforts in California.)

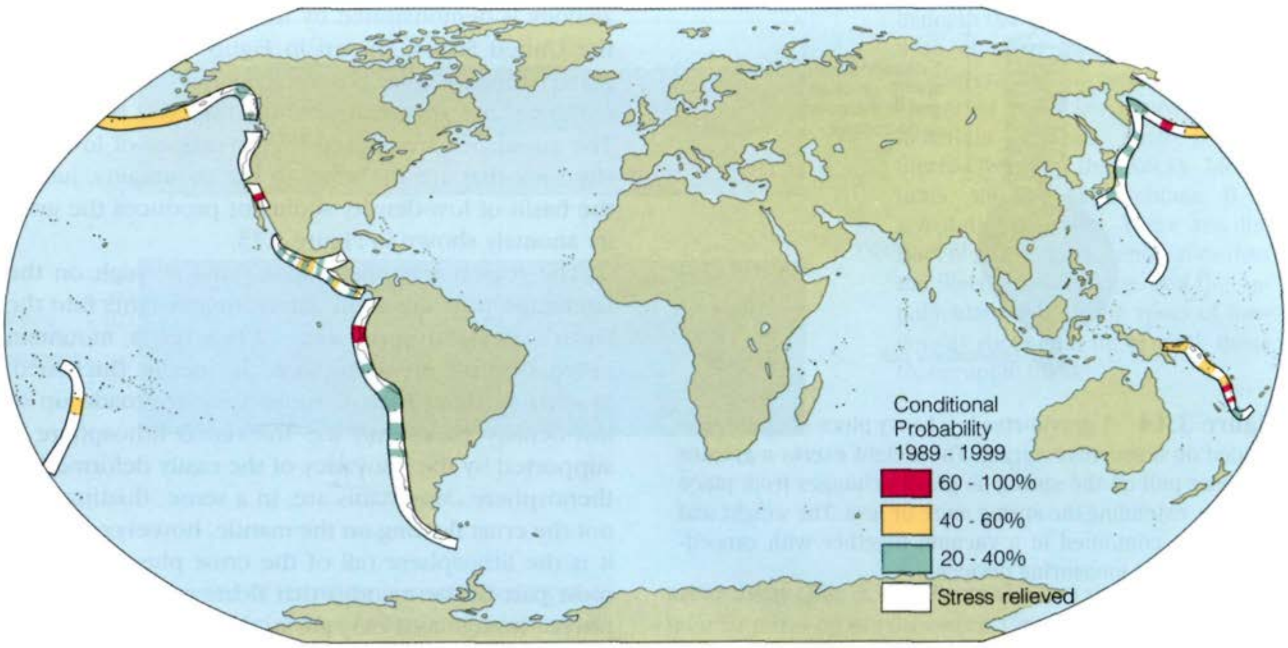


Figure 3.13 Seismic gaps in the circum-Pacific belt. In the areas indicated, earthquakes of magnitude 7.0 or greater are known to have occurred a long time ago, but have not occurred in recent times. Strain is now building up in each seismic gap, raising the probability that a large quake will occur before the year 2000.

GRAVITY ANOMALIES AND ISOSTASY

The Earth *looks* round, but in fact it's not a perfect sphere; careful measurement reveals that it is an ellipsoid that is slightly flattened at the poles and bulged at the equator. The radius at the equator is 21 km (13 mi) longer than at the poles.

Because the gravitational pull between two objects is inversely proportional to the square of the distance between their centers of mass, the pull exerted by the Earth's gravity on a body at the Earth's surface is slightly greater at the poles (because there the body is closer to the center of the Earth) than at the equator. Recall from high school science that your weight is a measure of how strongly the Earth's gravitational force is pulling your body toward the center of the Earth. Thus, a man who weighs 90.5 kg (199.5 lbs) at the North Pole would observe his weight decreasing to 90 kg (198.5 lbs) simply by traveling to the equator. If the weight-conscious traveler made very exact measurements as he traveled, however, he would observe that his weight changed irregularly rather than smoothly. From this irregular change, he could conclude that the pull of gravity must change irregularly. If the traveler went one step further and carried a sensitive device called a *gravimeter* (or *gravity meter)* for measuring the pull of gravity at any locality, he would indeed find an irregular variation. From those irregular variations, a great deal of important information about the interior of the Earth can be deduced.

Gravity Anomalies

A gravimeter, which is similar to an inertial seismograph, consists of a heavy mass suspended by a sensitive spring (Fig. 3.14). When the ground is stable and free from earthquake vibrations, the pull exerted on the spring by the mass provides an accurate measure of the Earth's gravitational pull on the mass. Modern gravimeters are incredibly sensitive. The most accurate devices in operation can measure variations in the force of gravity as tiny as one part in a hundred million.

In order to compare the pull of gravity at different points on the Earth, scientists must correct gravimeter measurements for changes in latitude and topography. The idea behind the corrections is to know the pull of gravity at a constant distance from the center of the Earth. Then, if the rock mass between the

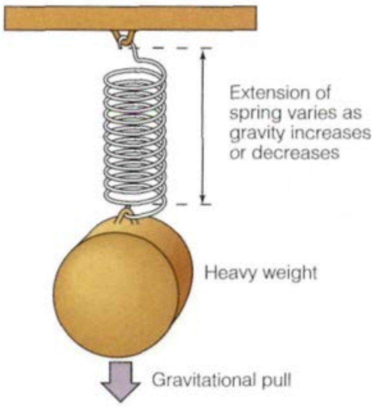


Figure 3.14 A gravimeter is a heavy piece of metal suspended on a sensitive spring. The weight exerts a greater or lesser pull on the spring as gravity changes from place to place, extending the spring more or less. The weight and spring are contained in a vacuum together with exceedingly sensitive measuring devices.

gravimeter and the center of the Earth were everywhere the same, the adjusted figures for the force of gravity might be expected to be the same at every place on the Earth. In fact, the adjusted figures reveal large and significant variations called **gravity anomalies.** Anomalies are due to bodies of rock having different densities, and a great deal of important information can be derived from them. A simple example of an anomaly is shown in Figure 3.15.

The thickness of the crust beneath the United States, as determined from seismic measurements of the moho, is shown in profile in Figure 3.16A. Beneath the three major mountain systems (the Appalachians, the Rockies, and the Sierra Nevada), the crust is thicker than in the nonmountainous regions of the country. The crust beneath the mountains resembles icebergs that have high peaks above the waterline but also massive roots below. The accuracy of this analogy is demonstrated by the gravity profile across the United States, shown in Figure 3.16B. Negative gravity anomalies are observed everywhere across the continent but are greatest where the crust is thickest. The anomalies are caused by the masses of low-density rock that are the roots of the mountains, just as the basin of low-density sediment produces the gravity anomaly shown in Figure 3.15.

The reason why mountains stand so high on the landscape provides some interesting insights into the Earth's physical properties. Wherever a mountain range occurs, the lithophere is locally thickened. Mountains stand high because they are made up of low-density rocks and the thickened lithosphere is supported by the buoyancy of the easily deformed asthenosphere. Mountains are, in a sense, floating. It is not the crust floating on the mantle, however. Rather, it is the lithosphere (all of the crust plus the uppermost part of the mantle) that floats on the asthenosphere. Strange as it may seem, the topographic variations observed at the Earth's surface arise not from the *strength* of the lithosphere but rather from the *weakness* and buoyancy of the asthenosphere.

Isostasy

The flotational balance among segments of the lithosphere is referred to as **isostasy.** The great ice sheets of the last ice age provide an impressive demonstration of isostasy. The weight of a large continental ice sheet, which may be 3 to 4 km (1.9 to 2.5 mi) thick, will depress the lithosphere. When the ice melts, the land surface slowly rises again. The effect is very much like pushing a block of wood into a bucket of thick, viscous oil. When you stop pushing, the wood slowly rises to an equilibrium position determined by

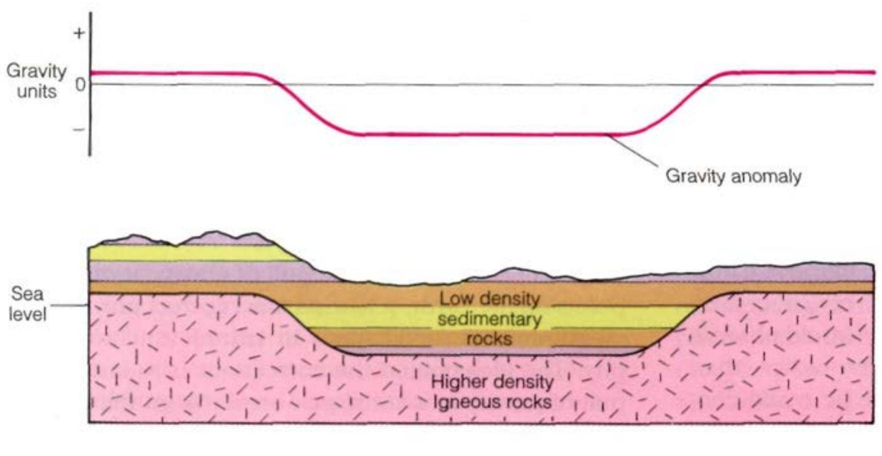


Figure 3.15 A gravity anomaly: a basin filled with low-density sedimentary rocks sitting on a basement of high-density igneous rocks. Gravity measurements reveal a pronounced gravity low throughout the basin. The magnitude of the anomaly can be used to calculate the thickness of the sedimentary rocks.

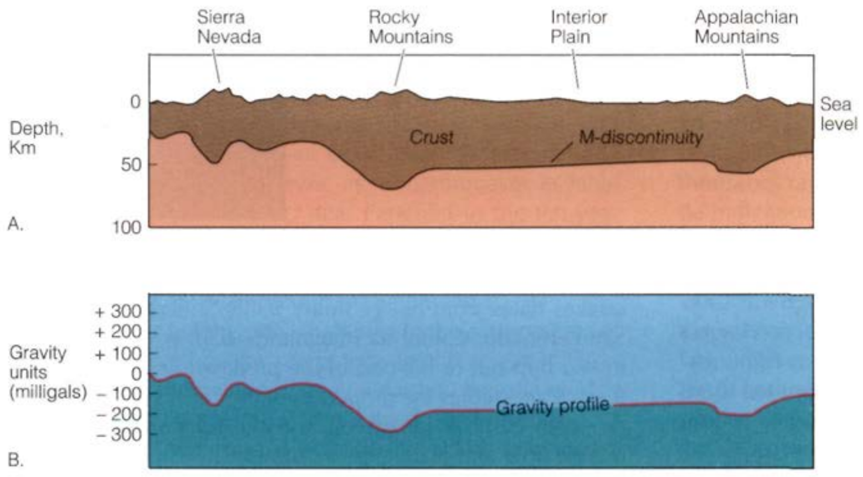


Figure 3.16 Profile of the crust beneath the United States. A. Thickness of crust is determined from measurements of seismic waves. The crust is thicker beneath major mountain masses, such as the Sierra Nevada, the Rocky Mountains, and the Appalachians. B. A gravitational profile. There are distinct negative gravity anomalies over the Sierra, the Rockies, and the Appalachians due to the roots of low-density rocks that lie beneath these topographic highs.

its density. The speed of its rising is controlled by the viscosity of the oil. Glacial depression and rebound of the lithosphere mean that rock in the asthenosphere must flow laterally when the ice depresses the lithosphere, and then must flow back again when the ice melts away (Fig. 3.17). From the fact that the land surface in parts of northeastern Canada and Scandinavia is still rising, even though most of the ice that covered these areas during the last ice age had melted away by 7000 years ago, we infer that the flow must be slow and therefore that the asthenosphere must be extremely viscous.

Continents and the mountains on them are composed of low-density rock, and they stand high because they are thick and light; ocean basins are topographically low because the oceanic crust is composed of denser rock. Isostasy and the fact that the continental crust is less dense than the oceanic crust are the reason the Earth has continents and ocean basins.

The important point to be drawn from this discussion of isostasy is that the lithosphere acts as if it were "floating" on the asthenosphere. *(Floating* is not the most precise word because the Earth is solid, but the lithosphere is buoyant and acts as though it were floating.) Sometimes gravity measurements suggest that a mountain has been pushed up so rapidly that it is top-

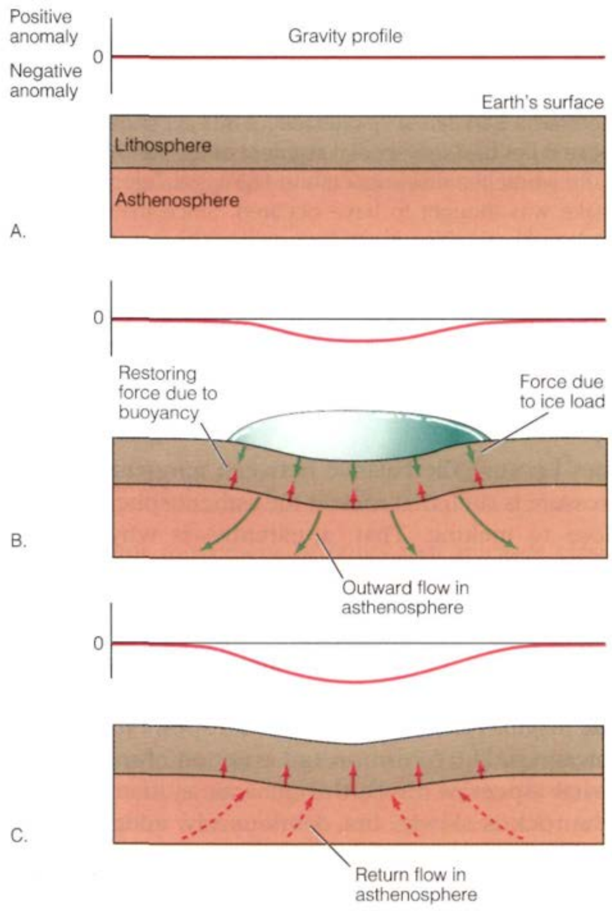


Figure 3.17 Depression of the lithosphere and asthenosphere by a continental ice sheet. A. Prior to formation of the ice sheet, there is no gravity anomaly. B. When the ice sheet forms, it depresses the lithosphere and the asthenosphere. At some depth in the asthenosphere, material must slowly flow outward to accommodate the sagging lithosphere. C. When the ice melts, buoyancy slowly restores the lithosphere and asthenosphere to their original levels. A negative gravity anomaly continues until the depression is removed. The viscosity of the asthenosphere controls the rate of flow and therefore the rate of recovery.

Guest Essay

Rethinking Earthquake Prediction

Jeffrey Park is an associate professor of geology and geophysics at Yale University. His interest in earthquakes dates from the February 9, 1971 earthquake in San Fernando, California (magnitude 6.5), which nearly tossed him out of his bed in the predawn twilight. Professor Park studies the properties of the Earth's deep interior with earthquake waves, and the relation between mantle flow and plate tectonic motions.

Perhaps seismologists were unlucky that the great 1906 San Francisco earthquake was one of the first large seismic events to be studied thoroughly. Although the 1906 quake was probably a typical example of the largest earthquakes that occur along the San Andreas Fault, residents in California may face greater risks from the more numerous smaller earthquakes, whose behavior is only now becoming understood.

For most of this century, earthquake prediction has depended on the elastic rebound theory developed by H. F. Reid. Based largely on field studies of ground motion along the San Andreas Fault after the great 1906 earthquake, Reid's theory holds that strain gradually builds in the rock surrounding a fault zone until a "threshold strain" is reached, the rock fails, and earthquake rupture occurs. In the 1960s plate tectonics theory allowed geophysicists to predict the rate at which relative motion occurs along plate boundaries. If the threshold-strain idea of the elastic rebound theory is correct, one can predict the timing of future earthquakes from the recurrence time of past events.

The San Andreas Fault system divides the North American Plate from the Pacific Plate, extends nearly the entire length of California, and is positioned to pose a serious hazard to the urban areas of San Francisco and Los Angeles. It came as no surprise in the 1970s that trenching experiments along the San Andreas Fault suggested that there was a characteristic interval (125 to 225 years) between Magnitude 8+ earthquakes along the southern part of the fault near Los Angeles. Because the last such earthquake near Los Angeles occured in 1857 near Fort Tejon, seismologists anticipate the next "Big One" within the next century. Along other major plate boundaries where no large earthquake had occured within the previous few decades, seismologists developed medium-term earthquake "forecasts" for these seismic gaps.

Still, this does not offer us "short-term" earthquake prediction, which would warn residents and public officials that a damaging earthquake is likely to occur in years or months, not decades. In an ambitious short-term earthquake prediction experiment, a large concentration of geophysical instruments was installed near Parkfield, California, to measure ground motion, tilt, strain, water level, and other suspected earthquake precursors, or phenomena that might help predict an earthquake in the short term. Parkfield was selected for this experiment because it lies beside a special segment of the San Andreas Fault, where the first break of the 1857 Fort Tejon earthquake was thought to have occured. Since that event, earthquakes of magnitude 6 or so have occured in the Parkfield segment in 1881, 1901, 1922, 1934, and 1966,

heavy and has too little root of low-density rock to counterbalance its upper mass. Sometimes it is observed that low-density crust has been dragged down so rapidly that it forms a root without a mountain mass above it. These and many other situations lead to local gravity anomalies. That the anomalies do not become very large suggests that the Earth is always moving toward an isostatic balance. Indeed, isostasy is the principal explanation for vertical motions of the Earth's surface, just as plate tectonics is the principal explanation for lateral motions.

The importance of the asthenosphere in determining the shape of the Earth's surface cannot be overemphasized. The asthenosphere has the properties it does because the balance between temperature and pressure is such that rock in the asthenosphere is very close to melting. That, apparently, is why the asthenosphere is so weak and why seismic wave speeds in the asthenosphere are low. The asthenosphere is also vitally important for another reason—wherever temperatures exceed rock-melting temperatures, magma (molten rock) is formed in the asthenosphere. The magma rises to the surface and spews forth from volcanoes. The formation and eruption of magma are a vital aspect of the Earth's dynamic system because new rock is slowly, but continuously, added to the Earth's surface from deep in the interior.

suggesting a regular recurrence time of 20 to 25 years. Seismologists extrapolated the sequence to forecast another magnitude 6 earthquake in 1988, with an estimated five-year margin of error. Nature's repeatability failed in this case, however, as no earthquakes as large as magnitude 5 occured near Parkfield in the ten-year period that ended December 30, 1992, when the Parkfield "prediction window" closed. Elastic rebound theory, at least in its simplest form, had failed.

This was not the only challenge to the assumptions of would-be earthquake predictors. In the ten-year Parkfield prediction window, several earthquakes occured in central and Southern California that forced seismologists to rethink their hazard assessments. Plate boundaries within continents are rarely sharp, and the associated deformation can be spread in a zone more than a hundred kilometers wide. As part of this complicated deformation, geologists estimate that the Los Angeles basin, south of the San Andreas Fault, is being compressed by 7 mm per year. This roughly north-south shortening leads to motion along thrust faults within the basin, which gives rise to the east-west-trending ranges of hills that partition the Los Angeles metropolitan area. In 1987 a magnitude 5.9 earthquake struck Whittier, California, within the Los Angeles metropolitan area and far south of the San Andreas Fault, along a thrust fault more than 10 km beneath the Montebello Hills. Serious property damage was caused in a localized region near the fault. In 1983 a similar, but larger (magnitude 6.5), earthquake devastated the town of Coalinga in central California. On January 1 7, 1994, yet another such earthquake (magnitude 6.6) ruptured under Northridge, a Los Angeles suburb, causing property damage estimated at over $15 billion.

That type of faulting in these events raised concern. Rupture occured along "blind-thrust" faults, where a dipping fault surface lies within deeply buried sediments. Such faults have no expression on the surface, aside from an anticlinal upwarp of the sedimentary layers above the fault. Blind thrusts have estimated recurrence intervals of thousands rather than hundreds of years, so they give little indication of their destructive potential in the hisorical record. In fact, the concept of recurrence interval may be inappropriate for blind-thrust earthquakes, for faulting on these geological structures may depend on the interaction of many tectonic movements within the Los Angeles basin. Scientists are now mapping blind-thrust faults in the basin with equipment and techniques similar to those used to discover petroleum and natural gas deposits.

On June 28, 1992, the San Andreas Fault was upstaged by a magnitude 7.3 earthquake near Landers, California, nominally within the North American Plate in the Mojave Desert. The largest earthquake in California since the 1906 San Francisco quake, the Landers quake ruptured a string of surface faults that previously were not thought to be linked. One dramatic result from the Landers event is the precise measurement of its total surface motion using Global Positioning System (GPS) satellite receivers. Under favorable conditions, a GPS receiver can determine its absolute position with an accuracy better than a centimeter by referencing itself to Earth-orbiting satellites. GPS-receiver networks in principle can measure the strain buildup and release associated with major earthquakes. If large-scale flexing of the crust along a fault zone occurs prior to large earthquakes, these types of measurements may provide seismologists with early warning signs. Seismologists hope to understand the earthquake process better by combining these new types of observations with traditional earthquake studies, thereby making short-term earthquake prediction more realistic.

Summary

1. Abrupt movement of faults that releases elastically stored energy is thought to cause earthquakes: this is known as the elastic rebound theory.
2. Earthquake vibrations are called seismic waves and are measured with seismographs.
3. Energy released at an earthquake's focus radiates outward as two kinds of body waves: P waves (Primary waves, which are compressional) and S waves (Secondary waves, which are shear waves). Earthquake energy also causes the surface of the Earth to vibrate because of surface waves.
4. From the study of seismic-wave refraction and reflection, scientists infer the internal structure of the Earth by locating discontinuities in its composition and physical properties. A pronounced compositional boundary occurs between the mantle and the outer core.
.5. At the mantle-crust interface is a pronounced seismic discontinuity called the Mohorovicic discontinuity. Crust thickness ranges from 30 to 70

km in continental regions but is only 8 km beneath oceans.

6. The core has a high density and is inferred to consist of iron as well as small amounts of nickel and other elements. The outer core must be molten because it does not transmit S waves. The inner core is solid.
7. From a depth of 100 km to 350 km is a zone of low seismic-wave speed. This low-speed zone is the asthenosphere. The lithosphere, which is rigid and on average 100 km thick, overlies the asthenosphere and consists of the upper part of the mantle and all of the crust. Below the asthenosphere, in the mesosphere, is a zone of higher seismic-wave speed than in the asthenosphere.
8. The amount of energy released during an earthquake is calculated on the Richter magnitude scale.
9. Earthquakes cause damage in six different ways: by ground motion; by faulting; by fires; by land movement and slope collapse; by liquefaction; and by tsunami.
10. The outer portions of the Earth are in approximate isostatic balance; in other words, like huge icebergs floating in water, the lithosphere "floats" on the asthenosphere.
11. When parts of the lithosphere are not in flotational equilibrium, gravity anomalies occur.

Important Terms to Remember

body wave (p. 73)
earthquake focus (p. 73)
elastic deformation (p. 71)
elastic rebound theory (p. 72)
epicenter (p. 73)
fault (p. 72)
gravity anomaly (p. 84)
isostasy (p. 84)
M-discontinuity (p. 77)
modified Mercalli scale (p. 81)
moho (p. 77)
Mohorovicic discontinuity (p. 77)
P (primary) wave (p. 75)
Richter magnitude scale (p. 78)
S (secondary) wave (p. 75)
seismic sea wave (p. 81)
seismic wave (p. 73)
surface wave (p. 73)
tsunami (p. 81)

Questions for Review

1. Explain how most earthquakes are thought to occur and why there seems to be a limit on earthquake magnitude.
2. What is the relation between an earthquake focus and the corresponding epicenter?
3. How are seismic waves recorded and measured? How would you locate an epicenter from seismic records? Explain how a focus depth is determined.
4. What are the differences between seismic body waves and surface waves? Identify two kinds of body waves and explain the differences.
5. Earthquakes can cause damage in many ways; name four. Where on the Earth was the most disastrous earthquake on record and how did the people die? Where was the biggest known earthquake in the United States?
6. What are reflection and refraction, and how do they affect the passage of seismic waves? How can refraction and reflection be used to define the mantle-crust boundary? The core-mantle boundary?
7. Briefly describe how seismic waves can be used to infer that the outer core is molten while the inner core is solid. What evidence indicates that the core is made largely of metallic iron?
8. Describe the Earth's three compositional layers.
9. The Earth is layered with respect to rock strength into five zones. Name and describe these zones.

10. What is the relationship between the crust, the mantle, and the lithosphere?
11. Why are ocean basins low spots on the Earth's surface and continents high places?
12. How do gravity anomalies arise and how can they be measured?
13. Describe some evidence that proves that isostasy is operating in the Earth. How is the Earth's surface topography related to isostasy?
14. Draw an east-west profile of the crust under the United States and indicate how isostasy plays a role in what you have drawn.

Questions for A Closer Look

1. How much energy can be stored in a cubic meter of elastically strained rock? How much energy can be released during a single big earthquake event?
2. Explain how seismologists use the Richter magnitude scale to estimate the energy released during an earthquake.
3. Why is the Richter magnitude scale logarithmic, and how big are the steps between magnitudes?

Questions for Discussion

1. The lithosphere of the Moon is about 1000 km thick and the asthenosphere only about 380 km thick. Would you expect isostasy to operate on the Moon? Could your hypothesis be tested by measuring gravity anomalies with an orbiting spaceship? Might gravity anomalies be something to measure if a spaceship were sent to investigate planets around some other sun?
2. Research the current work being done on earthquake prediction. How closely in time would a prediction of an earthquake have to be in order to be useful—a few hours; a few days; a few weeks? Be sure to explain your reasoning.

CHAPTER 4

Minerals and Rocks

Carajas iron mine in the tropical rain forest of northern Brazil, one of the world's largest and richest deposits of iron ore. When mining is finished, the hill of ore will be a great hole in the ground.

Minerals: A Linchpin of Society

Most minerals that are abundant in the Earth's crust have neither commercial value nor any particular use. Those few minerals that are the raw materials of industry, which we call *ore minerals,* tend to be rare and hard to find—gold, for instance, or sphalerite, the main zinc mineral. From the ore minerals we get the metals to make our machines and the ingredients for chemicals and fertilizers. Our modern society is totally dependent on an adequate supply of ore minerals. Without them we could not build planes, cars, televisions, or computers. Industry would falter and living standards would decline.

Can the ore minerals in the Earth's crust sustain both a growing population and a high standard of living for everyone? This difficult question has many experts worried. The minerals they worry most about are those used as sources of such important metals as lead, zinc, and copper. Metals, the experts point out, begin the chain of resource use. Without metals, we cannot make machines. Without machines, we cannot convert the chemical energy of coal and oil to useful mechanical energy. Without mechanical energy, the tractors that pull plows must grind to a halt; trains and trucks must stop running, and indeed our whole industrial complex must become still and silent.

Experts cannot tell how long the Earth's supplies of ore minerals will last because there is no way to "see" inside the Earth and know exactly what is there. Optimists point to the great success our technological society has enjoyed over the past two centuries as ever more remarkable discoveries have been made. Improved prospecting will keep up the success story, they insist. However, if mineral supplies do become limited, the experts suggest we will find ways to get around the limits by recycling, by substituting, and by discovering new technologies.

Many scientists have a more pessimistic opinion. Technologically advanced societies have faced mineral resource limits in the past, they point out, but the solution has always been to import new supplies rather than develop substitutes or effective recycling measures. England, for instance, was once a great supplier of metals. Today, its minerals are mined out, most of its mines are closed, and English industry runs on raw materials imported from abroad. The United States, too, was once self-sufficient in most minerals and an exporter of many. Slowly, the United States has become a net importer and now relies on supplies from such countries as Australia, Chile, South Africa, and Canada. Today no large industrial country can supply its own mineral needs. The only country that might be able to do so is Russia. But eventually the Russian mines, too, will be depleted, and so, too, will the mines of Australia and other countries. Where then does society turn?

The answer to this question is not obvious, but it must be found in the foreseeable future. It is highly likely that, within the lifetimes of the people who read this book, mineral limitations will occur. Which minerals, and therefore which metals, will be in short supply first is still an open question. How society will cope and respond, and when it will have to do so, are just two of the great social and scientific issues still to be solved.

MINERALS AND THEIR CHEMISTRY

The word *mineral* means different things to different people, but for scientists who study the Earth it has a very specific connotation. A **mineral** is any naturally formed, solid chemical element or chemical compound having a definite composition and a characteristic crystal structure. To take just one example, diamond is a mineral. It is naturally formed, it is a solid, it is made of the chemical element carbon, and all the carbon atoms are packed together in a regular and characteristic geometric array called the crystal structure of diamond.

Coal resembles diamond in that it is largely carbon, but coal is not a mineral; it is a rock. In addition to its carbon, coal contains many chemical compounds, and its composition varies from sample to sample so that it does not have a specific composition. Nor does coal have a characteristic crystal structure.

Rocks are aggregates of minerals: they are nature's books, and in them can be read the story of the way the Earth works: how continents move, how mountains form and then erode away. Minerals are the words used in nature's books. Minerals and rocks go together naturally, but before we can study rocks, we need to know something about minerals. The easiest way to introduce minerals is by examining their two most important characteristics:

1. Composition—the kinds of chemical elements present and their proportions.
2. Crystal structure—the way in which the atoms of the elements are packed together in a mineral.

Because most minerals contain several chemical elements, we begin our discussion by reviewing the way in which elements combine to form compounds.

Elements and Atoms

Chemical elements are the most fundamental substances into which matter can be separated by chemical means. For example, table salt is not an element because it can be separated into sodium and chlorine. Neither sodium nor chlorine can be further broken down chemically, however, and so each is an element.

Each element is identified by a symbol, such as H for hydrogen and Si for silicon. Some symbols, such as that for hydrogen, come from the element's English name. Other symbols come from other languages. For example, iron is Fe from the Latin *ferrum,* and sodium is Na from the Latin *natrium.* The 88 naturally occurring elements and their symbols are listed in Appendix B.

Even the tiniest piece of a pure element consists of a vast number of identical particles of that element called atoms. An **atom** is the smallest individual particle that retains all the properties of a given element. Atoms are so tiny they can be seen only by using the most powerful microscopes ever invented, and even then the image is imperfect because individual atoms are only about 10^{-10} m in diameter.

Atoms are built up from *protons* (which have positive electrical charges), *neutrons* (which, as their name suggests, are electrically neutral), and *electrons* (which have negative electrical charges that balance exactly the positive charges of protons). Protons and neutrons join together to form the core, or *nucleus,* of an atom. Electrons are much smaller than protons or neutrons; in an atom, they move in a distant and diffuse cloud around the nucleus (Fig. 4.1).

Protons give a nucleus a positive charge, and the number of protons in the nucleus of an atom is called the *atomic number* of the atom. The number of protons in the nucleus (in other words, the atomic number) is what gives the atom its special characteristics and what makes it a specific element. Thus, any and all atoms containing one proton in the nucleus are atoms of hydrogen; atoms containing two protons in the nucleus are helium; and so on. All atoms having the same atomic number are atoms of the same element. The atomic numbers of all naturally occurring elements are listed in Appendix B.

Because neutrons are electrically neutral, they cannot change the atomic number of an element. Neutrons can change the mass of an atom, however, and the sum of the neutrons plus protons in the nucleus of an atom is the *mass number.* As we learned in Chapter 2, the number of protons in the nucleus (the atomic number) is designated by a subscript before the chemical symbol, while the sum of the protons plus neutrons (the mass number) is indicated by a superscript: helium is written ${}^{4}_{2}He$. Most elements have several **isotopes;** these are atoms with the same atomic number and hence the same chemical properties, but different mass numbers. Carbon, for example, has three naturally occurring isotopes: ${}^{12}_{6}C$, ${}^{13}_{6}C$, and ${}^{14}_{6}C$. Some isotopes are radioactive, which means they transform spontaneously to another isotope of the same element or an isotope of a different element. Among the carbon isotopes only ${}^{14}_{6}C$ is radioactive, and it transforms to an isotope of nitrogen, ${}^{14}_{7}N$, by the spontaneous transformation of a neutron to a proton. There are 25 naturally occurring radioactive isotopes.

Energy-Level Shells

Electrons are confined to specific shells that are arranged at predetermined distances from the nucleus. Because the electrons in each shell have a specific amount of energy characteristic for that shell, the shell distances are commonly called **energy-level shells.** The maximum number of electrons that can occupy a given energy-level shell is fixed. As shown in Figure 4.1, shell 1, closest to the nucleus, is small and can accommodate only 2 electrons; shell 2, however, can accommodate up to 8 electrons; shell 3, 18; and shell 4, 32.

Ions

An energy-level shell filled with electrons is very stable; it is comparable to an evenly loaded boat. To fill their outermost energy-level shell and so reach a stable configuration, atoms either share or transfer electrons among themselves. In its natural state, an atom is electrically neutral because the positive electrical charge of its protons is exactly balanced by the negative electrical charge of its orbiting electrons. When an electron is transferred as part of the stabilizing of energy-level shells, this balance of electrical forces is upset. An atom that loses an electron has lost a negative electrical charge and is left with a net positive charge. An atom that gains an electron has a net negative charge. An atom that has excess positive or negative charges caused by electron transfer is called an **ion.** When the excess charge is positive (meaning that the atom gives up electrons), the ion is called a **cation;** when negative (meaning an atom adds electrons), the ion is an **anion.**

The most convenient way to indicate ionic charges is with superscripts. For example, Ca^{2+} is a cation (calcium) that has given up two electrons, and F^- is an anion (fluorine) that has accepted an electron.[1] Because the formation of ions involves only the energy-level shell electrons and not the protons and neutrons in the nucleus, it is common practice to omit the atomic and mass number symbols for reactions involving ions.

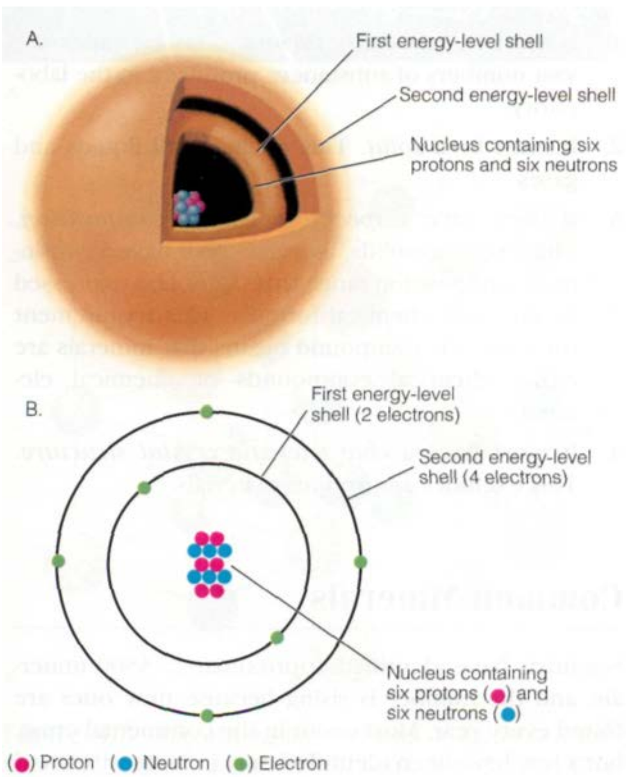


Figure 4.1 Schematic diagram of an atom of the element carbon. The nucleus contains six protons and six neutrons. Electrons orbiting the nucleus are confined to specific orbits called energy-level shells. A. Three-dimensional representation showing the first two shells. The first shell can contain two electrons, the second eight. B. Two-dimensional representation of the carbon atom to show the number of protons and neutrons in the nucleus and the number of electrons in the energy-level shells. The first energy-level shell is full because it contains two electrons. The second shell contains four electrons and so is half full.

Compounds

Chemical compounds form when one or more anions combine with one or more cations in a specific ratio. For example, two cations of H^+ combine with one anion of O^2 to make the compound H_2O (water). In a compound, the sum of the positive and negative charges must be zero.

The formula of a compound is written by putting the cations first and the anions second. The numbers of cations or anions are indicated by subscripts, and for convenience the charges of the ions are usually omitted. Thus, we write H_2O rather than $H_2{}^+O^2$-.

An example of the way electron transfer leads to formation of a compound is shown in Figure 4.2 for the elements lithium and fluorine. A lithium atom has energy-level shell 1 filled by two electrons but has only one electron in shell 2, even though shell 2 can accommodate eight electrons. The lone outer electron in shell 2 can easily be transferred to an element such as fluorine, which already has seven electrons in shell 2 and needs only one more to be completely filled. In this fashion, both a lithium cation and a fluorine anion end up with filled shells, and the resulting positive charge on the lithium and negative charge on the fluorine draw, or bond, the two ions together.

[1] Note that when the ionic charge Is 1, we omit the number. The symbol F^- means F^{1-}, and Li^+ means Li^{1+}. Most atoms are present in the Earth as ions rather than as electrically neutral atoms.

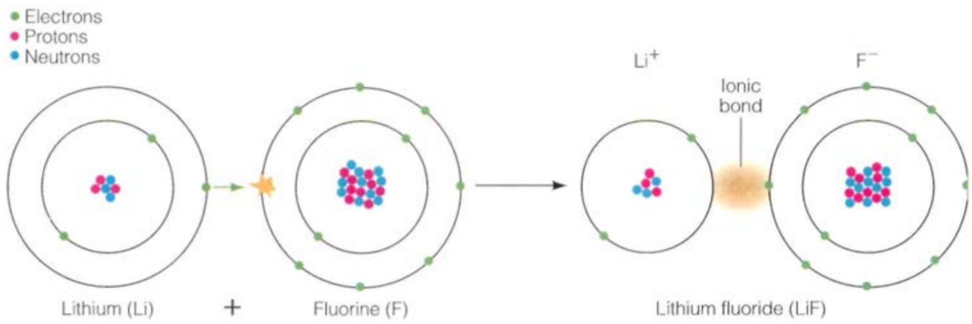


Figure 4.2 To form the compound lithium fluoride, an atom of the element lithium combines with an atom of the element fluorine. The lithium atom transfers its lone outer-shell electron to fill the fluorine atom's outer shell, creating an Li^+ cation and a F^- anion in the process. The electrostatic force that keeps the lithium and fluorine ions together in the compound lithium fluoride is an ionic bond..

Lithium and fluorine form the compound lithium fluoride, which is written LiF to indicate that for every Li ion there is one F ion. The combination of one Li ion and one F ion is called a molecule of lithium fluoride. A *molecule* is the smallest unit that retains all the properties of a compound. The properties of molecules are quite different from the properties of their constituent elements. The elements sodium (Na) and chlorine (Cl) are highly dangerous, for example, but the compound sodium chloride (NaCl, table salt) is essential for human health.

Complex Ions

Sometimes two kinds of ions form such strong bonds with each other that the combined ions act as if they were a single ion. Such a strongly bonded pair is called a *complex ion.* Complex ions act in the same way as single ions, forming compounds by bonding with other ions of opposite charge. For example, carbon and oxygen combine to form the complex carbonate anion CO_3^{2-} . The carbonate anion then bonds with cations such as Na^+ and Ca^{2+} to form compounds such as Na_2CO_3 (sodium carbonate) and $CaCO_3$ (calcium carbonate). Other important complex ions are the sulfate SO_4^{2-} , nitrate NO_3^- and silicate SiO_4^{4-} anions.

Crystal Structure

The ions in most solids are organized in the regular, geometric patterns of a crystal structure, like eggs in a carton, as shown in Figure 4.3. Solids that have such a crystal structure are said to be *crystalline,* whereas solids that lack a crystal structure are *amorphous* (Greek for without form). Glass and amber are examples of amorphous solids. All minerals are crystalline, and the crystal structure of a mineral is a unique property of that mineral. All specimens of a given mineral have identical crystal structure.

Before proceeding, let's review what is meant by the term *mineral.* To be called a mineral, a substance must meet four requirements.

1. It must be *naturally formed.* This excludes the vast numbers of substances produced in the laboratory.
2. It must be a *solid.* This excludes all liquids and gases.
3. It must have a *specific chemical composition.* This excludes solids, like glass, that have a continuous composition range that cannot be expressed by an exact chemical formula. This requirement for a specific compound means that minerals are either chemical compounds or chemical elements.
4. It must have a *characteristic crystal structure.* This excludes amorphous materials.

Common Minerals

Scientists have identified approximately 3500 minerals, and the number is rising because new ones are found every year. Most occur in the continental crust, but a few have been identified only in meteorites, and two new ones were discovered in the Moon rocks brought back by the astronauts. The total number of minerals may seem large, but it is tiny compared with the astronomically large number of ways a chemist can combine naturally occurring elements to form compounds. The reason for the disparity between nature and chemical experiment becomes apparent when we consider the relative abundances of the chemical elements. As Table 4.1 shows, only 12 elements occur in the continental crust in amounts equal to or greater than 0.1 percent. Together, these 12—

usually referred to as the abundant elements, with all others called the scarce elements—make up 99.23 percent of the continental crust mass. Therefore, the continental crust is constructed of 40 or 50 minerals, most of which contain one or more of the 12 abundant elements.

Minerals containing scarce elements certainly do occur, but only in small amounts, and those small amounts form only under special and restricted circumstances. A few scarce elements, such as hafnium and rhenium, are so rare that they are not known to form minerals under any circumstances; they occur only as trace impurities in common minerals.

As Table 4.1 shows, two elements, oxygen and silicon, make up more than 70 percent of the continental crust. Oxygen forms a simple anion, O^{2-}, and compounds that contain the O^{2-} anion are called oxides. Silicon forms a simple cation, Si^{4+}, and oxygen and silicon together form a strong complex ion, the **silicate anion $(SiO_4)^{4-}$.** Minerals that contain the silicate anion are complex oxides, and to distinguish them from simple oxides they are called **silicates.** The compound MgO is an oxide, but Mg_2SiO_4 is a silicate.

Silicates are the most abundant of all minerals, and simple oxides are the second most abundant group. Other mineral groups, all of them important but all less common than silicates and oxides, are sulfides, which contain the simple anion S^{2} ; carbonates $(CO_3)^{2-}$; sulfates $(SO_4)^{2-}$; and phosphates $(PO_4)^{3-}$.

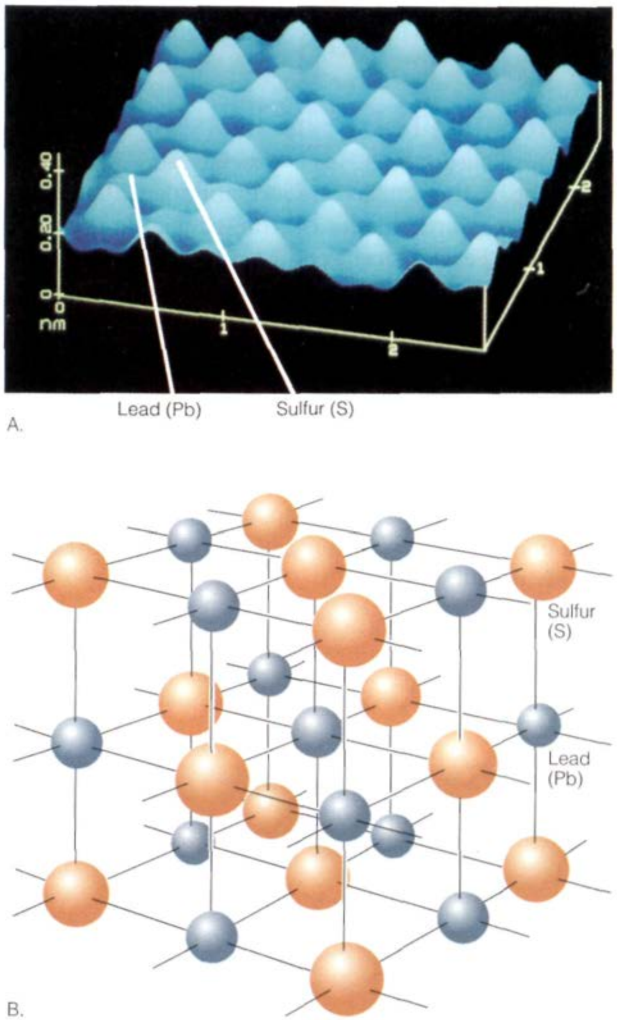


Figure 4.3 The arrangement of ions in the most common lead mineral, galena (PbS). Lead, the Pb part, is a cation with a charge of 2+, and sulfur, S, is an anion with a charge of 2-. To maintain a charge balance between the ions, there must be an equal number of Pb and S ions in the structure. Ions are so small that a cube of galena 1 cm on its edge contains 10^{22} ions each of lead and sulfur. A. Ions at the surface of a galena crystal revealed with a scanning-tunneling microscope. Sulfur ions are the large lumps, lead the smaller ones. B. The packing arrangement of ions is repeated continuously through a crystal. The ions are shown pulled apart along the black lines to demonstrate how they fit together.

The Silicates

The silicate anion $(SiO_4)^{4-}$ has the shape of a tetrahedron. The four relatively large oxygen anions surround and bond to the much smaller silicon cation as shown in Figure 4.4. All silicates contain the silicate anion as an integral part of the crystal structure. In many silicates, however, the anions actually join together by sharing their oxygens and so form chains, sheets, and three-dimensional networks of tetrahedra. (The process is called polymerization.) How this is done is shown in Figures 4.5 and 4.6. Polymerization plays a major role in determining the properties of silicates.

Table 4.1
The Most Abundant Chemical Elements in the Continental Crust

Element	Ion	Percent by weight
Oxygen (O)	O^{2-}	45.20
Silicon (Si)	Si^{4+}	27.20
Aluminum (Al)	Al^{3+}	8.00
Iron (Fe)	Fe^{2+} and Fe^{3+}	5.80
Calcium (Ca)	Ca2+	5.06
Magnesium (Mg)	Mg2+	2.77
Sodium (Na)	Na^{+}	2.32
Potassium (K)	K^{+}	1.68
Titanium (Ti)	Ti^{4+}	0.86
Hydrogen (H)	H^{+}	0.14
Manganese (Mn)	Mn^{2+} and Mn^{4+}	0.10
Phosphorus (P)	P^{3+}	0.10
All other elements		0.77
Total		100.00

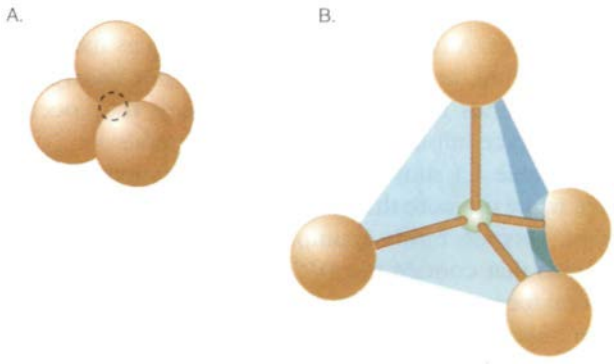


Figure 4.4 The tetrahedron-shaped silicate anion $(SiO_4)^{4-}$. A. Anion with the four oxygens touching each other in natural position. Silicon (dashed circle) occupies central space. B. Exploded view showing the relatively large oxygen anions at the four corners of the tetrahedron, equidistant from the relatively small silicon cation.

The silicates are by far the most abundant minerals in the continental crust, and among them the feldspars are the predominant variety—approximately 60 percent of all minerals in the Earth's crust are feldspars (Fig. 4.7). Indeed, the very name reflects how common feldspars are. The name is derived from two Swedish words, *feld* (field) and *spar* (mineral). Early Swedish miners were familiar with feldspar in their mines; they were also farmers, and they found the same minerals in the rocks they had to clear from their fields before they could plant crops. Struck by the abundance of feldspar, the miners chose a name to indicate that their fields seemed to be growing an endless crop of the minerals.

The second most abundant mineral in the crust is the silicate called quartz. Feldspar and quartz together account for 75 percent of the continental crust. All the silicates added together make up 95 percent or more of both the continental crust and the oceanic crust, and an even larger percentage of the mantle.

The Nonsilicates

The nonsilicate minerals are widespread and may at first sight be thought to be more abundant than they actually are. Three oxides of iron—hematite (Fe_2O_3), magnetite (Fe_3O_4), and goethite ($FeO{\cdot}OH$)—are estimated to be the most abundant nonsilicates. Other important nonsilicate mineral groups are the carbonates calcite ($CaCO_3$) and dolomite ($CaMg(CO_3)_2$), the sulfate gypsum ($CaSO_42H_2O$), and the sulfides pyrite (FeS_2), sphalerite (ZnS), galena (PbS), and chalcopyrite ($CuFeS_2$). Many of the less common nonsilicates are the minerals miners seek for the production of metals such as gold, silver, iron, copper, and zinc. For more information about minerals, see "A Closer Look: Identifying Minerals."

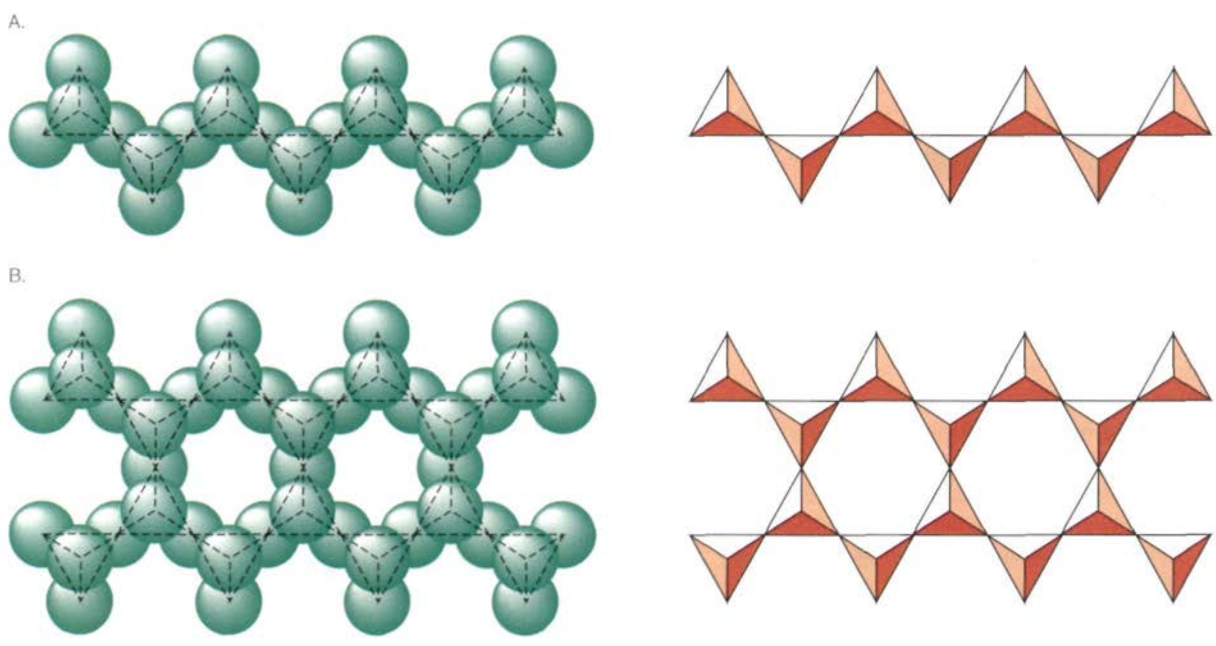


Figure 4.5 Polymerization of complex silicate anions. A. A polymer chain in which each silicate anion shares two of its oxygens with adjacent anions. A geometric representation of the chain is on the right. The formula of each basic unit in the chain is $(SiO0_3)^{2-}$. B. Double polymer chain for which the formula of the basic unit is $(Si_4O_{11})^{6-}$.

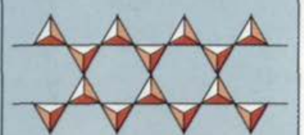
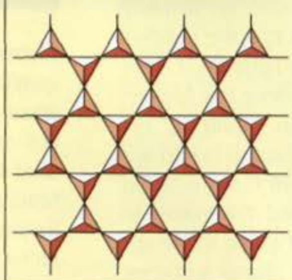
Mineral	Formula	Cleavage	Structure
Olivine	$(Mg,Fe)_2SiO_4$	None	Isolated tetrahedra 
Garnet	$Mg_3Al_2(SiO_4)_3$	None	
Pyroxene	$CaMg(SiO_3)_2$	Two planes at 90°	Chain of tetrahedra 
Amphibole	$Ca_2Mg_5(Si_4O_{11})_2(OH)_2$	Two planes at 120°	Double chain of tetrahedra 
Mica	$KAl_2(Si_3Al)O_{10}(OH)_2$	One plane	Sheet of tetrahedra 
Clay	$Al_4Si_4O_{10}(OH)_8$		
Feldspar	$KAlSi_3O_8$	Two planes at 90°	Three dimensional network too complex to be shown by a two dimensional drawing
Quartz	SiO_2	None	

Figure 4.6 Summary of the way silicate anions polymerize to form the common silicate minerals. The most important polymerizations are those that produce chains, sheets, and three-dimensional networks. Note the relationship between crystal structure and cleavage. (Cleavage is discussed in "A Closer Look: Identifying Minerals.")

Figure 4.7 The two most common minerals in the Earth's crust. Crystals of feldspar (green) and quartz (gray) from Pikes Peak, Colorado. This specimen is 20 cm across.

A Closer Look

Identifying Minerals

Mineral properties are determined by composition and crystal structure. It is not necessary, however, to analyze a mineral for its chemical composition or determine its crystal structure in order to discover its identity. Once we know which properties are characteristic of which minerals, we can use those properties to identify the minerals. The properties most often used to identify minerals are crystal form, growth habit, cleavage, luster, color, hardness, and specific gravity. Appendix C lists the properties of common minerals.

Crystal Form and Growth Habit

Ice fascinated the ancient Greeks. When they saw glistening needles of ice covering the ground on a frosty morning, they were intrigued by the fact that the needles were six-sided and had smooth, planar surfaces. Greek philosophers made many discoveries about the branch of mathematics called geometry, but they could not explain how three-dimensional, geometric solids could apparently grow spontaneously. The ancient Greeks called ice *krystallos,* and the Romans latinized the name to *crystallum.* Eventually, the word **crystal** came to be applied to any solid body that grows with planar surfaces. The planar surfaces that bound a crystal are called **crystal faces,** and the geometric arrangement of crystal faces, called the **crystal form,*** became the subject of intense study during the seventeenth century.

Seventeenth-century scientists discovered that crystal form could be used to identify minerals, but some aspects of crystal form were difficult for them to explain. Why, for example, did the size of crystal faces differ from sample to sample. Under some circumstances, a mineral may grow as a thin crystal; in other cases, the same mineral may grow as a fat one, as Figure C4.1 shows. Superficially, the two crystals of quartz in Figure C4.1 look very different, and this photograph illustrates that neither crystal size nor crystal face size is a unique property of a mineral.

The person who solved the mystery was a Danish physician, Nicolaus Steno. In 1669 Steno demonstrated that the unique property of crystals of a given mineral is not the relative face sizes, but rather the angles between the faces. It is this angle that gives each mineral a distinctive crystal form. The angle between any designated pair of crystal faces is constant, he wrote, and is the same for all specimens of a mineral, regardless of overall shape or size. Steno's discovery that interfacial angles are constant is made clear by the numbering in Figure C4.1. The same faces occur on both crystals. All the sets of faces are parallel: face 1 on the left is parallel to face 1 on the right, face 2 is parallel to face 2, and so forth. Therefore, the angle between any two equivalent faces must be the same on both crystals.

Steno speculated that constant interfacial angles must be a result of internal order, but the ordered particles—ions—were too small for him to see. Proof of internal order was only achieved in 1912 when Max von Laue, a German scientist, demonstrated, by use of X-rays, that crystals are made up of ions packed in fixed geometric arrays, as shown in Figure 4.4.

Crystals form only when a mineral can grow freely in an open space. Crystals are uncommon in nature because most minerals do not form in open, unobstructed spaces. Compare Figures C4.1 and C4.2. The crystals in Figure C4.1 grew freely into an open space, and so, well-developed crystal faces were able to form. The quartz in Figure C4.2, however, grew irregularly, without developing crystal faces, because it grew in an environment restricted by the presence of other minerals. We call such irregularly shaped mineral particles *grains.* Using X-ray techniques, it is easy to show that in both a crystal of quartz and an irregularly shaped grain of quartz, all the atoms present are packed in the same strict crystal structures. That is, both the quartz crystals and the irregular quartz grains are crystalline.

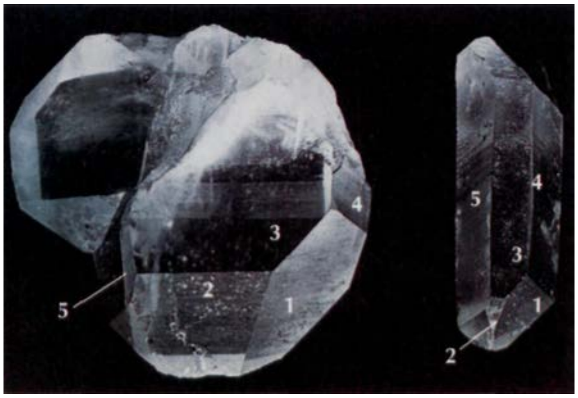


Figure C4.1 Because these two crystals are both quartz, they have the same crystal form. Although the sizes of the individual faces differ markedly between the two crystals, each numbered face on one crystal is parallel to an equivalent face on the other crystal. It is a fundamental property of crystals that, as a result of the internal crystal structure, the angles between adjacent faces are identical for all crystals of the same mineral.

* Crystal form refers to the arrangement of the crystal faces; crystal structure refers to the geometric packing of atoms in a crystal. We can macroscopically and microscopically observe crystal form, but we can only "see" crystal structure with X rays.

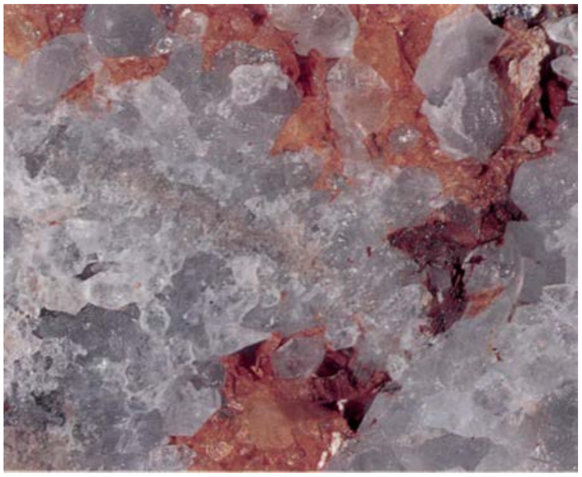


Figure C4.2 Quartz grains (colorless) that grew in an environment where other grains prevented development of well-formed crystal faces. The amber-colored grains are iron carbonate ($FeCO_3$). Compare with Figure C4.1, which shows quartz crystals that grew in open spaces, unhindered by adjacent grains.

Figure C4.3 Distinctive external shape of pyrite, FeS_2. The characteristic shape of pyrite is crystals with faces at right angles and with pronounced striations on the faces. The largest crystals in the photograph are 3 cm on an edge. The specimen is from Bingham Canyon, Utah.

Figure C4.4 Some minerals have distinctive growth habits, even though they do not develop well-formed crystal faces. The mineral chrysotile sometimes grows as fine, cottonlike threads that can be separated and woven into fireproof fabric. When chrysotile is used for this purpose, it is referred to as asbestos.

Every mineral has a characteristic crystal form. Some have such distinctive forms that we can use the property as an identification tool without having to measure angles between faces. For example, the mineral pyrite (FeS_2) is commonly (but not always) found as intergrown cubes (Fig. C4.3) with markedly striated faces. Cube-shaped crystals with striated faces are a reliable way to identify pyrite.

A few minerals develop distinctive growth habits when they grow in restricted environments, and these growth habits can be used for identification. For example, Figure C4.4 shows asbestos, a variety of the mineral serpentine that characteristically grows as fine, elongate threads. (See the "Guest Essay" at the end of this chapter for a discussion of asbestos.)

Cleavage

The tendency of a mineral to break in preferred directions along bright, reflective planar surfaces is called **cleavage.**

If you break a mineral with a hammer or drop a specimen on the floor so that it shatters, you will probably see that the broken fragments are bounded by cleavage surfaces that are smooth and planar, just like crystal faces. In exceptional cases, such as sodium chloride which is the mineral halite (NaCl), as shown in Figure C4.5, all of the breakage surfaces are smooth planar surfaces. (Don't confuse crystal faces and cleavage surfaces, however, even though the two often look alike. A cleavage surface is a breakage surface, whereas a crystal face is a growth surface.)

Many common minerals have distinctive cleavage planes. One of the most distinctive is found in mica (Fig. C4.6). Clay also has a distinctive cleavage; that is why it feels smooth and slippery when rubbed between the fingers.

Figure C4.5 Relation between crystal structure and cleavage. Halite, NaCl, has well-defined cleavage planes; it always breaks into fragments bounded by perpendicular faces.

Luster

The quality and intensity of the light reflected from a mineral produce an effect known as **luster.** Two minerals with almost identical color can have quite different lusters. The most important lusters are described as *metallic,* like that on a polished metal surface, and *nonmetallic.* Nonmetallic lusters are divided into *vitreous,* like that on glass; *resinous,* like that of resin; *pearly,* like that of pearl; and *greasy,* as if the surface were covered by a film of oil.

Color and streak

The color of a mineral, though often striking, is not a reliable means of identification. Color is determined by several factors, one of which is chemical composition, and even trace amounts of chemical impurities can produce distinctive colors.

Color in opaque minerals having a metallic luster can be very confusing because the color is partly a property of grain size. One way to reduce errors of judgment where color is concerned is to prepare a **streak,** which is a thin layer of powdered mineral made by rubbing a specimen on a nonglazed porcelain plate, called a 'streak plate'. The powder gives a reliable color effect because all the grains in a powder streak are very small and so the grain-size effect is reduced. Red streak characterizes hematite (Fe_2O_3), even though the specimen looks black and metallic (Fig. C4.7).

Hardness

The term **hardness** refers to the relative resistance of a mineral to being scratched. It is a distinctive property of minerals. Hardness, like crystal form and cleavage, is

Figure C4.6 Perfect cleavage of mica (variety muscovite) is illustrated by the planar flakes into which this specimen is being split. The cleavage flakes suggest leaves of a book, a resemblance embodied in the term *books of mica.*

Figure C4.7 Color contrast between hematite and a hematite streak. Massive hematite is opaque, has a metallic luster, and appears black. On a porcelain plate, however, this mineral gives a red streak.

governed by crystal structure and by the strength of the bonding forces that hold the atoms of the crystal together. The stronger the forces, the harder the mineral.

Relative hardness values can be assigned by determining the ease or difficulty with which one mineral will scratch another. Talc, the basic ingredient of most baby ("talcum") powder, is the softest mineral known, and diamond is the hardest. A scale called the *Moh's relative hardness scale* is divided into 10 steps, each marked by a common mineral (Table C4.1). These steps do not represent equal intervals of hardness, rather, any mineral on the scale will scratch all other minerals on the scale that have a lower number. Minerals on the same step of the scale can only scratch each other.

Density and Specific Gravity

We know that two identical baskets have different weights when one is filled with feathers and the other with rocks. The property that causes this difference is **density,** or the average mass per unit volume. The units of density are grams per cubic centimeter (g/cm^3).

Because density is difficult to measure accurately, we usually measure a property called specific gravity instead. **Specific gravity** is the ratio of the weight of a substance to the weight of an equal volume of pure water. Specific gravity is a ratio of two weights, and so it does not have any units. Because the density of pure water is 1 g/cm^3, the specific gravity of a mineral is numerically equal to its density.

Steps to Follow in Identifying Minerals

The following steps, used in conjunction with Table C4.2 and Appendix C, will help you identify common minerals.

1. Decide whether the mineral has a metallic or nonmetallic luster. If the mineral has a metallic luster, use the streak, hardness, and cleavage to decide which mineral it is.
2. If it has a nonmetallic luster, determine whether it is harder or softer than the blade of a pocket knife. (If harder, the mineral will scratch the blade; if softer, the blade will scratch the mineral.)
3. Once you determine hardness relative to the knife, decide whether the sample is dark or light in color. Go to the appropriate section of the table and use the cleavage data to determine which mineral you have.

Table C4.1
Moh's Scale of Relative Hardness[a]

	Relative Number in the Scale	Mineral	Hardness of Some Common Objects
Hardest	10	Diamond	
	9	Corundum	
	8	Topaz	
	7	Quartz	
	6	Potassium feldspar	
			Pocketknife; glass
	5	Apatite	
	4	Fluorite	
			Copper penny
	3	Calcite	
			Fingernail
	2	Gypsum	
Softest	1	Talc	

[a]Named for Friedrich Mohs, an Austrian mineralogist, who chose the 10 minerals of the scale.

Table C4.2

Reference Chart for the Identification of Common Minerals and a Guide to the Rock Types in which the minerals might be found

Metallic Luster [a]

Mineral	Streak	Rock Type[b]
Charcopyrite	Greenish yellow	O, I
Galena	Lead gray	O
Hematite	Reddish brown	O, M, S
Limonite	Yellow to brown	S, W
Magnetite	Black	I, M, S
Pyrite	Brass yellow	O, M, I, S
Sphalerite	Yellow to brown	O

Non-Metallic Luster

	Mineral	Cleavage	Rock Type
A. Harder Than a Knife Blade[c]			
Dark Colored			
	Amphibole	Perfect, two planes at 120°	I, M
	Garnet	None	M, I
	Olivine	None	I
	Pyroxene	Perfect, two planes at 90°	I, M
	Quartz	None	I, M, O
Light Colored			
	Feldspar	Perfect, two planes at 90°	I, M
	Quartz	None	I, M, S, O
B. Softer Than a Knife Blade			
Dark Colored			
	Chlorite	Perfect, one plane	M, S
	Hematite (earthly variety)	None	O, S, M
	Limonite (earthly variety)	None	W, S
	Mica (var. biotite)	Perfect, one plane	I, M, S
Light Colored			
	Apatite	Poor, one plane	I, M, S
	Calcite	Perfect, three planes	S, M, O, I
	Clay (var. kaolinite)	Perfect, one plane	W, S
	Dolomite	Perfect, three planes	S, M, O
	Fluorite	Perfect, four planes	O, S
	Gypsum	Perfect, one plane	S, W
	Halite	Perfect, three planes at 90°	S
	Mica (var. muscovite)	Perfect, one plane	I, M, S, O
	Talc	Perfect, one plane	M, S

[a]See Table C.1 in Appendix C for additional properties.
[b]I = igneous, M = metamorphic, O = ore, S = sedimentary, W = weathering product.
[c]See Table C.2 for additional properties.

ROCKS

The definition of a rock, given in the Introduction, is any naturally formed, nonliving, firm, and coherent aggregate mass of solid matter that constitutes part of a planet. The word *mineral* does not appear in the definition because rocks can be made of materials that are not minerals, such as natural glass (in the rock called obsidian), or bits of organic matter (in the rock called coal). Nevertheless, most rocks are made either entirely or almost entirely of minerals, and the relationship between kinds of rocks and the minerals they contain requires closer attention.

There are three large families of rock, each defined by the process that form the rocks:

1. **Igneous rock** (named from the Latin *igneus*, meaning fire) is formed by the cooling and consolidation of magma.
2. **Sedimentary rock** is formed either by chemical precipitation of material carried in solution in sea, lake, or river water, or by deposition of particles of regolith transported in suspension by water, wind, or ice. Soluble matter such as NaCl formed during weathering is the source of material transported in solution. Particulate matter transported in suspension originates in the regolith.
3. **Metamorphic rock** (from the Greek *tneta*, meaning change, and *morphe*, meaning form; hence, change of form) is either igneous or sedimentary rock that has been changed as a result of high temperatures, high pressures, or both. Metamorphism, the process that forms metamorphic rock, is analogous to the process that occurs when a potter fires a clay pot in an oven. The tiny mineral grains in the clay undergo a series of chemical reactions as a result of the increased temperature. New compounds form, and the formerly soft clay molded by the potter becomes hard and rigid.

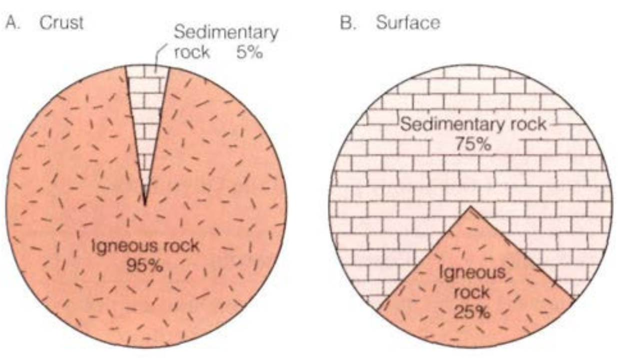


Figure 4.8 Relative amounts of sedimentary and igneous rock. A. The great bulk of the crust consists of igneous rock (95%), with sedimentary rock (5%) forming a thin veneer at the surface. B. Because the sedimentary rock veneer covers so much of the Earth's surface, it is mainly what we see. Thus, 75 percent of the surface is sedimentary rock. Igneous formations pushing through the sedimentary veneer account for the other 25 percent.

The crust is 95 percent igneous rock or metamorphic rock derived from igneous material. However, as Figure 4.8 shows, most of the rock that we actually see at the Earth's surface is sedimentary. The difference arises because sediments are products of weathering, and as a result they are draped as a thin veneer over the largely igneous crust below.

Features of Rocks

At first glance, rocks seem confusingly varied. Some are distinctly layered and have pronounced, flat surfaces covered with the silicate mineral called *mica*. Others are coarse and evenly grained and lack layering; yet, they may contain the same kinds of minerals present in the layered, micaceous (the adjective form of "mica") rock. Studying a large number of rock specimens soon makes it clear that no matter what kind of rock is being examined—sedimentary, metamorphic, or igneous—the differences between samples can be described in terms of two features.

The first feature is **texture**, by which is meant the overall appearance a rock has because of the size, shape, and arrangement of its constituent mineral grains. For example, the grains may be flat and parallel to each other, giving the rock a pronounced platy texture—like a pack of playing cards. In addition, the various minerals may be unevenly distributed and concentrated into specific layers. The rock texture is then both distinctly layered and platy. Specific textural terms are used for each rock family and will be introduced at the appropriate place in subsequent chapters.

Commonly, examination of a microscopic texture requires the preparation of a *thin section* of rock that must be viewed through a microscope. A thin section is prepared by first grinding and polishing a smooth, flat surface on a small fragment of rock. The polished surface is glued to a glass slide, and then the rock is ground away until the glued fragment is so thin that light passes through it easily. A polished surface and a thin section are shown in Figure 4.9.

The second feature used in differentiating rocks is the kinds of minerals present. A few kinds of rock contain only one mineral, but most rocks contain two or more minerals. The varieties and abundances of minerals present in a rock, commonly called the **mineral assemblage** of the rock, are important pieces of information for interpreting how the rock was formed.

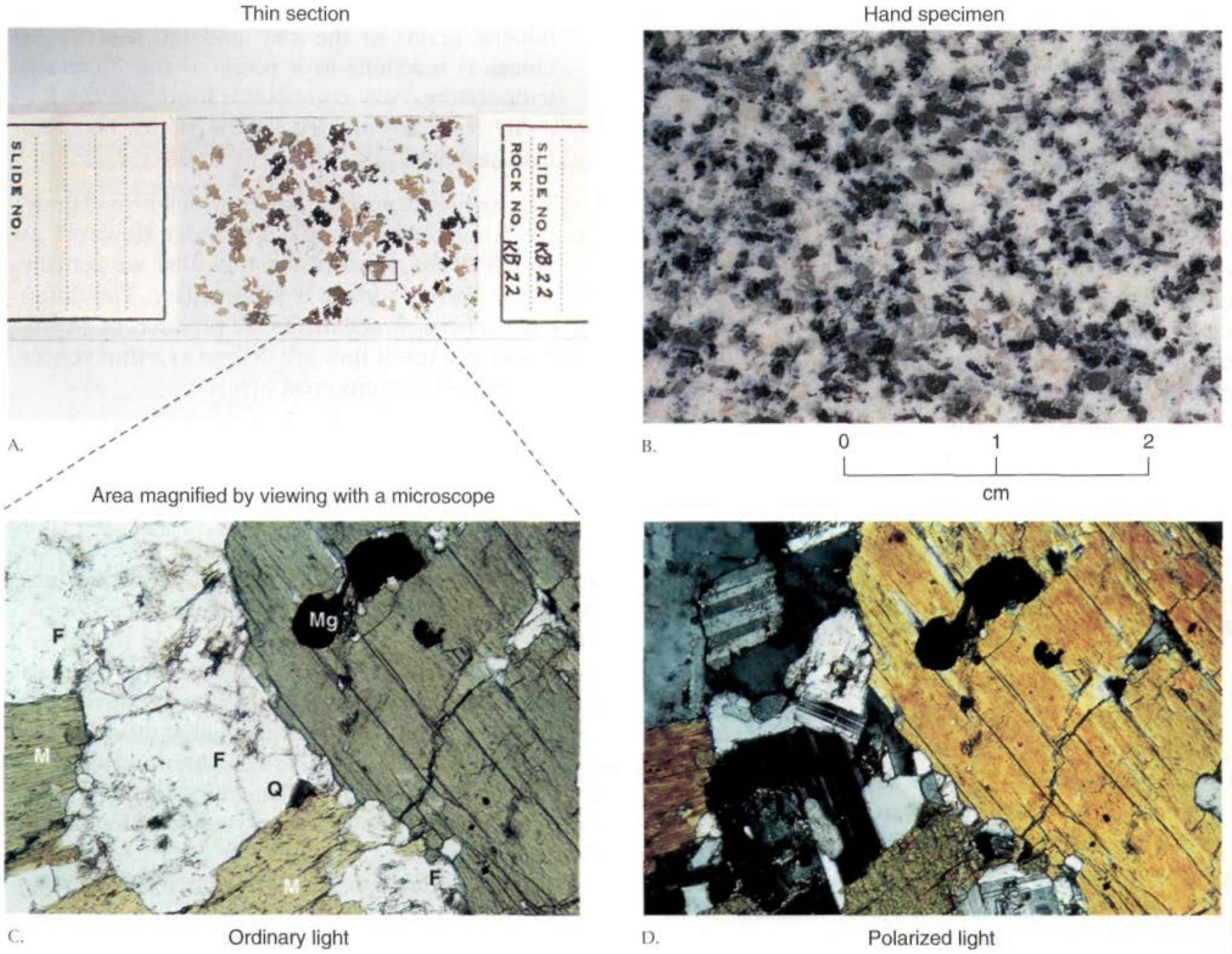


Figure 4.9 Polished surfaces and thin slices reveal textures and mineral assemblages to great advantage. The specimen here is an igneous rock containing quartz (Q), feldspar (F), amphibole (A), mica (M), and magnetite (Mg). A. A thin slice mounted on glass. The slice is 0.03 mm thick, and light can pass through the minerals. B. A polished surface. The dashed rectangle indicates the area used to make the thin slice shown in part A. C. An area of the thin slice as viewed under a microscope. The magnification is 25x. D. The same view as in part C seen through polarizers in order to emphasize the shapes and orientations of individual grains.

The rock in Figure 4.9 is an igneous rock called granite. The mineral assemblage is quartz, feldspar, amphibole, mica (variety, biotite), and magnetite. The texture is typical of granites and would be described as granitic, meaning the grains are uniform in size, intricately interlocked, irregular in shape, and randomly distributed. The minerals found most commonly in the three rock families are listed in Table 4.2.

Table 4.2
Minerals Most Commonly Found in the Three Rock Families

Rock Family	Common Minerals
Igneous	Feldspar, quartz, olivine, amphibole, pyroxene, mica, magnetite
Sedimentary	Clay, chlorite, quartz, calcite, dolomite, gypsum, goethite, hematite
Metamorphic	Feldspar, quartz, mica, chlorite, garnet, amphibole, pyroxene, magnetite

What Holds Rock Together?

The mineral grains in some kinds of rock are held together with great tenacity, whereas in other kinds of rock the grains are easily broken apart. The most tightly bound rocks are igneous and metamorphic because both types contain intricately interlocked mineral grains. During the formation of igneous and metamorphic rocks, the growing mineral grains crowd against each other, filling all spaces and forming an intricate, three dimensional jigsaw puzzle. A similar interlocking of grains holds together steel, ceramics, and bricks.

The forces that hold the grains of sedimentary rocks together are less obvious. Sediment is a loose aggregate of particles, and it must be transformed into sedimentary rock. Sediment becomes sedimentary rock in two ways.

1. *Deposition of a cement.* Water circulating slowly through the open spaces in a sediment deposits new materials such as calcite, quartz, and goethite, which cement the sediment grains together.
2. *Recrystallization.* As a result of the geothermal gradient (see the Introduction), temperature increases with depth. Thus, as layer after layer of sediment is deposited, the deeper layers of sediment are subjected to rising temperatures. In response to increased temperatures, mineral grains in deeply buried sediment begin to recrystallize, and the growing grains interlock and form strong aggregates. The process is the same as when ice crystals in a snow pile recrystallize to form a compact mass of ice.

Both mineral assemblage and texture reflect the conditions under which a rock formed. In the next chapter we will see how these properties can be used to recover information from the most abundant and most important family of rocks, the igneous rocks.

The Rock Cycle

When any type of rock erodes, the eroded particles form sediment. The sediment may eventually become cemented and thereby converted to sedimentary rock. In some places where sedimentary rock forms, the base of the pile of sediment can reach depths at which pressure and heat create new compounds, so that the sedimentary rock becomes metamorphic rock. Sometimes metamorphic rock settles so deep that the high temperatures melt it and magma is formed. This magma moves upward through the crust, cools, and forms igneous rock. Eventually, this is subjected to erosion, the eroded particles form new sediment, and the cycle repeats itself. As James Hutton recognized, this cycle has been continuous throughout the Earth's long history.

The *rock cycle* depicted in Figure 4.10 is one of the

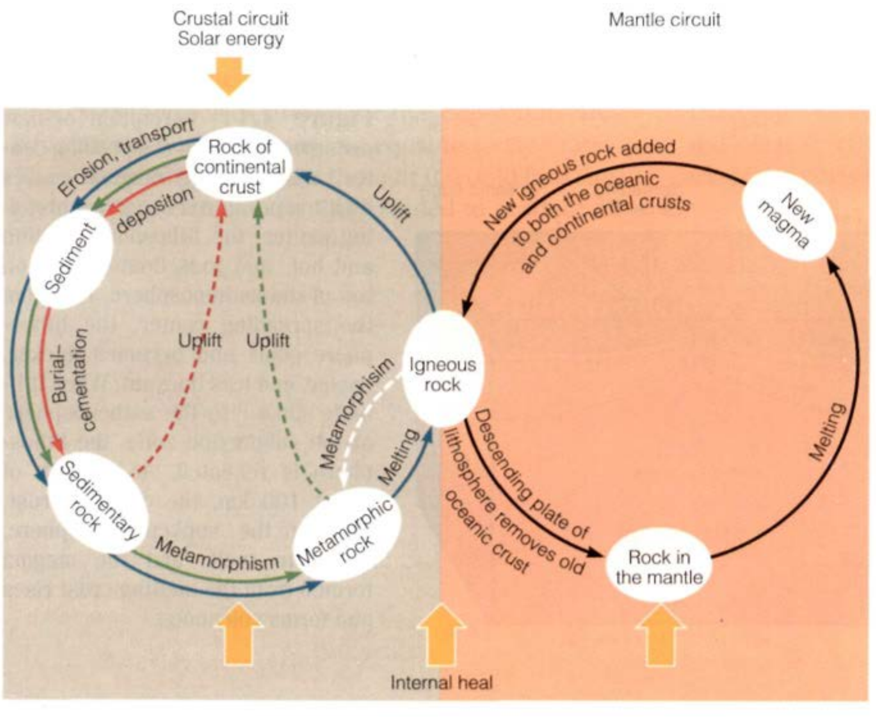


Figure 4.10 The rock cycle, an interplay of internal processes driven by the Earth's internal heat and external processes driven by the Sun's heat energy. Rock in the continental crust can follow any of the arrows from one phase to another. At one time or another, it has followed all of them. In the mantle circuit, magma rises from deep in the mantle and forms new igneous rock, partly in the continental crust and partly in the oceanic crust. The old oceanic crust descends again to the mantle, where it is eventually remixed. The igneous rock added to the continental crust then undergoes the processes of uplift, erosion, sedimentation, and so on in the continental crust circuit.

important components of the Earth system; it is the way the internal activities of the solid Earth interact with external activities involving the hydrosphere, atmosphere, and biosphere. The cycle has two parts, one involving continental crust and the other involving oceanic crust. The example involving igneous, sedimentary, and metamorphic rock described above is simply one circuit (the blue circuit in Fig. 4.10) among many that occur in the rock cycle of the continental crust.

As the three colors of pathways in Figure 4.10 show, other circuits involve sedimentary rock that is not metamorphosed (the orange path) and metamorphic rock that is not melted before being uplifted in a mountain range and eroded. Whether the circuits are long or short, the continental crust is being endlessly recycled. The forces that produce the basins where sediment accumulates and that cause mountains to be uplifted and thereby eroded are the same internal forces that move tectonic plates around. Thus, the rock cycle in the continental crust operates the way it does because of plate tectonics. Because the mass of the continental crust is large, the average time a rock takes to complete the cycle is long. Time estimates vary, and they are difficult to make, but the average age of all rock in the continental crust seems to be about 650 million years.

The rock cycle in the oceanic crust also works the way it does because of plate tectonics. The basic concept of plate tectonics is that magma rises from the asthenosphere through great rifts (long, narrow fractures) in the seafloor and forms new oceanic crust (Fig. 4.11). These great rifts run down the center of a vast submarine mountain chain, known both as the **oceanic ridge** and the **midocean ridge.**

Because they are places where two plates of lithosphere are moving away from each other, the rifts in the center of the oceanic ridge are called **spreading centers.** Far away from a spreading center the lithosphere, the upper part of which is the oceanic crust, sinks back into the asthenosphere. The place where a plate sinks is called a **subduction zone.** The subducted lithosphere heats up as it sinks, and at a depth of about 100 km (62 mi) the oceanic crust portion of the sunken lithosphere starts to melt and new magma is formed. The new magma rises to create volcanoes.

Figure 4.11 makes it clear that, through plate tectonics, both the mantle and the crust are involved in the rock cycle. What is not clear is how the mantle influences the composition of seawater. The magma that rises to form new oceanic crust becomes hot igneous rock that reacts with seawater. Some constituents of the hot rock, such as calcium, are dissolved in the seawater, and some constituents already in the seawater, such as magnesium, are deposited in the igneous rock. Thus, via the reactions between hot, newly formed crust and seawater, the mantle plays a role in determining the composition of seawater.

The mantle also plays a role in determining the composition of the atmosphere and the viability of the biosphere. All magma contains some dissolved gas. During an eruption the dissolved gas bubbles out of the magma and mixes with the atmosphere. Because most magma originates in the mantle, eruptions are the means by which gas is transferred from the mantle to the atmosphere.

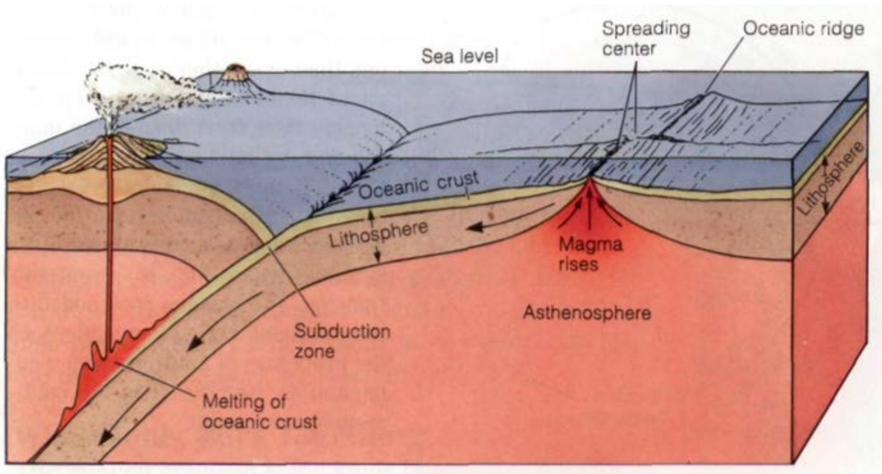


Figure 4.11 Formation of new oceanic crust from a spreading center. Where rising magma causes high temperatures near the spreading center, the lithosphere is thin and hot, and thus floats easily on top of the asthenosphere. Far from the spreading center, the lithosphere cools and becomes thicker, cooler, and less buoyant. When it finally sinks into the asthenosphere at the subduction zone, the lithosphere is reheated. At a depth of about 100 km, the oceanic crust, part of the sunken lithosphere, starts to melt, and the magma formed from the melting crust rises and forms volcanoes.

The Laki eruption discussed in the Introduction is an example of the way an eruption can influence the composition of the atmosphere. The effects of the Laki eruption did not last long. There have been times in the Earth's long history, however, when so many volcanoes were erupting, and so much gas was being added to the atmosphere, that very long-lasting climatic effects resulted. An unusually high rate of volcanic activity occurred between 135 and 115 million years ago when a vast, submarine lava plateau was formed by eruption in the southwest Pacific. So much carbon dioxide was released during eruption that it is estimated that the atmospheric concentration of carbon dioxide was about 20 times higher than today's level. As a result, the global temperature rose about 10°C (18° F) and several million years of scorching climates followed, causing drastic changes in the distribution of plants and animals.

Uniformitarianism and the Rate of the Rock Cycle

As mentioned in the Introduction, it was James Hutton who recognized that the same external and internal processes occurring today have been operating throughout the Earth's long history and therefore that the present is the key to the past. This is the Principle of Uniformitarianism.

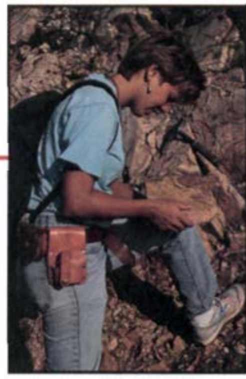
During the nineteenth century, with the Principle of Uniformitarianism generally accepted, geologists tried to estimate how long the rock cycle has been going on by estimating the thickness of all sediments laid down through geological time. They assumed that the principle applied to process rates as well as to the processes themselves, and hence that deposition rates have always been constant and equal to today's rates. Thus, these early geologists thought it would be a simple matter to estimate the time needed to produce all the sediments. The results, we now know, were greatly in error. One of the reasons for the error was the assumption of rate constancy. The more we learn of the Earth's history and the more accurately we determine the timing of past events through radiometric dating (Chapter 7), the clearer it becomes that cycle rates have not always been the same.

One reason the rate of the rock cycle has changed through time is that the Earth is very slowly cooling as its internal heat leaks away. The Earth's internal temperature is maintained by natural radioactivity. Because radioactive isotopes transform spontaneously to nonradioactive isotopes, the Earth's natural radioactivity is slowly declining. Early in its history, the Earth must therefore have contained more radioactive atoms than there are today, and so more heat must have been produced than is produced today. Therefore, the internal processes, because they are driven by the Earth's internal heat, must have been more rapid during Earth's early history than they are today. It is possible that 3 billion years ago oceanic crust was created at a faster rate than now, that tectonic plates moved rapidly, that volcanoes were more active, and that continental crust was uplifted and eroded at a faster rate. Any or all of these actions would cause the rock cycle to speed up.

At the same time, the rates of external processes have also varied. Long-term changes in rates seem to have occurred because of slow increases in the luminosity of the Sun and because of the gradual slowing of the Earth's rotation. (Scientists estimate that 600 million years ago there were 400 days in the year and 2 billion years ago there were 450.) In other words, even though the rock cycle has been continuous, it has not maintained a constant rate through time. Therefore, we now turn to a consideration of magma and what it tells us about the rates of the Earth's internal activities today.

Guest Essay

Asbestos: To Understand the Science Is to Understand the Politics



Jill S. Schneiderman is Associate Professor of Geology at Pomona College in Claremont, California. Winner of a 1992-1993 Wig Distinguished Teaching Award, she is dedicated to the realization of accessible science education. Though a metamorphic petrologist by training, she has received funding from the National Science Foundation and the Smithsonian Institution for her current research in climate change in North Africa.

"Asbestos" is a commercial rather than a mineralogical term applied to a variety of fibrous silicate minerals of different chemical compositions. Because of an unusual combination of useful properties such as the ability to divide into fine fibers, high strength and flexibility, high chemical and mechanical durability, low thermal and electric conductivity, and relative incombustibility, asbestos is used widely in industrial products and processes. Roofing and flooring materials, automobile brake linings, cement and mortar, and building insulation commonly contain asbestos.

Asbestiform describes the tendency of any mineral to break into fine fibers. Dozens of minerals have asbestiform habits under some physical conditions; however, only six asbestiform minerals attract commercial interest. These six minerals—chrysotile, amosite, crocidolite, fibrous anthophyllite, fibrous tremolite, and fibrous actinolite—are all referred to as asbestos. The majority of asbestos (95%) is chrysotile, $Mg_3Si_2O_5(OH)_4$, also known by the generic name serpentine, which is formed by the linking of silicate tetrahedra (Figs. 4.5 and 4.6) to form sheets. (Note that, as discussed in the essay in Chapter 6, serpentine provides a major component in serpentinite, which is formed by the hydration of olivine in oceanic crust.) In chrysotile, the sheets of tetrahedra curl up to form hollow cylinders, thus producing the asbestiform habit. Most of the chrysotile for commercial asbestos comes from mineral deposits in Canada. The five other minerals that form asbestos—amosite, crocidolite, fibrous anthophyllite, fibrous tremolite, and fibrous actinolite—are all amphiboles; amphiboles are silicate minerals of variable composition that have a crystal structure formed of two chains of silicon tetrahedra linked together to form a double chain (Figs. 4.5 and 4.6). These five minerals are, respectively:

$Fe^{2+}{}_7Si_8O_{22}(OH)_2$,
$Na_2Fe^{2+}{}_3Fe^{3+}{}_2Si_8O_{22}(OH)_2$,
$Mg_7Si_8O_{22}(OH)_2$, $Ca_2Mg_5Si_8O_{22}(OH)_2$, and
$Ca_2(Fe^{2+},Mg)_5Si_8O_{22}(OH)_2$.

The double-chain structure also gives rise to the asbestiform habit.

During the 1960s researchers conducted studies on rates of mesothelioma, or cancer of the lining of the chest or abdomen, which is nearly always fatal, among insulation workers. The results convincingly showed that exposure to asbestos caused this disease. As a result, strict regulations to control the amount of airborne asbestos fibers were enacted to protect the health of construction workers, firefighters, and maintenance workers as **well** as others occupationally exposed to asbestos. However, epidemiological studies during the 1970s suggested that the type of fiber comprising asbestos affects the degree of health risk. In particular, some researchers concluded that chrysotile asbestos is far less carcinogenic than amphibole asbestos.

Currently, researchers studying the health hazards of asbestos remain severely divided; one group maintains that chrysotile asbestos is not a health risk in nonoccupational environments such as schools and houses. In fact, they state that since 95 percent of the asbestos used commercially in the United States is chrysotile asbestos, the health risk posed by asbestos in buildings is much less than other environmental health hazards like radon or tobacco smoke. Hence, they say that chrysotile asbestos should be regulated differently than amphibole asbestos. Furthermore, since current regulations also protect the occupants of buildings that contain asbestos by requiring asbestos removal, they advocate removal of only amphibole asbestos. The other group maintains that fiber-type studies are uncertain and that epidemiological data show that exposure to asbestos, whether it is chrysotile asbestos or amphibole asbestos, leads to asbestosis, a disease that causes scarring of the lungs, and lung cancer. They advocate the removal of both chrysotile asbestos and amphibole asbestos from buildings.

An important question further complicates this issue: what acts as a greater airborne health risk, leaving asbestos in place or disturbing it through removal? Because of its fibrous structure, asbestos easily becomes an airborne particle, readily dispersed to be inhaled and ingested. Therefore, removal of asbestos may produce a health hazard that otherwise would not exist if asbestos remained entombed in ceiling tiles, cement, and mortar. Answers offered to the question posed are frequently motivated by politics and economics. For example, school districts that have paid enormous sums of money to remove asbestos from school buildings will be able to recoup their costs from manufacturers if indeed it turns out that all asbestos is deemed a substantial health hazard. Asbestos manufacturers stand to gain by being absolved of responsibility for removal costs if chrysotile asbestos is as innocuous as some claim. Certainly, the debate about asbestos has created a huge industry in asbestos removal, and thousands of legal cases regarding asbestos have kept lawyers busy and prosperous. Whose interests are being served? As the debate continued in September 1993, the New York City Board of Education was forced to delay by three weeks the opening of schools so that asbestos inspection could be completed. Whatever the outcome, a knowledge of mineralogy makes the debate comprehensible.

Summary

1. Minerals are naturally formed, solid chemical elements or compounds having a definite composition and a characteristic crystal structure. Crystal structure is the geometric array of atoms in a crystalline solid.
2. Minerals are formed through the bonding together of cations and anions of different chemical elements.
3. Silicates are the most common minerals (95% of the crust), followed by oxides, carbonates, sulfides, sulfates, and phosphates.
4. The basic building block of silicate minerals is the silicate tetrahedron, a complex anion in which an Si^{4+} ion is bonded to four O^{2-} ions. The four O^{2-} ions sit at the apexes of a tetrahedron, with the Si^{4+} at its center. Adjacent silicate tetrahedra can bond together to form polymers by sharing oxygens.
5. The feldspars are the most abundant group of minerals in the Earth's crust (60%). Quartz is the second most common mineral in the crust (15%).
6. The principal properties used to characterize and identify minerals are crystal form, growth habit, cleavage, luster, color and streak, hardness, and specific gravity.
7. There are three families of rocks: sedimentary, igneous, and metamorphic. Igneous are the most common kinds of rock in the continental crust.
8. The differences between rocks can be described in terms of texture and mineral assemblage.
9. Sediment is transformed to sedimentary rock by cementation or recrystallization of the sediment particles.
10. The rock cycle arises from the interactions of the Earth's internal and external processes. Igneous rock is eroded, creating sediment, which is deposited in layers that become sedimentary rock. When sedimentary rock is buried, changes in temperature and pressure cause it to convert to metamorphic rock. Eventually, temperatures and pressures may become so high that metamorphic rock melts and forms new magma. The magma rises, forms new igneous rock, and the cycle is repeated.
11. The rock cycle in the oceanic crust interacts with that in the continental crust through the agency of plate tectonics.
12. The Principle of Uniformitarianism is an accurate guide to understanding geological processes throughout the Earth's long history, but the principle is not a guide to the rates of geological processes. Many rates have varied considerably.

Questions for Review

1. What is a mineral? Give three reasons why the study of minerals is important.
2. Approximately how many common minerals are there? Which is the most common one?
3. Name five minerals that are found in the area in which you live. Are any minerals mined in the area in which you live? What are they and for what are they mined?
4. Describe the structure of the silicate anion.
5. Describe how silicate anions join together to form silicate minerals.
6. Describe the polymer structure of pyroxenes, micas, and feldspars.
7. Name five minerals that are not silicates and name the anion each contains.
8. Can a rock be uniquely defined on the basis of its mineral assemblage? If not, what additional information is needed?
9. Describe two ways by which loose aggregates of sediment are transformed into sedimentary rock.
10. What holds together the mineral grains in metamorphic and igneous rocks?
11. Why are sedimentary rocks so abundant at the Earth's surface when igneous rocks make up most of the crust?
12. What is the rock cycle? How does oceanic crust interact with continental crust through the rock cycle?

Questions for A Closer Look

1. What properties besides composition and crystal structure can be used to identify minerals?
2. If you found a mineral that met the following description—metallic luster and lead gray streak—what might it be?
3. If you found a mineral that was light in color, had a single perfect cleavage, and was softer than a knife blade, what might it be?

Questions for Discussion

1. Within the room in which you are sitting, identify all the objects that are derived in some way from minerals. What would happen to society if all mining were stopped?
2. Would you expect other planets in the solar system to have the same kinds of minerals as on the Earth? How about planets around other suns? Be sure to say why you think there may be similarities or differences.

Important Terms to Remember

Terms in italic are defined in A Closer Look

anion (p. 93)
atom (p. 92)
cation (p. 93)
cleavage (p. 100)
crystal (p. 98)
crystal face (p. 98)
crystal form (p. 98)
crystal structure (p. 98)
density (p. 101)
element (chemical) (p. 92)
energy-level shell (p. 93)
hardness (of a mineral) (p. 100)
igneous rock (p. 103)
ion (p. 93)
isotope (p. 92)
luster (p. 100)
metamorphic rock (p. 103)
midocean ridge—oceanic ridge (p. 106)
mineral (p. 92)
mineral assemblage (p. 103)
oceanic ridge—midocean ridge (p. 106)
sediment (p. 103)
sedimentary rock (p. 103)
silicate (p. 95)
silicate anion (SiO_4^{-4}) (p. 95)
specific gravity (p. 101)
spreading center (p. 106)
streak (p. 100)
subduction zone (p. 106)
texture (of a rock) (p. 103)

CHAPTER 5

The Heat Within: Magma and Volcanoes

A painting of a violent eruption of Mount Vesuvius, Italy, in 1737, taken from an old-time lantern slide. The Roman statesman, Pliny, died during a similar eruption in 79 A.D.; the rising column of gas and glowing fragments is now called a plinian column.

Violent Eruptions

"During the summer of 1883, an apparently dormant Indonesian volcano called Krakatau started to emit steam and ash. Krakatau was an island off the western end of Java. On Sunday, August 26, activity increased, and on the next day Krakatau blew up: the island disappeared. As a telegram of the time tersely reported, "Where once Mount Krakatau stood, the sea now plays." Noise from the paroxysmal explosion was heard on an island in the Indian Ocean, 4600 km (2,900 mi) away. As the island blew apart, it created *tsunami* (gigantic waves), some as high as 40 m *(44* yd); tsunami spread out from the site of the explosion and crashed into the shores of Java and Sumatra, the two closest Indonesian islands. Thirty-six thousand people lost their lives.

The effect of Krakatau was felt around the world. About 20 km^3 (5 cu mi) of volcanic debris was ejected during the eruption, some blasted as high as 50 km (31 mi) into the stratosphere. Within 13 days the stratospheric dust had encircled the globe, and for months there were strangely colored sunsets—sometimes green or blue and other times scarlet or flaming orange. One November sunset over New York City looked so much like the glow from a massive fire that fire engines were called out. The suspended dust made the atmosphere so opaque to the sun's rays that the temperature around the Earth dropped an estimated 0.5°C during 1884. It was five years before all the dust fell to the ground and the climate returned to normal.

In March 1980 the sudden, violent eruption of Mount St. Helens, a long quiescent volcano in the state of Washington, reminded us once again of the enormous magnitude of volcanic forces. Mount St. Helens was known to have been active in historic times, but it had not erupted for more than 200 years. Then, in early 1980 people living near the volcano began reporting frequent small earthquakes. On March 27 steam and volcanic ash puffed from the summit.

Monitoring by geologists working for the U.S. Geological Survey quickly revealed that Mount St. Helens was swelling like a balloon: the north face was watched especially closely because by early May it was bulging outward at a rate of 1.5 m/day (1.6 yd/day). From an observation post several kilometers north of the volcano, geologists, stationed at Vancouver, Washington, mounted a round-the-clock watch. On Sunday, May 18, 1980, David A. Johnson was on duty, and at 8:32 A.M. he shouted into his microphone, "Vancouver, Vancouver, this is it." They were Johnson's last words. A devastating eruption was under way. A gigantic mass of volcanic particles and very hot gases blasted out sideways, directly toward Johnson. No trace of him or the observation post has ever been found. At least 62 other people were killed by the eruption, but the total would have been much higher had the authorities not heeded the geologist's early warnings and kept people far away. Mount St. Helens didn't exactly disappear the way Krakatau did, but it was certainly beheaded. Originally a little more than 2900 m (1.8 mi) high, it is only 2490 m (1.5 mi) high today.

Scientists recognize several kinds of volcanoes, each characterized by a distinct kind of magma and a distinct eruption style. Magma is the molten material that forms when rock melts. We can learn a great deal about the processes taking place deep inside the Earth by studying what kind of rock melts, in which part of the Earth's interior the melting happens, and what kind of volcanic edifice marks the place where the magma reaches the surface.

It is important to study volcanism and the deep-seated processes that give rise to it because volcanism plays an important role in Earth system science. Volcanic eruptions, particularly large ones such as the Krakatau eruption of 1883, can change climates around the world. Through volcanism, events that happen deep inside the Earth influence what happens on the Earth's surface.

PROPERTIES OF MAGMA

One of the best ways to learn about volcanoes and the Earth's interior is to study magma, the material that volcanoes erupt. Magma was briefly defined in Chapter 3 as molten rock, but in fact it is a much more complex material. A complete definition of **magma** is the mixture of molten rock, suspended mineral grains, and dissolved gases that forms in the crust or mantle when temperatures are sufficiently high. Magma reaches the Earth's surface through a **volcano,** which is a vent from which magma, solid rock debris, and gases are erupted. The term *volcano* comes from the name of the Roman god of fire, Vulcan, and it conjures up visions of streams of **lava**—magma that reaches the Earth's surface—pouring out over the landscape. Although some lava does flow as hot streams, magma can also be erupted as clouds of tiny, red-hot fragments, as was the case at Mount St. Helens.

Volcanoes are the only places we can see and study magma, and so we start this chapter by gaining some insight into volcanoes and the properties of magma.

By observing lava, we can to draw three important conclusions concerning magma:

1. Magma is characterized by a *range of compositions* in which silica (SiO_2) is always predominant.
2. Magma has the properties of a liquid, including the *ability to flow.* This is true even though most magma is a mixture (often referred to as *melt)* of suspended crystals, dissolved gases, and molten rock and, in some instances, almost as stiff as window glass.
3. Magma is characterized by *high temperatures.*

Composition

Magma composition is determined by the common chemical elements in the Earth—silicon (Si), aluminum (Al), iron (Fe), calcium (Ca), magnesium (Mg), sodium (Na), potassium (K), hydrogen (H), and oxygen (O). Because O^{2-} is by far the most abundant anion and is therefore the anion that balances the charges on all the cations, scientists usually express magma composition in terms of charge-balanced oxides, such as SiO_2 and $A1_2O_3$. The most abundant component of magma is silica, SiO_2.

Small amounts of gas (0.2 to 5% by weight) are dissolved in all magma and play an important role in eruptive processes. The principal gas is water vapor, which, together with carbon dioxide, accounts for more than 98 percent of all gases emitted from volcanoes. The remaining 2 percent is nitrogen, chlorine, sulfur, and argon.

Magma gases are also called volcanic gases. They are important in Earth system science because:

1. Volcanic gases, especially carbon dioxide and sulfur dioxide, influence the composition of the atmosphere and thereby the climate.
2. The rate at which gas bubbles out of a magma controls the violence of an eruption; rapid bubbling means a violent and therefore hazardous eruption.
3. Violent, gas-driven eruptions, such as the eruptions of Tamboro discussed in the Introduction, and the eruptions of Krakatau and Mount St. Helens mentioned in this chapter's opening essay, can blast such massive amounts of volcanic dust into the atmosphere that the global temperature can drop. The drop was 0.5°C (0.9° F) in the case of Krakatau, but could be 1°C (1.8°F) or more for a huge eruption.

Three distinct types of magma are more common than all others: basaltic, andesitic, and rhyolitic (Fig. 5.1).

1. Basaltic magma contains about 50 percent SiO_2 and very little dissolved gas. The two common igneous rocks derived from basaltic magma are **basalt** and **gabbro.**[1]
2. Andesitic magma contains about 60 percent SiO_2 and a lot of dissolved gas. **Andesite** and **diorite** are the common igneous rocks derived from andesitic magma.

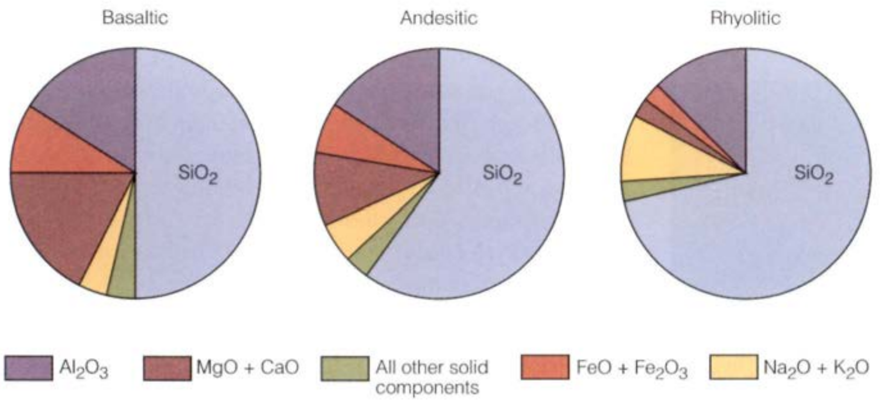


Figure 5.1 The average composition of the solid part of the three principal kinds of magma. In addition to the solid materials, the magmas also contain dissolved gases. Basaltic magma has a low content of dissolved gas; andesitic and rhyolitic magmas tend to be very gassy.

3. Rhyolitic magma contains about 70 percent SiO_2 and more gas than either basaltic or andesitic magma. **Rhyolite** and **granite** are the common igneous rocks derived from rhyolitic magma.

The three magmas are not formed in equal abundance. Approximately 80 percent of all magma erupted by volcanoes is basaltic, with andesitic and rhyolitic each about 10 percent. Hawaiian volcanoes, such as Kilauea and Mauna Loa, are basaltic. Mount St. Helens and Krakatau are both andesitic volcanoes, and the now dormant volcanoes at Yellowstone National Park are rhyolitic. As we will see in the next chapter, the locations of the different kinds of volcanoes are closely related to plate tectonics.

Viscosity

Dramatic pictures of lava flowing rapidly down the side of a volcano prove that some magma is very fluid. Basaltic lava moving down a steep slope on Mauna Loa in Hawaii has been clocked at 16 km/h (9.9 mi/h). Such fluidity is rare, however, and flow rates are more commonly measured in meters per hour or even meters per day. As suggested by the scene in Figure 5.2, which shows basaltic lava destroying a house in Hawaii, flow rates are usually slow enough so that people can easily get out of the way.

The property that causes a substance to resist flowing is **viscosity.** The more viscous a magma, the less fluid it is. Magma viscosity depends on temperature and composition, especially the SiO_2 content. The higher the silica content, the more viscous the magma. For this reason, rhyolitic magma is always more viscous than basaltic, and andesitic magma has a viscosity intermediate between the two. As we will see, magma viscosity is one factor that affects the violence of a volcanic eruption.

Figure 5.2 An advancing tongue of basaltic lava setting fire to a house in Kalapana, Hawaii, during an eruption of Kilauea volcano in June 1989. Flames at the edge of the flow are due to burning lawn grass.

Temperature

Magma temperature can sometimes be measured during a volcanic eruption. Because volcanoes are dangerous places and because scientists who study them are not eager to be roasted alive, measurements must be made from a distance using optical devices. Magma temperatures determined in this manner during eruptions range from 1000° to 1200°C (1832° to 2192°F).

The higher the temperature, the lower the viscosity of a magma and the more readily it flows. In Figure

[1] Basalt and gabbro contain the same minerals; the two rocks differ only in grain size; with basalt containing small mineral grains, and gabbro coarse grains. Andesite and rhyolite are fine-grained rocks; diorite and granite are their respective coarse-grained equivalents. See "A Closer Look: Naming Igneous Rocks" for more information on rock names.

Figure 5.3 Lava flow rate is controlled by viscosity, which in turn is controlled by temperature. The formation, on which the geologist is standing, is pahoehoe lava formed from a very hot, low-viscosity, and therefore fast-moving lava that was erupted in 1959. The upper flow (the one being sampled), which is relatively cool and therefore very viscous and slow moving, is an aa lava erupted from Kilauea volcano in 1989. They have the same basaltic composition.

5.3, the smooth, ropy-surfaced lava on which the geologist is standing, called *pahoehoe* (a Hawaiian word pronounced pa-ho-e-ho-e), formed from a hot, very fluid basaltic magma. The rubbly, rough-looking lava piled up at the center and left formed from a cooler basaltic magma that had a higher viscosity. Hawaiians call this rough lava *aa* (pronounced ah'-ah).

No matter how hot a magma is when it exits a volcano, the lava soon cools, becomes more viscous, and eventually slows to a complete halt.

ERUPTION OF MAGMA

Magma, like most other liquids, is less dense than the solid matter from which it forms. Therefore, once formed, lower density magma exerts an upward push on any enclosing higher density rock and slowly forces its way up. There is, of course, a reverse pressure on a rising mass of magma due to the weight of all the overlying denser rock. Because this pressure is proportional to depth, it decreases as a magma rises upward.

Pressure controls the amount of gas a magma can dissolve; more gas is dissolved at high pressure, less at low. Gas dissolved in a rising magma acts the same way as gas dissolved in soda water. When a bottle of soda is opened bubbles form because the pressure inside the bottle has dropped, allowing gas to come out of solution. Gas dissolved in an upward-moving magma also comes out of solution and forms bubbles as the pressure on the magma decreases. What happens to these bubbles determines whether an eruption will be explosive or nonexplosive.

Nonexplosive Eruptions

It is understandable why people tend to regard any volcanic eruption as a hazardous event and active volcanoes as dangerous places that should be avoided. However, geologists have discovered that basaltic volcanoes, such as those in Hawaii, are comparatively safe because they generally erupt nonexplosively.

The differences between nonexplosive and explosive eruptions are largely a function of magma viscosity and dissolved gas content. Nonexplosive eruptions are favored by low-viscosity magmas and low-dissolved gas levels. Basaltic magma has a lower SiO_2 content, a higher temperature, and therefore a lower viscosity than andesitic or rhyolitic magmas; it also has a lower content of dissolved gas than either of the other magma types. Eruptions of basaltic magma are rarely explosive.

Even though they are nonexplosive, basaltic magma eruptions can be spectacular during their initial stages. Gas bubbles in a low-viscosity basaltic magma will rise rapidly upward, like the gas bubbles in a glass of soda water. If basaltic magma moves rapidly through the Earth's surface, which means that the pressure exerted by overlying rock drops quickly, gas can come out of solution so rapidly that when the magma starts erupting the froth of bubbles can cause spectacular fountaining (Fig. 5.4). When fountaining dies down because most of the dissolved gas has come out of solution and escaped, the hot, fluid lava emerging from the vent flows rapidly downslope (Fig. 5.5). As the lava cools, it continues to lose dissolved gases, its viscosity increases and the character of flow changes. The very fluid initial lava forms thin pahoehoe flows, but with increasing viscosity the rate of movement slows and the cooler, stickier lava is transformed into a slow-moving aa flow. Thus, during a single, nonexplosive, Hawaiian-type eruption, pahoehoe and aa may be formed from the same batch of magma.

As a hot, low-viscosity basaltic lava cools and the viscosity rises, the gas bubbles find it increasingly dif-

Figure 5.4 Fountaining starts an eruption of Kfafla, a basaltic volcano in Iceland. Use of a telephoto lens foreshortens the field of view. The geologist in a protective suit is making measurements several hundred meters away from the fountain.

Figure 5.5 This stream of low-viscosity (and therefore very hot) basaltic lava moving smoothly away from an eruptive vent demonstrates how fluid and free flowing lava can be. The temperature of the lava is about 1100°C. The eruption occurred in Hawaii in 1983.

ficult to escape. When the lava finally solidifies to rock, the last-formed bubbles become trapped and their form is preserved. These bubble holes are called *vesicles,* and the texture they produce in an igneous rock is said to be *vesicular.*

Explosive Eruptions

Viscous magmas—both andesitic and rhyolitic—with higher silica content and lower temperature than basaltic magma also tend to have high dissolved-gas contents; the combination is a recipe for an explosive eruption. As the gas-charged, viscous magma rises, the gas comes out of solution and bubbles form but cannot escape from the sticky, viscous magma. If the rate at which the magma rises (and hence the rate at which bubbles form) is rapid, the bubbles can shatter a viscous magma into a cloud of tiny, red-hot fragments.

A fragment of hot, shattered magma, or any other fragment of rock ejected during an explosive volcanic eruption, is called a **pyroclast** (named from the Greek words *pyro,* meaning fire, and *klastos,* meaning broken). Geologists refer to a deposit of unconsolidated (loose) pyroclasts as **tephra,** a Greek term for ash; when the pyroclasts in a deposit of tephra are consolidated (cemented together), the result is a **pyroclastic rock.** The terms used to describe tephra of different sizes are listed in Table 5.1 and illustrated in Figure 5.6.

Before proceeding, let's review the different materials erupted from volcanoes.

1. Lava is magma that oozes out of a volcano and flows over the landscape.
2. Pyroclasts are broken bits of rock as well as hot fragments of viscous magma shattered as a result of gas escape. An unconsolidated deposit of pyroclasts is called tephra.
3. Volcanic gases, which are mainly water vapor, carbon dioxide, and sulfur, are emitted before, during, and after the eruption of lava and pyroclasts.

Table 5.1
Names for Tephra and Pyroclastic Rock

Average Particle Diameter (mm)	Tephra (unconsolidated material)	Pyroclastic Rock (consolidated material)
>*64* mm	Bombs	Agglomerate
2-64 mm	Lapilli	Lapilli tuff
<2 mm	Ash	Ash tuff


A.


B.


C.

Figure 5.6 Tephra. A. Large spindle-shaped pyroclasts up to 50 cm in length cover the surface of a tephra cone on Halcakala volcano, Maui. B. Intermediate-sized tephra called lapilli cover the Kau Desert, Hawaii. The coin is about 1 cm in diameter. C. Volcanic ash, the smallest-sized tephra, covers leaves in a garden in Anchorage, Alaska, following the eruption of Mount Spurr.

Eruption Columns and Tephra Falls

As rising magma approaches the Earth's surface, the rapid drop in pressure causes dissolved gas to bubble furiously, like a violently shaken bottle of soda. As a result of the bubbling, a viscous magma can break into a mass of hot, glassy pyroclasts; the resulting mixture of hot gas and pyroclasts produces a violent upward thrust that culminates in an explosive eruption. The hot, turbulent mixture of gas and pyroclasts rises rapidly in the cooler air above the volcano to form an *eruption column* that may reach as high as 45 km (28 mi) in the atmosphere. The opening figure for this chapter shows a huge eruption column rising from Mount St. Helens. The rising, buoyant column is driven by heat energy released from hot, newly formed pyroclasts. At a height where the density of the column equals that of the surrounding atmosphere, the column begins to spread laterally to form a mushroom-shaped eruption cloud.

As an eruption cloud begins to drift with the upper atmospheric winds, the pyroclasts fall out and eventually accumulate on the ground as tephra deposits. During exceptionally explosive eruptions, tephra can be spread over distances of 1500 km (932 mi) or more. Some eruption columns reach such great heights that winds are able to transport the pyroclasts and gases completely around the world. This was the case in the eruptions of Tamboro and Krakatau. As mentioned in this chapter's opening essay, such atmospheric pollution, by blocking incoming solar energy, can lower average temperatures at the land surface for a year or more and cause spectacular sunsets as the sun's rays are refracted by the airborne particles.

Pyroclastic Flows

A hot, highly mobile flow of tephra that rushes down the flank of a volcano during a major eruption is called a *pyroclastic flow.* Such a flow, which is often referred to by the French term *nuee ardente* (glowing cloud), occurs when the mixture of red-hot tephra and searing gases is too dense to rise upward. Pyroclastic flows are among the most devastating and lethal forms of volcanic eruptions. The worldwide geologic record of historic pyroclastic flows shows that they can travel 100 km (62 mi) or more from source vents and reach velocities of more than 700 km/h (435 mi/h). One of the most destructive, on the Caribbean island of Martinique in 1902, rushed down the flanks of Mount Peleé volcano and overwhelmed the city of Saint Pierre, instantly killing 29,000 people. The term *nuee ardente* was coined by the French geologists who investigated the disaster at Saint Pierre.

Lateral Blasts

The eruption of Mount St. Helens in 1980 displayed many features of a typical large, explosive eruption. Nevertheless, the magnitude of the event caught geologists by surprise. The events leading to this eruption are shown diagrammatically in Figure 5.7. As magma moved upward under the volcano, the northern flank of the mountain began to bulge upward and outward. Finally, the slope became unstable, broke loose, and quickly slid toward the valley as a gigantic landslide of rock and glacier ice. The landslide exposed the mass of hot magma in the core of the volcano. With the lid of rock removed, dissolved gases bubbled so furiously that a mighty blast resulted, blowing a mixture of pulverized rock, pyroclasts, and hot gases sideways as well as upward. The sideways blast, initially traveling at the speed of sound, roared across the landscape, killing David Johnson and others in the blast zone. Within the devastated area which extends as much as 30 km (19 mi) from the crater and covers some 600 km^2, (232 mi^2) trees were blasted to the ground and covered with hot debris.

Although Mount St. Helens provides the best documented recent example of a lateral blast, a closely similar 1956 eruption of Bezmianny volcano in Kamchatka (eastern Siberia) also produced a devastating lateral blast, a high eruption column, and associated pyroclastic flows.

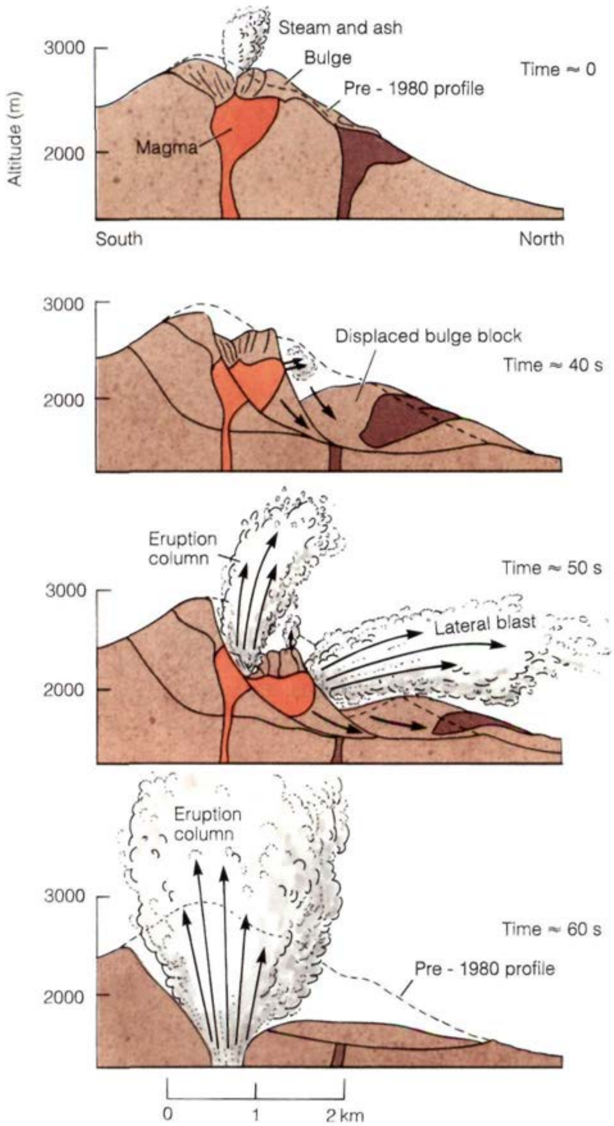


Figure 5.7 Sequence of events leading to the eruption of Mount St. Helens on May 18, 1980. Time approximately Os: Earthquakes and then puffs of steam and ash indicate that magma is rising: the north face of the mountain bulges alarmingly. Time approximately 40s: an earthquake shakes the mountain, and the bulge breaks loose and slides downward. This reduces the pressure on the magma and initiates the lateral blast that killed David Johnson. Time approximately 50s: the violence of the eruption causes a second block to slide downward, exposing more of the magma and initiating an eruption column. Time approximately 60s: the eruption increases in intensity. The eruption column carries volcanic ash as high as 19 km into the atmosphere.

Types of Volcanoes

The name *volcano* is applied to any vent hole or opening from which lava, pyroclasts, or gases are erupted. The volcanic shape has a lot to do with the kind of magma erupted and with the relative proportions of lava and pyroclasts. There are three shape-related terms more specific than the general term *volcano:* shield volcano, tephra cone volcano, and stratovolcano.

The kind of volcano that is easiest to visualize is one built up of successive flows of very fluid lava. Such lavas are capable of flowing great distances down gentle slopes and of forming thin sheets of nearly uniform thickness. Eventually, the pile of lava builds up a **shield volcano,** which is a broad, roughly dome-shaped formation with an average surface slope of only 5° near the summit and about 10° on the flanks (Fig.5.8). Shield volcanoes are characteristically formed by the eruption of basaltic lava; the proportions of ash and other pyroclasts are small. Hawaii, Tahiti, Samoa, the Galapagos, and many other oceanic islands are the upper portions of large shield volcanoes.

Rhyolitic and andesitic volcanoes tend to eject large quantities of pyroclasts and therefore to be surrounded by layers of tephra. As the debris showers down, a steep-sided volcano, composed entirely of tephra and therefore called a **tephra cone,** builds up around the vent (Fig. 5.9). The slope of the cone is determined by the size of the pyroclasts. Sunset crater in Arizona is a tephracone volcano.

Large, long-lived volcanoes, particularly those of andesitic composition, emit a combination of lava flows and pyroclasts. **Stratovolcanoes** are defined as steep conical mounds consisting of layers of lava and tephra. The volume of tephra may equal or exceed the volume of the lava. The slopes of stratovolcanoes, which may be thousands of meters high, are steep like those of tephra cones. The slope is about 30° near the summit of a stratovolcano and 6° to 10° at the base.

The beautiful, steep-sided cones of stratovolcanoes are among Earth's most picturesque sights (Fig. 5.10). The snow-capped peak of Mount Fuji in Japan has inspired poets and writers for centuries. Mount Rainier and Mount Baker in Washington and Mount Hood in Oregon are majestic examples in North America.

Many large volcanoes, especially shield and strato-

Figure 5.8 Mauna Kea, a 4200-m-high shield volcano on Hawaii, as seen from Mauna Loa. Note the gentle slopes formed by highly fluid basaltic lava. The view is almost directly north. A pahoehoe flow is in the foreground on the northeast flank of Mauna Loa.

A.

B.

Figure 5.9 Tephra cones. A. Two small tephra cones forming as a result of an eruption of andesitic lava in Kivu, Zaire. Arcs of lights are caused by the eruption of red-hot lapilli and bombs. B. Tephra cone in Arizona built from lapilli-sized basaltic tephra. Note the small basaltic lava flow coming from the base of the cone.

volcanoes, are marked near their summit by a large depression. This is a **caldera,** a roughly circular, steep-walled basin several kilometers or more in diameter. Calderas form after the partial emptying of a magma chamber in a volcanic eruption (Figs. 5.11 and 5.12). Rapid ejection of magma during an eruption can leave the magma chamber empty or partly empty. The now unsupported roof of the chamber slowly sinks under its own weight, like a snow-laden roof on a shaky barn, dropping downward on a ring of steep vertical fractures. Subsequent eruptions commonly occur along these fractures. Crater Lake, Oregon, occupies a circular caldera 8 km (5 mi) in diameter, formed after an explosive pyroclastic eruption about 6600 years ago (Fig. 5.12). The volcano that erupted has been posthumously named Mount Mazama.

Sometimes lava reaches the Earth's surface through a vent that is an elongate fracture in the crust rather than through a small circular opening. Extrusion of lava from an extended fracture is called *a.fissure erup-*

Figure 5.10 Mount Fuji, Japan, a snow-clad giant that towers over the surrounding countryside, displays the classic profile of a stratovolcano.

Figure 5.11 Crater Lake, Oregon, occupies a caldera 8 km in diameter that crowns the summit of a once lofty stratovolcano, posthumously called Mount Mazama. Wizard Island is a small tephra cone that formed after the collapse that created the caldera.

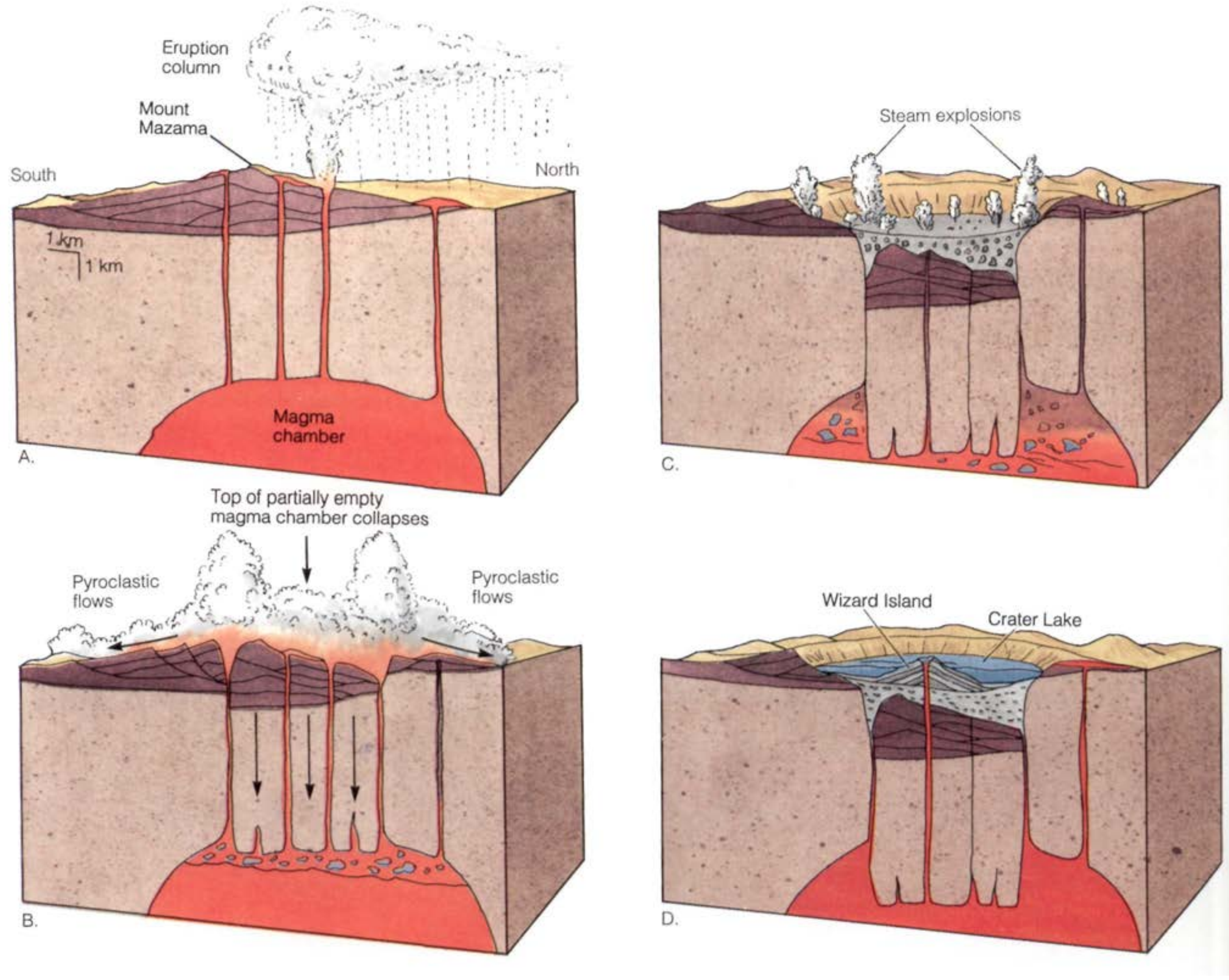


Figure 5.12 Sequence of events that formed Crater Lake following the eruption of Mount Mazama 6600 years ago. A. An eruption column of tephra rises from the flank of Mount Mazama. B. The eruption reaches a climax. Dense clouds of ash fill the air, and the hot pyroclastic flows sweep down the mountain side. C. The top of Mount Mazama collapses into the partly empty magma chamber, forming a caldera 10 km in diameter. D. During a final phase of eruption, Wizard Island formed. The water-filled caldera is Crater Lake, shown in Figure 5.11.

tion. (In this case the fissure is the volcano.) Such eruptions, which are often very dramatic, are characteristically associated with basaltic magma, and the lavas resulting from a fissure eruption on land tend to spread widely and to create flat lava plains.

The 1783 Laki eruption in Iceland that Benjamin Franklin identified as the cause of the unusually cold winter of 1783-1784 (see the opening essay of the Introduction) was a fissure eruption occurring along a fracture 32 km (20 mi) long. Lava flowed 64 km (40 mi) outward from one side of the fracture and nearly 48 km (30 mi) outward from the other side. Altogether it covered an area of 588 km^2 (227 mi^2). The volume of the lava extruded was 12 km^2 (4.6 mi^2), making this the largest lava flow in historic times. It was also one of the most deadly, destroying homes and food supplies and covering vast areas of farmlands. Famine followed and 9336 people died.

There is good evidence that larger fissure eruptions occurred in prehistoric times. The Roza flow, a great sheet of basaltic lava in eastern Washington State, can be traced over 22,000 km^2 (8500 mi^2) and shown to have a volume of 650 km^3 (156 mi^3).

Volcanic Hazards

Volcanic eruptions are not rare events. Every year about 50 volcanoes erupt somewhere on Earth. Eruptions of basaltic volcanoes are seldom dangerous, and basaltic eruptions are far more common than andesitic and rhyolitic eruptions. Though less common, pyroclastic eruptions from andesitic or rhyolitic stratovolcanoes, such as Mount St. Helens and Krakatau, do occur and can be disastrous, since millions of people live on or close to stratovolcanoes.

Eruptions of stratovolcanoes present five kinds of hazards:

1. Hot, rapidly moving pyroclastic flows *(nuée ardente)* and lateral blasts may overwhelm people before they can run away. The tragedies of Mount Peleé in 1902 and Mount St. Helens in 1980 are examples.
2. Tephra and hot, poisonous gases may bury people or suffocate them. Such a tragedy occurred in A.D. 79 when Mount Vesuvius, the supposedly dormant volcano in southern Italy, burst to life. First, hot, poisonous gases killed people in the nearby Roman city of Pompeii, and then tephra buried them (Fig. 5.13).
3. Tephra can be dangerous long after an eruption has ceased. Rain or melt water from snow can loosen tephra piled on a steep volcanic slope and start a deadly mudflow. In 1985, following a small and otherwise nonthreatening eruption of the Colombian stratovolcano Nevado del Ruiz, massive mudflows were formed by melting glaciers. The mudflows moved swiftly down the mountain and killed 20,000 people.
4. Violent undersea or coastal eruptions can cause tsunami. As noted earlier, set off by the eruption of Krakatau, tsunami killed more than 36,000 coast dwellers on Java and Sumatra.
5. A tephra eruption may wreak such havoc on agricultural land and livestock that people die from famine. Tephra can also overwhelm cities and other sites of human activities. A 1990 eruption of Mount Pinatubo in the Philippines, which caused the destruction of a nearby U.S. airforce base, is an example.

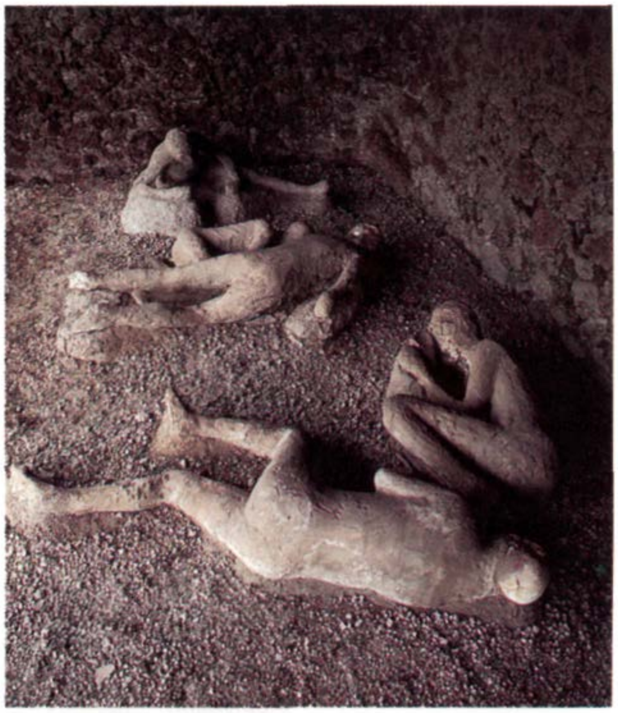


Figure 5.13 Evidence of an ancient disaster. Casts of bodies of five citizens of Pompeii, Italy, killed during the eruption of Mount Vesuvius in A.D. 79. Death was caused by poisonous gases, and then the bodies were buried by lapilli. Over the centuries, the bodies decayed, but the body shapes were imprinted in the tephra blanket. Excavators who discovered the imprints carefully recorded them with plaster casts.

Table 5.2
Volcanic Disasters Since A.D. 1800 in Which a Thousand or More People Lost Their Lives

			Primary Cause of Fatalities			
Volcano	**Country**	**Year**	**Pyroclastlc Eruption**	**Mudflow**	**Tsunami**	**Famine**
Mayon	Philippines	1814	1,200			
Tambora	Indonesia	1815	12,000			80,000
Galunggung	Indonesia	1822	1,500	4,000		
Mayon	Philippines	1825		1,500		
Awu	Indonesia	1826		3,000		
Cotopaxi	Ecuador	1877		1,000		
Krakatau	Indonesia	1883			36,417	
Awu	Indonesia	1892		1,532		
Soufriere	St. Vincent	1902	1,565			
Mt. Peleé	Martinique	1902	29,000			
Santa Maria	Guatemala	1902	6,000			
Taal	Philippines	1911	1,332			
Kelud	Indonesia	1919		5,110		
Merapi	Indonesia	1930	1,300			
Lamington	Papua-New Guinea	1951	2,942			
Agung	Indonesia	1963	1,900			
El Chichón	Mexico	1982	1,700			
Nevado del Ruíz	Colombia	1985		25,000		

Source: From a Report by the Task Group for the International Decade of Natural Disaster Reduction. Published in Bull. Volcano. Soc. Japan, Series 2, Vol. 35, #1, 1990, p. 80-95, 1990.

Since A.D. 1800 there have been 18 volcanic eruptions in which a thousand or more people died (Table 5.2). It is certain that other violent and dangerous eruptions will occur in the future. Likely candidates for dangerous eruptions in the United States are volcanoes in the states of Oregon, Washington, and Alaska. Potentially dangerous volcanoes are also to be found in Japan, the Philippines, New Guinea, New Zealand, Indonesia, the countries of Central and South America, and the Caribbean islands. To some extent, volcanic hazards can be anticipated, provided experts can gather data before, during, and after eruptions. The experts can then advise civil authorities when to implement hazard warnings and when to move endangered populations to areas of lower risk.

After an Eruption

When a volcanic eruption spreads lava or tephra across the land, a blanket of fresh new rock is found. In this manner, volcanism renews the land surface. Because volcanism occurs on each of the terrestrial planets, surface renewal by volcanism is an important planetary process.

Surface renewal is an especially important process for life on Earth. New rock means new supplies of the fertilizer elements needed by planets. Lava and volcanic ash, when subjected to weathering, produce very fertile soils. Some of the richest agricultural land in Italy, near Naples, has volcanic soil developed on tephra. Japan, the North Island of New Zealand, the Hawaiian islands, the Philippines, and Indonesia are other places where rich volcanic soils produce high agricultural yields.

It is remarkable how quickly the land recovers after an eruption. Within a year of the Mount St. Helens eruption, trees and other plants had started to sprout in the area devastated by the lateral blast of 1980; animals started to return to the area to graze as soon as the plants were big enough.

Agricultural land can also be worked very quickly after an eruption. Although the eruption of Mount Pinatubo caused great distress to the local population, local farmers planted crops in the volcanic ash as soon as the eruption stopped. Plants can even grow on recently erupted lava. In Hawaii, papaya trees have been reported to grow on basaltic lava within two years after the lava was erupted.

From the human viewpoint, volcanism has bad features and good. The bad are the hazards and dangers associated with eruptions and the effects of volcanic dust on the climate. The good is the production of **volcanic rock** from which rich volcanic soils develop.

INTRUSION OF MAGMA

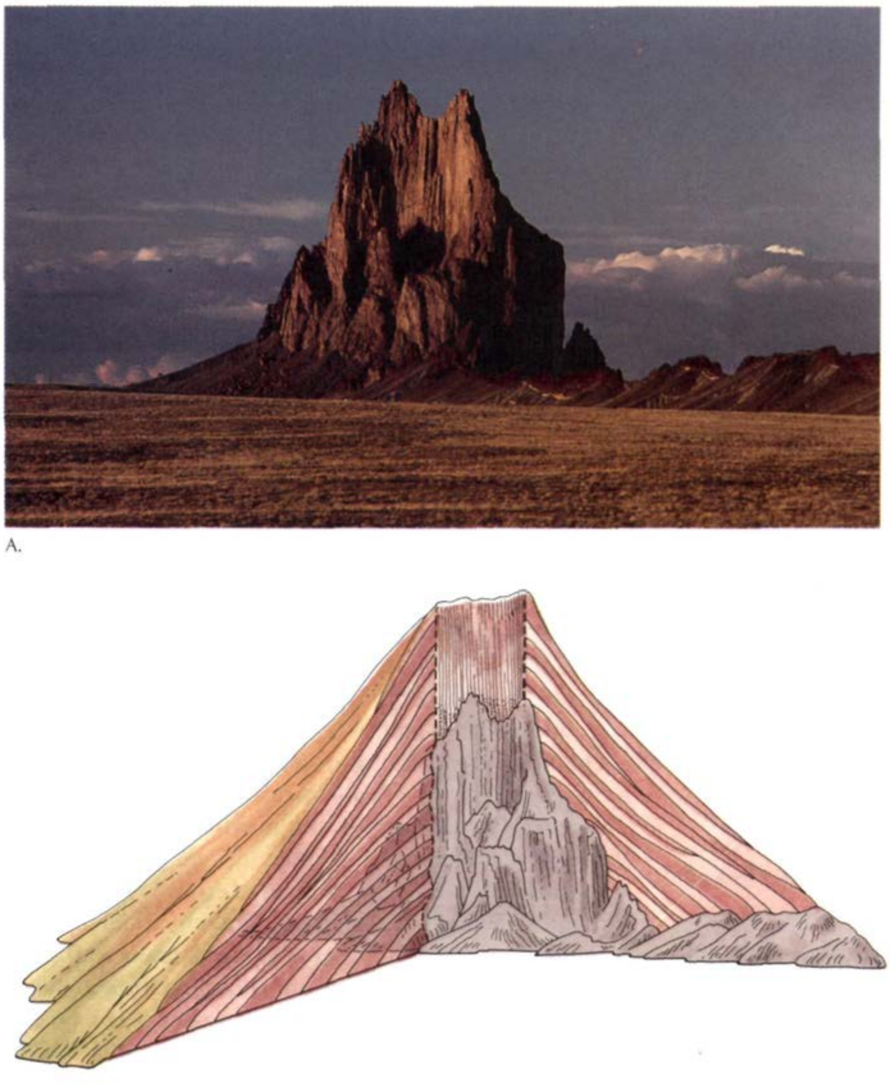
Now that we have considered what happens to magma when it erupts on the Earth's surface, we turn to the question of how magma works its way upward and how igneous rocks are formed from magma.

Beneath every volcano lies a complex of chambers and channelways through which magma reaches the surface. Naturally, we cannot study the magmatic channels of an active volcano, except at the earliest stages, when only heat and hot water are escaping. (See the "Guest Essay" at the end of this chapter for an example.) However, we can look at ancient cones that have been laid bare by erosion, as seen in Figure 5.14. What we find is that these ancient channelways are filled by igneous rock because they are the underground sites where magma solidified.

All bodies of what we call intrusive igneous rock, regardless of shape or size, are called plutons, after Pluto, the Greek god of the underworld. The magma that formed a pluton did not originate where we now find the pluton. Rather, the magma was squeezed upward from the place where it formed, thereby intruding the overlying rock (thus, the term *intrusive igneous rock*).

Plutons are given special names depending on shape and size (Fig. 5.15):

1. Dikes, sills, and laccoliths: tabular, parallel-sided



A.

B.

Figure 5.14 Shiprock, New Mexico. A. The conical tephra cone that once surrounded this volcanic neck has been removed by erosion. B. Diagram of the way the original volcano may have appeared prior to erosion.

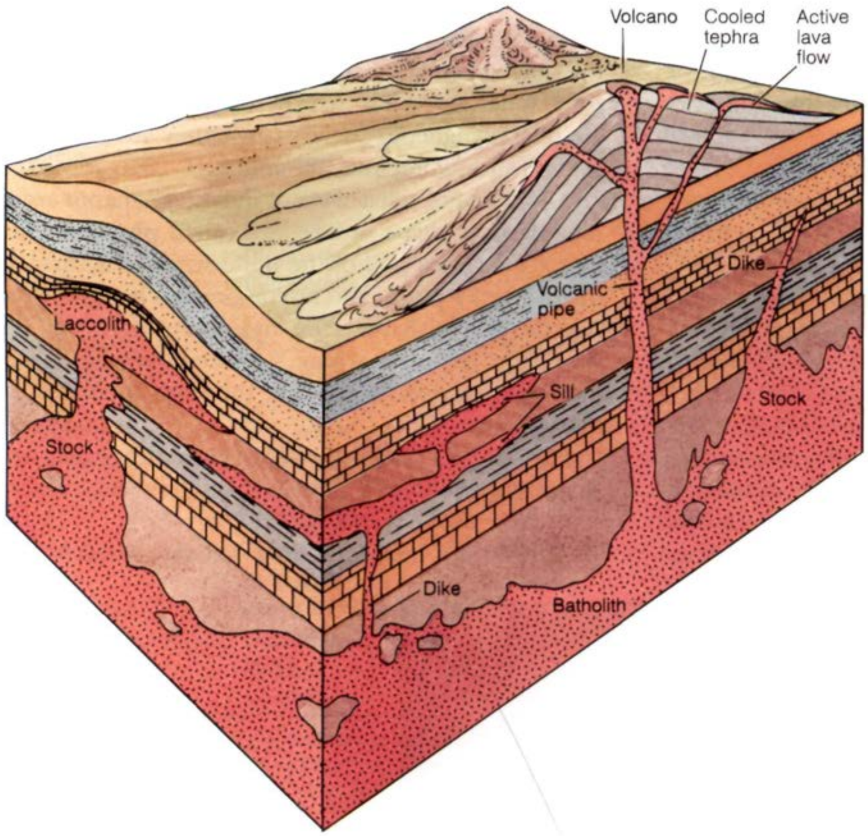


Figure 5.15 Diagrammatic section of part of the crust showing the various forms assumed by plutons.

sheets of igneous rock. **Dikes** cut across the layering of the intruded rock, **sills** are parallel to the layering, and **laccoliths** are sills that cause the intruded rocks to bend upward.

2. Volcanic pipes and necks: a **volcanic pipe** is a cylindrical conduit of igneous rock below a volcanic vent. A **volcanic neck** is a pipe laid bare by erosion (Fig. 5.14).
3. Batholiths and stocks: intrusive igneous bodies of irregular shape that cut across the layering of the intruded rock. **Stocks** are small (less than 10 km (6.2 mi) in maximum dimension), and **batholiths** are huge (up to 1000 km /621 mi in length and 250 km/155 mi wide).

THE ORIGIN OF MAGMAS AND IGNEOUS ROCKS

We come now to the most difficult but also two of the most interesting questions concerning magma, volcanoes, and igneous rocks: where do magmas form, and why are there three major kinds of magma (basaltic, andesitic, and rhyolitic)? A great many hypotheses have been offered over the years, but only two have stood the repeated test of investigation and experimentation.

The first hypothesis arose from the pioneering studies carried out by a Canadian scientist, Norman L. Bowen, early in the twentieth century. Bowen discovered by laboratory experiment that, as a magma cools, minerals crystallize in a specific sequence. Furthermore, he discovered that the first-formed minerals later react with the cooler, remaining magma to form different minerals. These reactions, he discovered, also follow a definite sequence. The sequence of reactions is now called **Bowen's reaction series.**

Bowen reasoned that his reaction series could account for the different magmas in the following way. Suppose that the only primary magma is basaltic magma, formed deep in the mantle, and suppose further that mineral grains, once formed in this primary magma, are somehow separated from the remaining magma. What kind of rock, Bowen asked himself, might form from the separated minerals, and what kind of magma might remain? He called the separation process *fractional crystallization* (Fig. 5.16), and he used his laboratory research to demonstrate that, by separating minerals and magma at different stages in the crystallization and reaction process, a single

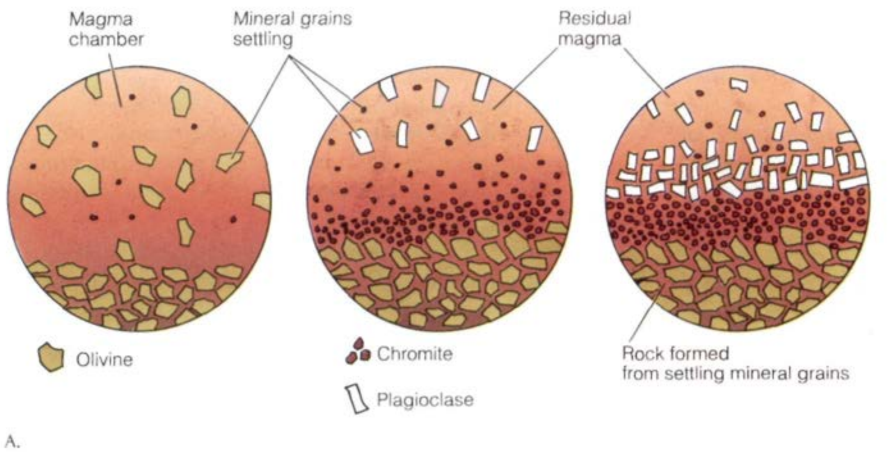


Figure 5.16 Fractional crystallization. A. Grains of three minerals—olivine, chromite (chromium-iron oxide), and feldspar—settle one after the other to the bottom of a magma chamber, producing three types of rocks whose compositions differ considerably from that of the parent magma. B. Layers of plagioclase (light gray) and chromite (black) formed by fractional crystallization in the Bushveld Igneous Complex, South Africa.

magma could be changed from basaltic to andesitic and rhyotitic, as shown in Figure 5.17.

The second hypothesis, for the origins of the three major magma types, which arose through the work of many people, is the reverse of the Bowen hypothesis. Instead of seeking the answer in cooling processes (the Bowen approach), they looked at melting processes—*fractional melting* instead of fractional crystallization. As so often happens, both hypotheses turned out to be right in part. As discussed below, fractional melting does indeed seem to be the reason there are three major kinds of magma. However, many of the subtle variations in igneous rocks are best explained by the reaction series identified by Bowen. (See "A Closer Look: Naming Igneous Rocks.")

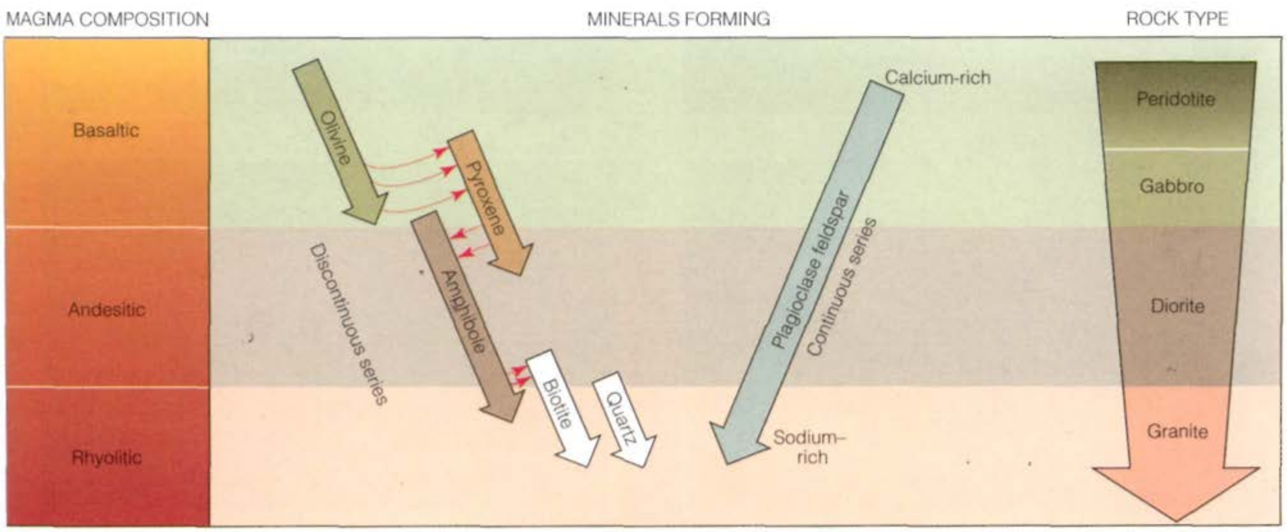


Figure 5.17 Bowen's reaction series demonstrates how the cooling and crystallization of a primary magma of basaltic composition, through reactions between mineral grains and magma followed by separation of mineral grains and magma, can change from basaltic to andesitic to rhyolitic. Bowen identified two series of reactions: A *continuous series* in which one mineral, feldspar, changes from an initial calcium-rich form to a sodium-rich one; and a *discontinuous series* in which minerals change abruptly—for example, from olivine to pyroxene.

A Closer Look

Naming Igneous Rocks

Igneous rock forms by the cooling and solidification of magma. **Extrusive igneous rocks** are those formed by solidification of lava; **intrusive igneous rocks** are those formed when magma solidifies within the crust or mantle. Both extrusive and intrusive igneous rocks are classified and named on the basis of rock texture and mineral assemblage.

Naming by Texture

The most obvious textural feature of an igneous rock is the size of its mineral grains. Lava cools so rapidly that mineral grains do not have sufficient time to become large. As a result, extrusive igneous rocks are fine-grained (individual grains less than 2 mm (0.08 in) in diameter). Some lavas cool so rapidly that the rocks they form are glassy. Figures C5.1 A and B are examples of a glassy and a fine-grained igneous rock, respectively.

Intrusive igneous rock tends to be coarse-grained because magma that solidifies in the crust or mantle cools slowly and has sufficient time to form large mineral grains. Figure C5.1C is an example of a coarse-grained igneous rock, respectively.

One special texture involves a distinctive mixture of large and small grains. Rock of such a texture is called a **porphyry,** meaning an intrusive igneous rock consisting of coarse mineral grains scattered through a mixture of fine mineral grains, as shown in Figure C5.1D, like raisins in rice pudding. The large grains in a porphyry are formed in the same way that those of any other coarse-grained igneous rocks are formed—by slow cooling of magma in the crust or mantle. The fine-grained mass, called groundmass, that encloses the coarse grains provides evidence that partly solidified magma moved quickly upward. In the new setting, the magma cooled rapidly, and as a result, the later mineral grains are all tiny.

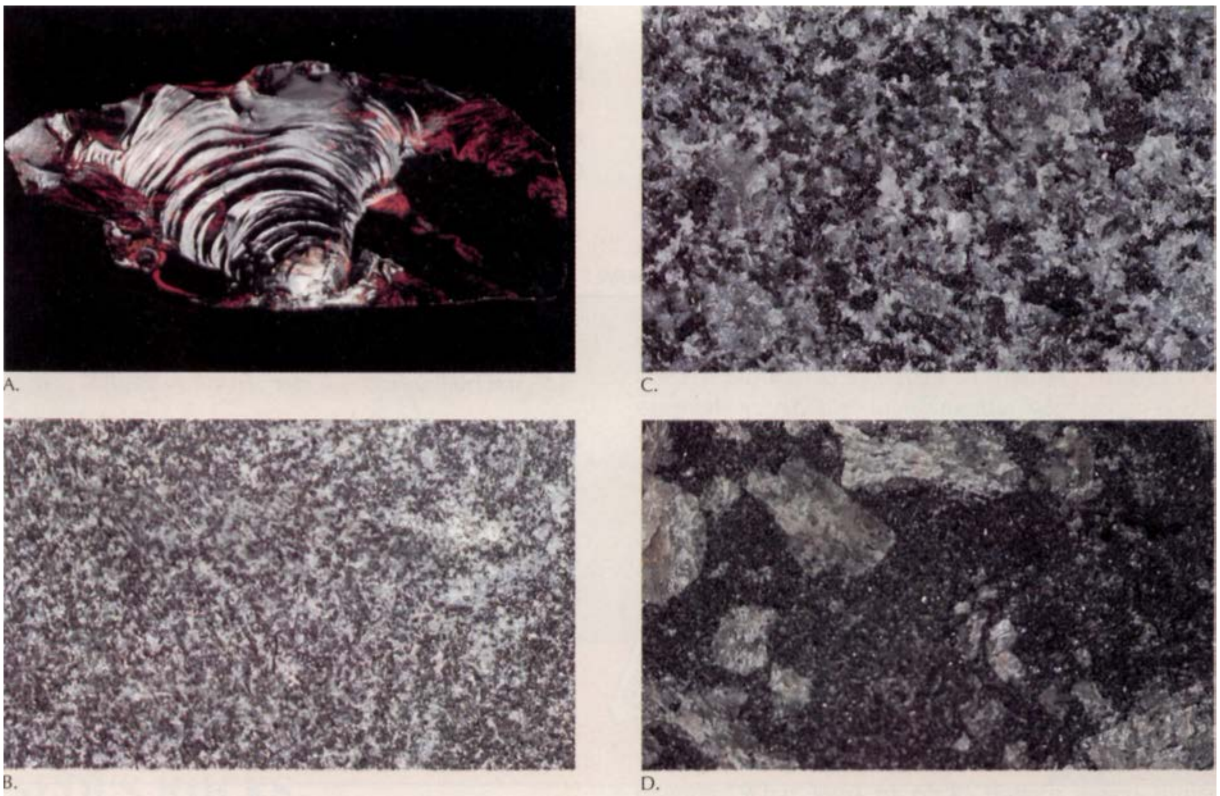


Figure C5.1 Different textures in igneous rock. A. Obsidian, a wholly glassy igneous rock (extrusive). B. Basalt, a fine-grained igneous rock (extrusive). C. Gabbro, a coarse-grained igneous rock (intrusive). D. Basalt porphyry (extrusive). Sample A has the composition of a rhyolite (Table C5.1), but B, C, and D have the same mineral assemblage—feldspar (white), pyroxene (dark green to black), and olivine (pale brown).

Naming by Mineral Assemblage

When magma or lava of a given composition solidifies, the mineral assemblage that forms is the same for intrusive and extrusive rocks; the only differences are textural. Once the texture of an igneous rock has been determined, therefore, specimens are named on the basis of mineral assemblage, as shown in Table C5.1 and Figure C5.2.

All common igneous rocks are composed of one or more of these six minerals or mineral groups: quartz, feldspar, mica (both muscovite and biotite), amphibole, pyroxene, and olivine. Although mineral assemblages are gradational, the common igneous rocks can be divided into four families based on one key feature: the presence or absence of quartz and olivine (Table C5.1).

Table C5.1
Mineral Assemblages of Common Igneous Rocks

		Name[a]	
Key Feature	**Mineral Assemblage**	**Coarse (intrusive)**	**Fine (extrusive)**
Quartz yes, olivine no.	Quartz, feldspar, muscovite, biotite, amphibole	Granite	Rhyolite
Quartz no, olivine no.	Feldspar, amphibole, pyroxene, biotite	Diorite	Andesite
Quartz no, olivine yes.	Feldspar, pyroxene, olivine, biotite	Gabbro	Basalt
Quartz no, olivine yes.	Olivine (abundant), pyroxene, feldspar	Peridotite	(none)

[a] When a rock has a porphyritic texture, we use the name determined by the mineral assemblage as an adjective and the term for the texture of the groundmass as the noun. For example, if the groundmass is fine, we call it a rhyolite porphyry; if the groundmass is coarse, we call it a granite porphyry.

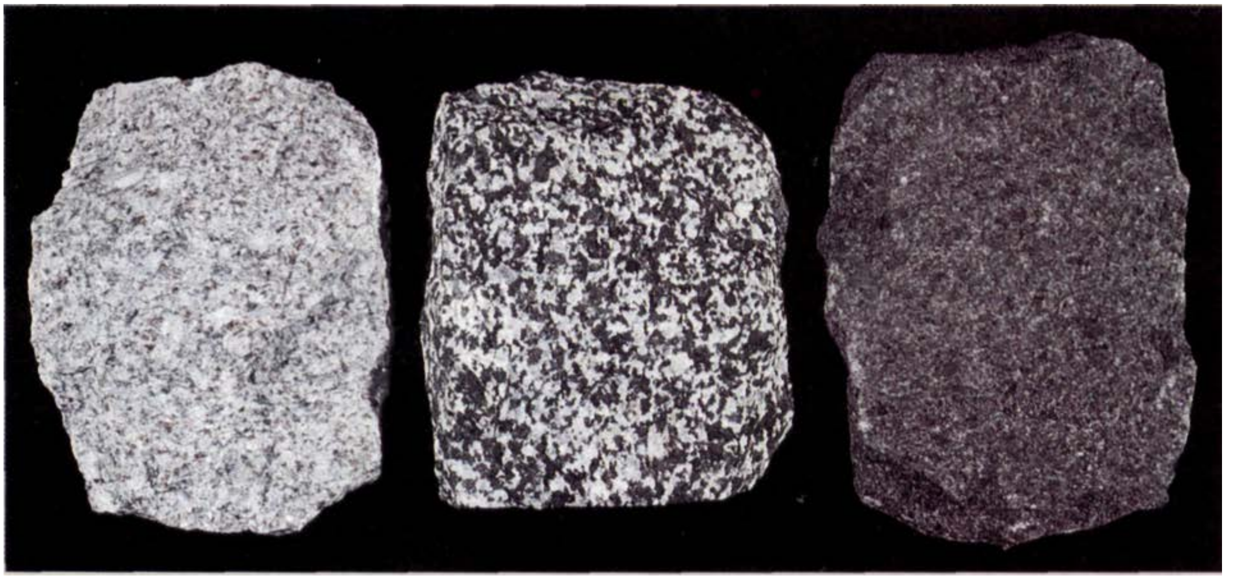
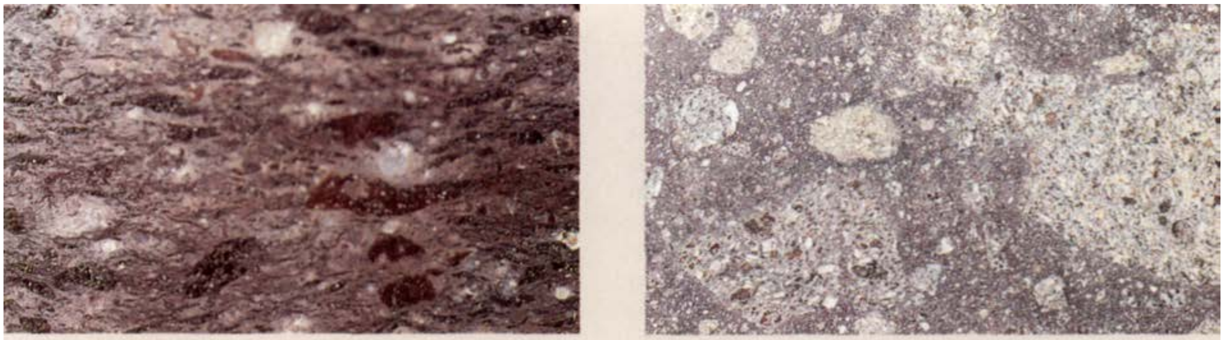


Figure C5.2 Three coarse-grained igneous rocks. Compare their mineral assemblages by using Table C5.1. Note the change in color from granite (left), which is light colored because it is rich in feldspar and quartz, through diorite (center), to gabbro (right), which is quartz-free and rich in pyroxene and olivine and therefore darker in color. Each specimen is 7 cm across.



A. B.

Figure C5.3 Two ways of forming a pyroclastic rock. A. Rhyolite tuff, formed by cementation of lapilli and ash, from Clark County, Nevada. B. Welded tuff from the Jemez Mountains, New Mexico. The dark patches are glassy fragments flattened during welding. Note the fragments of other rocks in the specimen. Both samples are 4 cm across.

Varieties of Pyroclastic Rocks

There is an old saying that pyroclasts are igneous on the way up and sedimentary on the way down. As a result, pyroclastic rocks are transitional between igneous and sedimentary. They are called **agglomerates** when tephra is bomb sized, or **tuffs** when the pieces are small, either lapilli or ash. The igneous origin of a pyroclastic rock is indicated by the name for the mineral assemblage. For example, we refer to a rock of appropriate mineral assemblage as an andesite tuff.

Tephra is converted to pyroclastic rock in two ways. The first, and most common, way is through the addition of a cementing agent, such as quartz or calcite introduced by groundwater. Figure C5.3A is an example of a rhyolitic tuff formed by cementation. The second way tephra is transformed to pyroclastic rock is through the welding of hot, glassy, ash particles. When ash is very hot and plastic, the individual particles can fuse together to form a glassy pyroclastic rock. Such a rock is called **welded tuff** (Fig. C5.3B).

Uses of Igneous Rock

Igneous rocks have many uses. Granites, diorites, and gabbros are very attractive when polished and are widely used for facing buildings and for ornamental purposes. Basalt is a very tough rock, and fragments of crushed basalt are widely used as the aggregate in concrete and for roadways. Crushed basalt of commerce is sometimes called trap rock. Rhyolite and andesite are sometimes used for roadways, but when rhyolite is glassy (obsidian) it is also used for jewelry and other ornamental purposes.

Fractional Melting and Magma Types

Evidence supporting the hypothesis that the three types of magma originate through fractional melting comes from both laboratory experiment and the distribution of the different kinds of volcanoes on our planet. A close relationship exists between plate tectonics and the volcano locations; a summary of present thinking about the distribution is presented in Figure 5.18 as well as in the following discussion, which should be read with frequent references to Figure 5.18.

It has long been known that volcanoes that erupt rhyolitic magma are abundant on the continental crust but are not known on the oceanic crust. This observation suggests that the processes that form rhyolitic magma do not occur in the mantle and must be confined within the continental crust. (If the processes that form rhyolitic magma did occur in the mantle, the rhyolitic magma would rise to the surface regardless of the kind of crust above and therefore be found in volcanoes on oceanic crust.)

Volcanoes that erupt andesitic magma are found on both the oceanic crust and the continental crust. This suggests that andesitic magma must form in the mantle and rise up regardless of the nature of the overlying crust. However, andesitic volcanoes have a restricted geographic distribution. For example, as shown in Figure 5.19, a ring of andesitic volcanoes surrounds the Pacific, forming the so-called Ring of Fire. The Ring of Fire, which geologists also call the *Andesite Line,* is exactly parallel to the plate subduction margins, the places where lithosphere capped by oceanic crust sinks down into the asthenosphere.

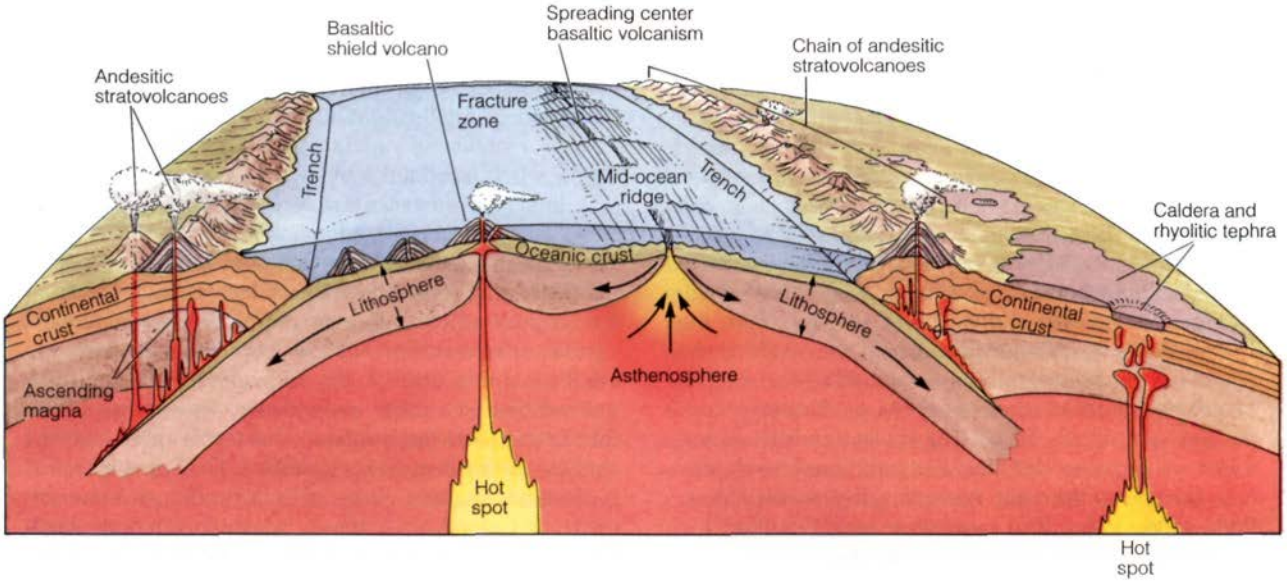


Figure 5.18 Diagram illustrating the locations of the major kinds of volcanoes in a plate tectonic setting.

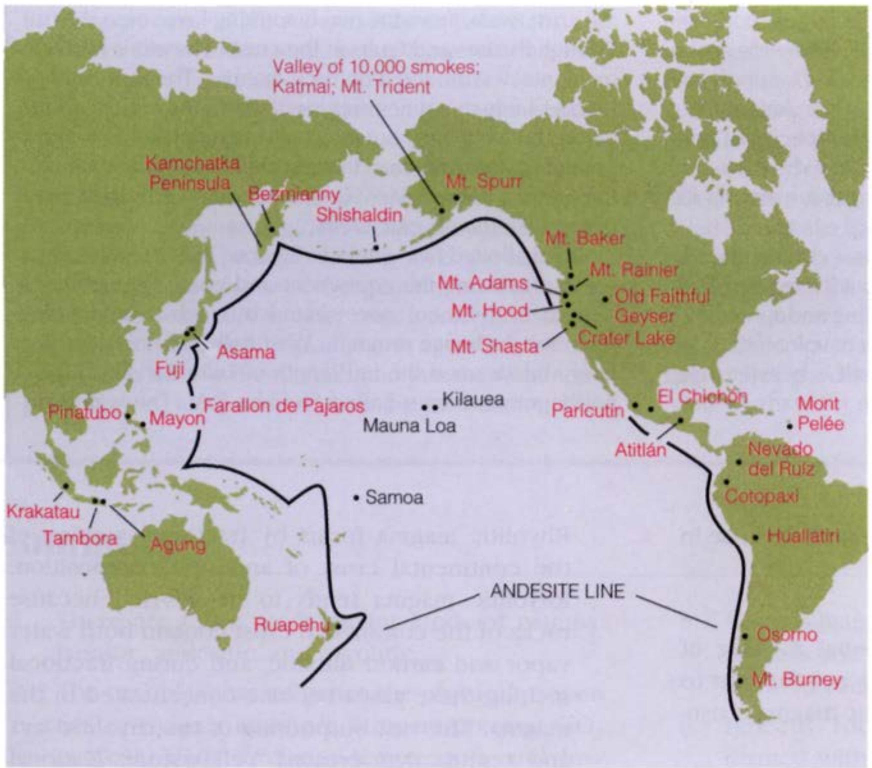


Figure 5.19 The Ring of Fire around the Pacific Ocean basin is formed by andesitic volcanoes. The Ring of Fire is coincident with subduction zones where lithosphere capped with oceanic crust is being subducted into the asthenosphere. Volcanoes in the Pacific Ocean, such as Mauna Loa, erupt basaltic magma but not andesitic magma.

This suggests that (1) andesitic magma forms as a consequence of plate tectonics, and (2) andesitic magma must be formed by the fractional melting of subducted oceanic crust.

Volcanoes that erupt basaltic magma also occur on both the oceanic and the continental crust. The source of basaltic magma, like the source of andesitic magma, must, therefore, be in the mantle. The geographic distribution of basaltic volcanoes does not, however, seem to be related to specific features of the crust or plate tectonics, such as subduction zones. This suggests that basaltic magma must be formed by the melting of the mantle itself and that the magma must rise up regardless of what lies above. The most likely hypothesis for the generation of basaltic magma, therefore, is that it is formed by the melting of the mantle as a result of deep-seated convection currents.

Guest Essay

Lake Baikal, One of the World's Underwater Wonders

Kathleen Crane received a B.A. in geology from Oregon State University and a Ph.D. in oceanography from the Scripps Institute of Oceanography. She has conducted significant research on seafloor hydrothermal vents and continental rifts, dived in the Alvin and the Russian Mirs, utilized deep-sea sonar and video systems, searched for the *Titanic,* and led multinational teams of researchers studying Arctic heat flow. At present, Dr. Crane serves as professor at Hunter College of CUNY, adjunct senior scientist at Columbia University's Lamont-Doherty Earth Observatory, and visiting senior scientist at the Naval Research Laboratory, where she is completing an Arctic Environmental Atlas.

In September 1989 the Soviet Academy of Sciences extended an invitation to the National Geographic Society to orgainize the first major visit of scientists from the United States to the deepest lake in the world, Lake Baikal. In the geological community there had been a heated debate about whether or not Lake Baikal was rifting and volcanically active. The Soviet Academy of Sciences wanted me to lead this visit and to resolve the debate one way or the other. Preferably, they wanted me to find evidence for heat, and that meant finding hydrothermal vents on the bottom of the deepest lake in the world. This seemed an awesome task. I had no maps; a decent echosounding chart of the lake had never been produced.

Located 200 kilometers north of Mongolia in remotest Siberia, the lake is the world's largest body of fresh water, 636 kilometers long and 1640 meters deep, and contains about one-fifth of the world's supply of fresh water, excluding that locked up in the Antartic ice sheet. It also contains an estimated 1700 species of animals and plants, many of them found nowhere else in the world. Fresh water seals and sponges are among its unusual inhabitants.

Geophysicists suggest that Baikal is a continental rift that is spreading apart at a rate of about 2 cm/year. This is how oceans form, by continental rifting and spreading. Involved in this process is the injection of volcanic basalt deep into the rift valley. Because basalt is heavier than the neighboring continental rock, the rift starts to sink and water collects. During the cracking and stretching of the crust, water from the newly forming lake seeps down through fissures and faults in the crust, where it is heated by contact with the hot rock below. The hot water leaches large quantities of minerals from the molten rock before spewing back out the vents into the lake. Through time, if rifting continues, the lake will become a sea (like the young Norwegian-Greenland Sea) and then an ocean (like the Atlantic Ocean).

I had allotted five weeks in Siberia. The first week we spent organizing the equipment and going through the diplomatic protocol necessitated by such a landmark scientific exchange program. We finally left port a week later and steamed the full length of Lake Baikal, to the northernmost corner called Frolikha Bay. The water in

The three magma types are now thought to arise in the following ways:

1. Basaltic magma forms by fractional melting of rock in the mantle. The mantle is not thought to be gas-rich, and that is why basaltic magma is usually gas-poor.
2. Andesitic magma forms by the fractional melting of oceanic crust (which is basaltic in composition) once that crust has been subducted into the mantle. Oceanic crust is in contact with seawater, and some of the water is carried down with the subducted crust. When fractional melting of this subducted oceanic crust starts at a depth of 80 to 100 km (50 to 62 mi), gas-rich (mainly water vapor) andesitic magma is the result.
3. The continental crust is andesitic in composition. Rhyolitic magma forms by fractional melting of the continental crust of andesitic composition. Rhyolitic magma tends to be gas-rich because rocks of the continental crust contain both water vapor and carbon dioxide, and during fractional melting these gases become concentrated in the magma. The vast outpouring of the rhyolitic lava and tephra that created Yellowstone National Park, for example, apparently occurred because the base of the continental crust was heated by basaltic magma in the mantle. The heating caused fractional melting and the production of rhyolitic magma from the andesitic continental crust. Scientists are now testing the hypothesis that the entire continental crust may have formed through a sequence of fractional melting processes—first to form a basaltic oceanic crust and then to form andesitic continental crust by subduction.

this lake is the clearest in the world. Horizontal visibility surpasses 28 meters! Surrounding its banks are dense forests called the Taiga and Mongolian-like Steppes, home of the famous Ghengis Khan. Our survey was in a water depth of 365 to 640 meters, but because the walls of the lake were so steep we were very close to land.

After a week of photographing mud and measuring subtle temperature anomalies, we chanced upon a temperature spike of 0.1°C, which is a very significant change in deep-water characteristics. At the same time, our underwater camera showed mats of luminescent white bacteria covering an area the size of two football fields. Needless to say, we were jubilant. In addition to the thick mats of bacteria, which signify that hot water laden with sulfur is bubbling out from below (a telling characteristic of heat vents), we detected numerous white encrusting sponges.

Six hundred and forty kilometers away, Emory Kristof and Mikhail Grachev, the director of the Soiviet Limnological Institute, anxiously awaited news from our team. In the event that hydrothermal vents were to be found, they had convinced the Shirsov Institute of Oceanology to transport two PISCES submersibles to Lake Baikal. It had taken two weeks to ready one of these subs for work in fresh water. (Fresh water is lighter than saltwater, so a submersible would sink like a stone in it unless the buoyancy were changed.)

I had brought with me a new heat-flow instrument developed by the Woods Hole Oceanographic Institution for use from a submersible. This would allow us to quantify just how much heat was venting out of the bottom of Frolikha Bay. *The National Geographic* attached video and still cameras to the other manipulator arm of the submersible. Detailed measurements revealed that the vent area is laden with minerals and that the water temperature was 13°C warmer than the surrounding lake water. The area registers more than 900 times the normal Earth heat flow!

Whether this vent community consists of vent-specific animals or is merely a dense concentration of background lake fauna is not known, because the deeper parts of Lake Baikal are so poorly mapped. However, 85 percent of the photographs suggest that no similar concentrations of organisms or sponges with the same growth form are seen in areas removed from the vent field. Many of the organisms have thrived on an unprecedented scale, forming a multitude of new species and genera, and many of the animals strangely resemble ocean species. How did they evolve in this lake? Was it once connected to the sea, or has life evolved separately in this freshwater environment? This question is of incredible scientific significance. How the animals inhabiting the Lake Baikal hydrothermal vents fit into the biogeographical arena of vent communities needs more detailed investigation.

Furthermore, the discovery of the first deep-freshwater venting in Lake Baikal suggests that the Baikal Rift may be very active and that the rift is gradually turning into an ocean. In addition, the vents give us a chance to investigate the evolution of life forms that are not dependent on the sun for their energy.

Unlocking the treasures in Lake Baikal helps us to see the biological diversity built up over millions of years of evolution. Scientists and environmentalists within Siberia hope that the Western world will become educated about Lake Baikal so that an international ecological center can be established on its shores, protecting the life within from the onslaught of twentieth-century pollution and ignorance. It has now been declared one of the seven underwater wonders of the world, and this is a step in the right direction.

Summary

1. There are three predominant kinds of magma: basaltic, andesitic, and rhyolitic.
2. The variables that influence the physical properties of magma most are temperature and SiO_2 content. High formation-temperature and low SiO_2 content result in fluid, nonviscous magma (basaltic). Lower formation-temperature and high SiO_2 content result in viscous magma (andesitic and rhyolitic).
3. Volcano sizes and shapes depend on the kind of material erupted, viscosity of the lava, and explosiveness of the eruptions.
4. Low-viscosity magma, low in SiO_2, erupts as fluid lavas that build gently sloping shield volcanoes.
5. Viscous magma, rich in SiO_2, erupts mainly as pyroclasts and builds steep sided tephra cones or stratovolcanoes.
6. Igneous rock may be intrusive (meaning it formed within the crust) or extrusive (meaning it formed on the surface). The texture and grain size of igneous rock indicate how rapidly, and where, the rock cooled.
7. Basaltic magma forms by fractional melting of rock in the mantle. Andesitic magma forms during subduction by fractional melting of basalt in oceanic crust. Rhyolitic magma forms by fractional melting of rock in the continental crust.
8. Two opposing hypotheses have been proposed

to explain the origin of the three predominant magmas: Bowen's hypothesis called on fractional crystallization of a single-parent magma, and the opposing hypothesis calls on fractional melting. Of the two, the fractional melting hypothesis is the more widely accepted.

Important Terms to Remember

Terms in italic are defined in A Closer Look

agglomerate (p. 130)
andesite (p. 114)
basalt (p. 114)
batholith(p. 126)
Bowen's reaction series (p. 126)
caldera(p. 121)
dike (p. 126)
diorite (p. 114)
extrusive igneous rock (p. 128)
gabbro (p. 114)
granite (p. 115)
intrusive igneous rock (p. 128)
laccolith (p. 126)
lava (p. 114)
magma (p. 114)
pluton (p. 125)
porphyry (p. 128)
pyroclast (p. 117)
pyroclastic rock (p. 117)
rhyolite (p. 115)
shield volcano (p. 120)
sill (p. 126)
stock (p. 126)
stratovolcano (p. 120)
tephra(p. 117)
tephra cone (p. 120)
tuff (p. 130)
viscosity (p. 115)
volcanic neck (p. 126)
volcanic pipe (p. 126)
volcanic rock (p. 125)
volcano (p. 114)
welded tuff (p. 130)

Questions for Re\1ew

1. What's the difference between magma and lava? Between lava and pyroclasts? Between pyroclasts and tephra?
2. Is the major oxide component of magma SiO_2, MgO, or $A1_2O_3$? Describe the effect of SiO_2 content on magma fluidity. What effect does temperature have on viscosity?
3. Where inside the Earth does basaltic magma form? What does the term *fractional melting* mean, and what role does it play in the formation of basaltic magma?
4. What is the origin of andesitic magma? With what kind of volcanoes are andesitic eruptions associated?
5. What is the origin of rhyolitic magma? Where do you expect to find rhyolitic volcanoes?
6. What are the differences between a shield volcano and a stratovolcano? Between a tephra cone and a stratovolcano?
7. The island of Moorea on the cover of this book is a volcanic island in the Pacific Ocean. What kind of volcano formed Moorea?
8. How might it be possible for fractional crystallization to produce more than one kind of igneous rock from a single magma?
9. Why does a shield volcano like Mauna Loa in Hawaii have a gentle surface slope, while a stratovolcano such as Mount Fuji in Japan has steep sides?
10. How do calderas form?
11. Name some ways in which volcanoes affect life on earth.
12. Why are some volcanic eruptions violent and others not?

13- What is the difference between a dike and a sill? A batholith and a stock?

Questions for A Closer Look

1. What are the distinguishing features of pyroclastic rocks? How might you tell the difference between a rhyolite that flowed as a lava and a rhyolitic tuff formed as a result of a pyroclastic eruption?
2. How does a welded tuff differ from a cemented tuff? Could both rocks form as a result of eruptions from the same volcano?
3. If you were vacationing near Mount St. Helens and picked up an igneous rock, what would you name a sample that had the following qualities?

Texture—fine-grained; mineral assemblage—feldspar 1 amphibole 1 pyroxene 1 biotite.

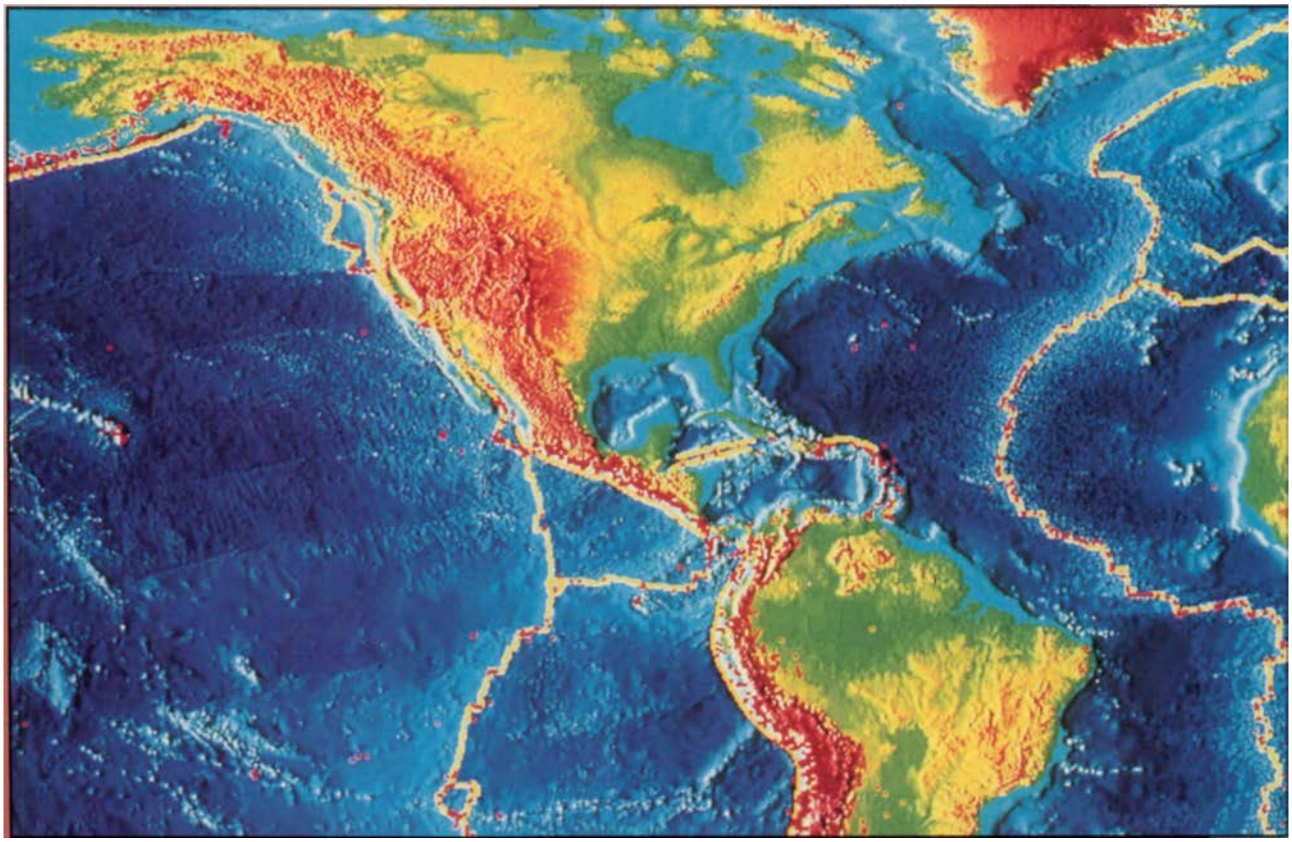
4. On another vacation, this time in Hawaii, you find an igneous rock with the following characteristics: texture—a porphyry, ground-mass coarse-grained; mineral assemblage—feldspar 1 pyroxene 1 olivine 1 biotite. What name would you give to the rock?

Questions for Discussion

1. Spaceships have landed on Venus, Mars, and the Moon. In each case, basaltic igneous rocks have been found, but rhyolitic or andesitic rocks, common on the Earth, have not been found. What hypothesis can you suggest about the evolution of Venus, Mars, and the Moon to explain this observation? Can you suggest why the Earth seems to be so different?
2. There are two different hypotheses to explain the sudden disappearance of the dinosaurs. Both suggest that some massive event (or sequence of events) so changed the atmosphere that large animals like dinosaurs could not survive. One hypothesis is that the impact of a giant meteorite was the cause; the other hypothesis is that a series of great volcanic eruptions was the culprit. Suggest ways by which the two hypotheses could be tested.
3. Do some research to see if Mount Vesuvius is still an active volcano. When did it last erupt, and what danger did it pose for people who live near the volcano? How many people live in the area?

CHAPTER 6

The Principles of Plate Tectonics



The Americas and the adjacent oceans. Yellow bands are mid-ocean ridges. Shape of the eastern Coasts of North and South America were determined by the breakup of a giant continent (Pangea) about 200 million years ago. The line along which Africa and Europe broke apart from tire Americas is the Mid-Atlantic Ridge. After breakup, the Americas moved west, Europe and Africa moved east.

Do We Drift or Don't We? That Was the Question.

In 1508, Leonardo da Vinci, recognized that some fossils he had collected were the remains of seashells. He immediately understood the significance of his observations. The rocks in which the fossils were embedded, now high in the mountains of Italy, must once have been on the seafloor; either the world, including mountains, had once been covered by the sea, or the seafloor must have been locally uplifted. Because fossil seashells are not found everywhere across the land surface, Leonardo realized that seafloor uplift must have taken place. Three centuries later James Hutton incorporated the idea of uplift into the concept of the rock cycle; uplift, he explained, raises rocks and exposes them to weathering. Charles Darwin recorded important evidence in favor of uplift when, on his voyage in the *Beagle* during the 1830s, he observed that part of the coastline of Chile had been raised up as a result of a great earthquake. By the middle of the nineteenth century, the concept of vertical movement (uplift) was widely accepted. A little bit of evidence—mainly the parallelism of the coastlines on either side of the Atlantic Ocean—suggested that lateral (sideways) movement was possible, but at that time the idea of moving continents (continental drift, as it came to be called) was too much for most scientists to accept.

In a paper published in 1910, Frank B. Taylor, an American geologist, offered the hypothesis that lateral movements do occur, and he suggested that two ancient continental masses, one over the South Pole and the other over the North Pole, had broken up and that the pieces had slid slowly to their present sites. This hypothesis, he argued, provided a neat explanation for the formation of mountain ranges and oceanic mountain ranges (which we now know to be mid-ocean ridges).

A more persuasive, and in the long run more important, supporter of continental drift was Alfred Wegener, a German scientist. Wegener, who was unaware of Taylor's work, published a book in 1914 in which he tried to explain such phenomena as the parallelism of the Atlantic coastlines and the fact that similar plant and animal fossils could be found on different continents. At some time in the distant past, Wegener suggested, all of the world's landmasses were formed together in a single huge continent, so that plants and animals could spread freely. To this ancient, huge continent Wegener gave the name Pangaea (pronounced Pan-jee-ah, meaning all lands). According to Wegener's hypothesis, Pangaea was somehow disrupted and its fragments (the continents of today) slowly drifted to their present positions. Proponents of the theory likened the process to the breaking up of a sheet of ice that floats in a pond. The broken pieces, they argued, should all fit back together again, like pieces of a jigsaw puzzle.

Although Wegener presented impressive evidence that continental drift may have happened, the hypothesis was not widely accepted in his lifetime because no one could explain how a solid, rocky continent could possibly overcome friction and slide across the oceanic crust. The process, said his critics, was like trying to slide two sheets of coarse sandpaper past each other.

Wegener died in 1930. Debate about continental drift slowed down because some of the supporting evidence Wegener had gathered was found, on close examination, to be explainable in other ways, and still no one was able to explain the way drift occurred. By 1939, when the Second World War broke out, the continental drift hypothesis had few supporters, and most of these few lived in the southern hemisphere where some of the best evidence was to be found. Discoveries about the Earth's magnetism made possible by the technologies spawned by the war effort revived the debate on a worldwide basis in the 1950s and eventually led to the hypothesis of plate tectonics and an explanation of how continents can move. Wegener was right, the continents do move, but not for the reasons he proposed—they move, rather, because of plate tectonics.

THE SOLID EARTH AND PLATE TECTONICS

The solid Earth, the largest of the four reservoirs of the Earth system, may seem to be constant and unchanging, but nothing could be further from the truth. This constancy is an illusion. Measurements show that the solid Earth is ever changing, that mountains are slowly rising, continents are drifting, and ocean basins are continually changing their shapes and sizes. Plate tectonics, it is now recognized, is the principal mechanism by which such changes happen to the face of the Earth. Through plate tectonics the Earth's surface is slowly but continually renewed. Understanding plate tectonics is the key to understanding the solid Earth's role in the Earth system.

MAGNETISM AND THE REVIVAL OF THE CONTINENTAL DRIFT HYPOTHESIS

Wegener's hypothesis of continental drift was largely abandoned by the 1940s because the movement of an entire continent seemed to be physically impossible. However, as so often happens in science, the hypothesis was revived as a result of accidental discoveries in another field—studies of the Earth's magnetism.

The source of the Earth's magnetism lies in the molten outer core. As a result of the Earth's rotation, the molten iron of the outer core flows continually around the solid inner core. The flowing stream of molten iron causes an electrical current to flow in the outer core, and the electrical current in turn creates the magnetic field. Because the magnetism is a result of the Earth's rotation, the north magnetic pole is close to the North Pole and the south magnetic pole is close to the South Pole.

Because the Earth is a gigantic magnet, it is surrounded by an invisible magnetic field that permeates everything placed in the field. If a small magnet is allowed to swing freely in the Earth's magnetic field, the magnet will become oriented so that its axis points to the Earth's magnetic north pole (Fig. 6.1). This is true for all places on the Earth, all free-swinging magnets will point to the north magnetic pole.

Certain rocks became permanent magnets as a result of the way they formed, and like free-swinging magnets they point to the Earth's north magnetic pole. Investigation of the properties of natural magnetism in rocks led to the revival of the continental drift hypothesis in the following manner.

Magnetism in Rocks

Magnetite and certain other iron-bearing minerals can become permanently magnetized. This property develops because the electrons spinning around an atomic nucleus create a tiny atomic magnet. In minerals that can become permanent magnets, the atomic magnets line up in parallel arrays and reinforce each other. In nonmagnetic minerals, the atomic magnets are oriented in random directions.

Above a temperature called the **Curie point,** the thermal agitation of atoms is such that permanent magnetism of the kind found in magnetite is impossible. The Curie point for magnetite is 580°C (1076°F). Below the Curie point, adjacent atomic magnets reinforce each other (Fig. 6.2). When the Earth's magnetic field permeates the magnetite, all magnetic domains (regions in which the atomic magnets point in the same direction) parallel to the Earth's magnetic field become larger and expand at the expense of adjacent, nonparallel domains. Quickly, the parallel domains become predominant, and a permanent magnet is the result.

Consider what happens when lava cools. All the minerals crystallize at temperatures above 700°C (1292°F)—well above the Curie point of magnetite. As the crystallized lava continues to cool and the temperature drops below the Curie point, all the mag-

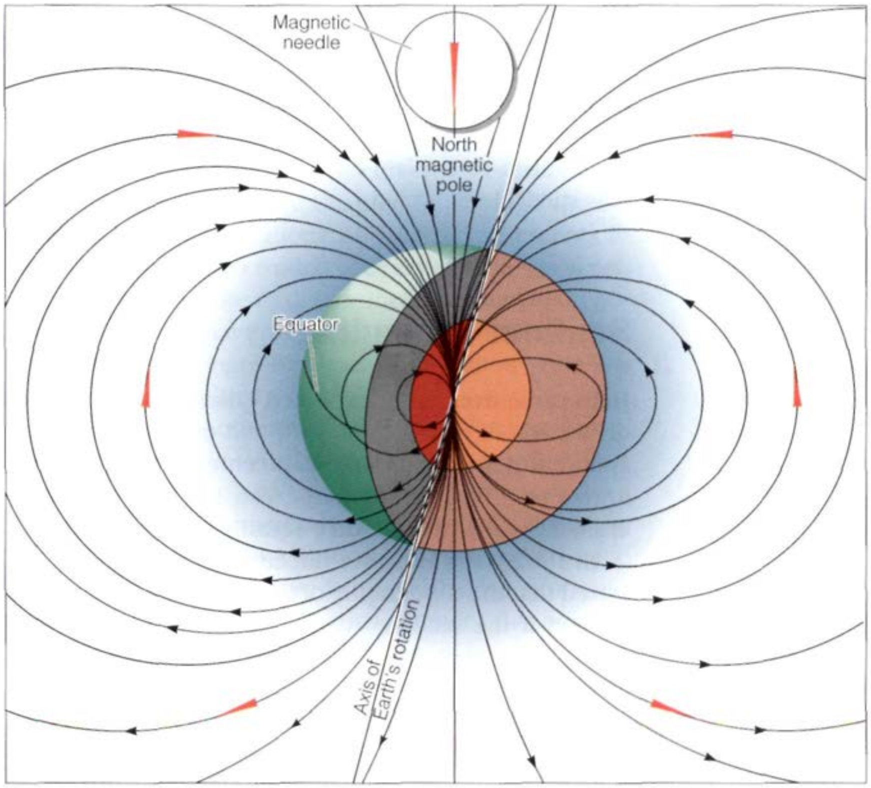


Figure 6.1 The Earth is surrounded by a magnetic field generated by the magnetism of the core. A magnetic needle, if allowed to swing freely, will always line up parallel to the magnetic field and point to the magnetic north pole.

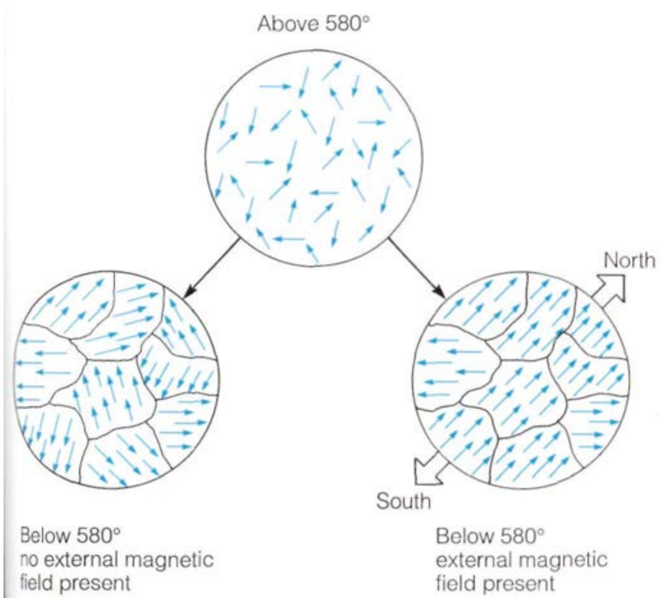


Figure 6.2 Magnetization of magnetite. Above 580°C (the Curie point), the vibration of atoms is so great that the magnetic poles of individual atoms, shown as arrows, point in random directions. Below 580°C, atoms in small domains reinforce one another and form tiny magnets. In the absence of an external field, the domains are randomly oriented. In the presence of a magnetic field, most domains tend to be parallel to the external field and the material becomes permanently magnetized.

netite grains in the rock become tiny permanent magnets having the same polarity as the Earth's field. Grains of magnetite locked in a lava cannot move and reorient themselves in the same way that a freely swinging magnet can. As long as that lava lasts (until it is destroyed by weathering or metamorphism), it will carry a record of the Earth's magnetic field at the moment it cooled below the Curie point.

Sedimentary rocks can also acquire weak but permanent magnetism through the orientation of magnetic grains during sedimentation. As sedimentary grains settle through ocean or lake water, or even as dust particles settle through the air, any magnetite particles act as freely swinging magnets and orient themselves parallel to the Earth's magnetic field. Once locked into a sediment, the grains make the rock a weak permanent magnet.

Apparent Polar Wandering

During the 1950s scientists started measuring past directions of the Earth's magnetic field by using ancient lavas. **Paleomagnetism** is the magnetism in rocks that records the directions of ancient magnetic fields at the time of rock formation, just as a free-swinging magnet indicates the direction of today's magnetic field. Two pieces of information can be obtained from

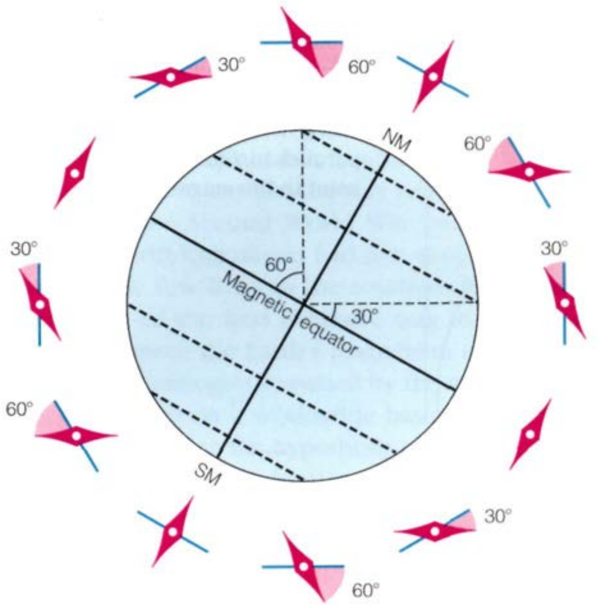


Figure 6.3 Change of magnetic inclination with latitude. The solid red diamonds show the magnetic inclinations of a free-swinging magnet. The solid blue line indicates a horizontal surface at each point.

paleomagnetism. The first is the direction of the magnetic field at the time the rock became magnetized. The second provides the data needed to say how far from the point of rock formation the magnetic poles lie; this is the *magnetic inclination*, which is the angle with the horizontal assumed by a freely swinging bar magnet. Note in Figure 6.3 that inclination varies regularly with latitude, from zero at the magnetic equator to 90° at the magnetic pole. The paleomagnetic inclination is therefore a record of the place between the pole and the equator (that is, the *magnetic latitude)* where the rock was formed. Once we know the magnetic latitude of a rock and the direction of the Earth's magnetic field at the time the rock was formed, we can determine the position of the magnetic poles at that time.

Geophysicists studying paleomagnetic pole positions during the 1950s found evidence suggesting that the poles wandered all over the globe. They referred to the strange plots of paleopole positions as *apparent polar wandering.* This evidence puzzled the geophysicists because the Earth's magnetic poles and the poles of the Earth's rotation axis should always be close together. Determination of the magnetic latitude of any rock should therefore be a good indication of the geographic latitude at which the rock was formed. When it was discovered that the path of apparent polar wandering measured in North America differed from that in Europe (Fig. 6.4), geophysicists were even more puzzled. Somewhat reluctantly, they concluded that, because it is unlikely that the magnetic poles moved, the continents—and with them the magnetized rocks—must have moved. In this way, the hypothesis of continental drift was revived, but a mechanism to explain how the movement occurred was still lacking.

Seafloor Spreading

Help came from an unexpected quarter. All the early debate about continental drift, and even the data on apparent polar wandering, had centered on evidence drawn from the continental crust. But if continental crust moves, why shouldn't oceanic crust move too?

In 1962 Harry Hess of Princeton University hypothesized that the topography of the seafloor could be explained if the seafloor were moving sideways, away from the oceanic ridge. His hypothesis came to be called **seafloor spreading,** and strong evidence in favor of it was soon found. (See the "Profile" at the end of this chapter).

From studies of paleomagnetism, geophysicists had discovered an extraordinary and still poorly understood phenomenon—some rocks contain a record of

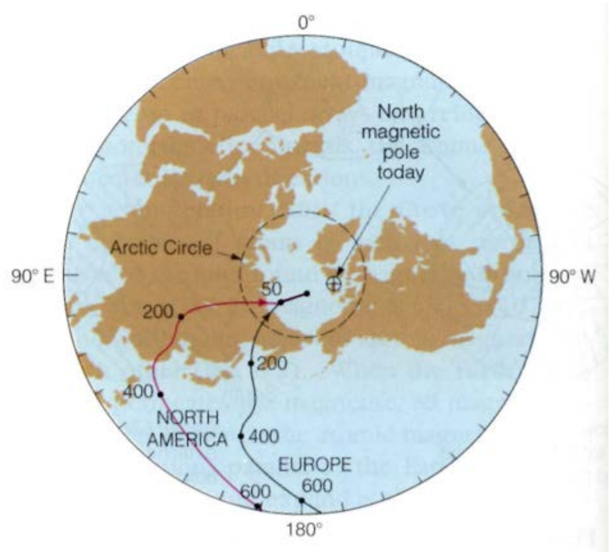


Figure 6.4 Apparent path of the north magnetic pole through the past 600 million years. Numbers are millions of years before the present. The curve determined from paleomagnetic measurements in North America (red) differs from that determined from measurements in Europe (black). Wide-ranging movement of the pole is unlikely; therefore, scientists conclude that it was the continents, not the pole, that moved.

reversed magnetic polarity. That is, some lavas indicate a south magnetic pole where the north magnetic pole is today, and vice versa (Fig. 6.5). Just why the Earth's poles reverse polarity is not yet understood, but the fact that they do provided some very interesting information. The ages of certain lavas can be accurately determined (see Chapter 7). Through age determinations and magnetic polarity measurements in thick piles of lava extruded over several million years, it has been possible to determine when magnetic polarity reversals have occurred (Fig. 6.6). A detailed record of all changes for nearly 200 million years has now been assembled.

The Hess hypothesis of seafloor spreading postulated that oceanic crust moved sideways, away from the oceanic ridge, and that basaltic magma rose from the mantle and formed new oceanic crust along the ridge. Although Hess could not explain what made the oceanic crust move, he nevertheless proposed that it did and that, as a consequence, the oceanic crust far from any ridge must be older than crust nearer the ridge. In the 1960s a powerful test of the Hess hypothesis was proposed by three geophysicists: Frederick Vine, (a student at the time); Drummond Matthews, Vine's mentor; and Lawrence Morley, a Canadian scientist. The Vine-Matthews-Morley suggestion concerned paleomagnetism, magnetic polarity reversals, and the oceanic crust.

When lava is extruded at the oceanic ridge, the rock it forms becomes magnetized and acquires the magnetic polarity that exists at the time. Thus, oceanic crust should contain a record of when the Earth's magnetic polarity was reversed. The oceanic crust should be, in effect, a very slowly moving magnetic tape recorder. In fact, two oceanic tape recorders commence at the midocean ridge, one on each side of the ridge. In these "recorders," successive strips of oceanic crust are magnetized with normal and reversed polarity (Fig. 6.7). It was a straightforward matter to match this magnetic pattern with a record of magnetic polarity reversals, such as that shown in Figure 6.6. The magnetic striping allowed the age of any place on the seafloor to be determined.

Because the ages of magnetic polarity reversals had been so carefully determined, magnetic striping provided a means of estimating the speed with which the seafloor had moved. In some places today, such movement is remarkably fast: as high as 9 cm/yr (3.5 in/yr).

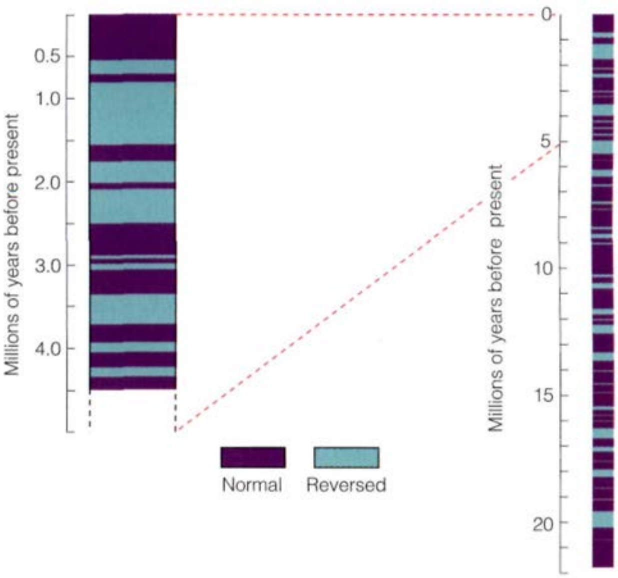


Figure 6.6 Polarity reversals during the past 20 million years.

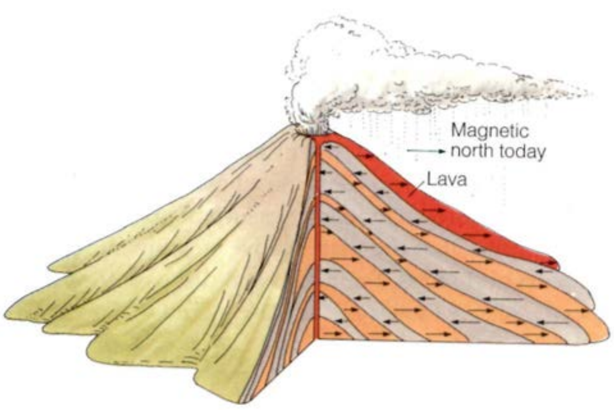


Figure 6.5 Lavas retain a record of the polarity of the Earth's magnetic field at the instant they cool through the Curie point. A pile of lava flows, like those in the volcanoes of the Hawaiian islands, may record several field reversals.

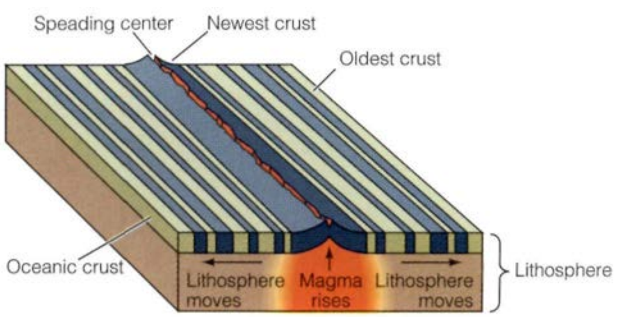


Figure 6.7 Schematic diagram of oceanic crust. Lava extruded along a spreading center at the midocean ridge forms new oceanic crust. As the lava cools, it becomes magnetized with the polarity of the Earth's magnetic field. Successive strips of oceanic crust have opposite polarities.

PLATE TECTONICS

Proof that the seafloor moves was the spur needed for the emergence of the theory of plate tectonics, which, as mentioned on page 15 in the Introduction, states that segments (plates) of the Earth's hard, outermost shell (the lithosphere) move slowly sideways.[1] The two essential points in formulating the theory were, first, that the zone of low seismic-wave velocities between 100 and 350 km (62 to 218 mi) deep (as discussed in Chapter 3) is exceedingly weak, viscous, and fluidlike. It was quickly realized that the asthenosphere, postulated many years earlier in order to explain isostasy (the flotational property of the lithosphere), but never proved to exist, and the zone of low seismic-wave velocities in the upper mantle must be one and the same. The second point was that the rigid lithosphere is strong enough to form coherent slabs (plates) that can slide sideways over the weak, underlying asthenosphere. These two points answered the main objection to Wegener's hypothesis—movement must occur with minimal resistance from friction. The crust—both oceanic and continental—is part of the lithosphere. Thus, one consequence of plate tectonics is that, as the plates of lithosphere move, the crust on the plate is rafted along as a passenger. Continents move, to be sure, but they do so only as portions of larger plates.

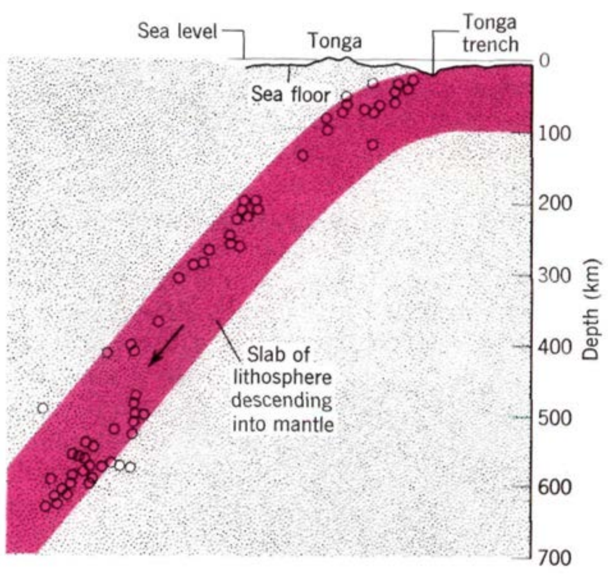


Figure 6.8 Deep focus earthquakes define the Benioff zone near the island of Tonga, in the Pacific Ocean. Each circle represents a single earthquake in 1965. The earthquakes are generated by downward movement of a comparatively cold slab of lithosphere.

[1]The word tectonics is derived from the Greek *tekton,* carpenter or builder. Thus the plate tectonics theory seeks to explain how the surface of our planet has been built up.

The theory of plate tectonics provides a solution for one of the puzzles about seafloor spreading. If, as seafloor spreading theory requires, new oceanic crust is being created along the midocean ridge, either the Earth's surface must be expanding and the ocean basins getting larger, or else an equal amount of old crust must be being destroyed. The answer to the puzzle was provided by previously unexplained regions inside the Earth, called Benioff zones, where very deep focus earthquakes occur. These slanting zones are the places where lithosphere capped by old, cold oceanic crust is sinking into the asthenosphere (Fig. 6.8). Destruction of old oceanic crust and creation of new oceanic crust are in balance.

Earthquake foci quickly provided evidence to support the hypothesis that the lithosphere is broken into six large and many small plates, each about 100 km (60 mi) thick. As Figure *6.9* demonstrates, most of the Earth's seismicity occurs in sharply defined belts, and it is these earthquake belts that outline the plates.

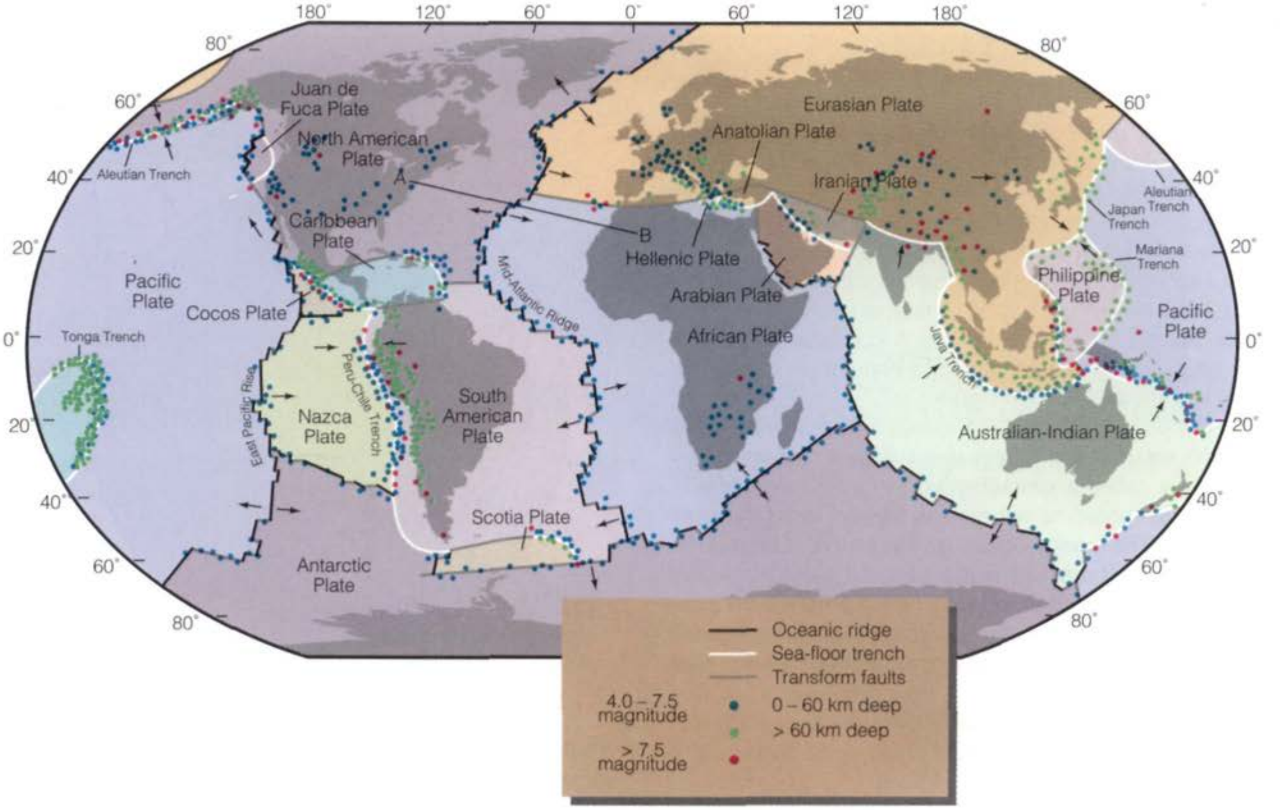
Plate Motions

As a plate moves, everything on it moves too. If the top of the plate is partly oceanic crust and partly continental crust, then both the ocean floor and the continent move with the same speed and in the same direction. Although the first clear evidence that seafloor and continent on the same plate of lithosphere move together came from paleomagnetism, a series of remarkable measurements have recently provided even more convincing evidence.

The new evidence of plate motion comes from satellites. Using laser beams bounced off satellites, we can measure the distance between two points on the Earth with an accuracy of about 1 cm (0.4 in.). Thus, we can monitor any change in distance between, for example, Los Angeles on the Pacific plate and San Francisco on the North American plate. By making distance measurements several times a year, therefore, we can measure present-day plate velocities directly. As seen in Figure 6.10, plate speed based on satellite measurements agrees very closely with speeds calculated from paleomagnetic measurements. The agreement implies that the plates move steadily rather than by starts and stops. For additional discussion of plate speeds, see "A Closer Look: Plate Speeds."

(Facing page, top.)

Figure 6.9 The Earth's seismicity outlines plate margins. This map shows earthquakes of magnitude 4.0 or greater from 1960 to 1989. Six large plates of lithosphere and several smaller ones are present. Each plate moves slowly but steadily in the direction shown by the arrows. The profile shown in Figure 6.13 lies along the line A-B.



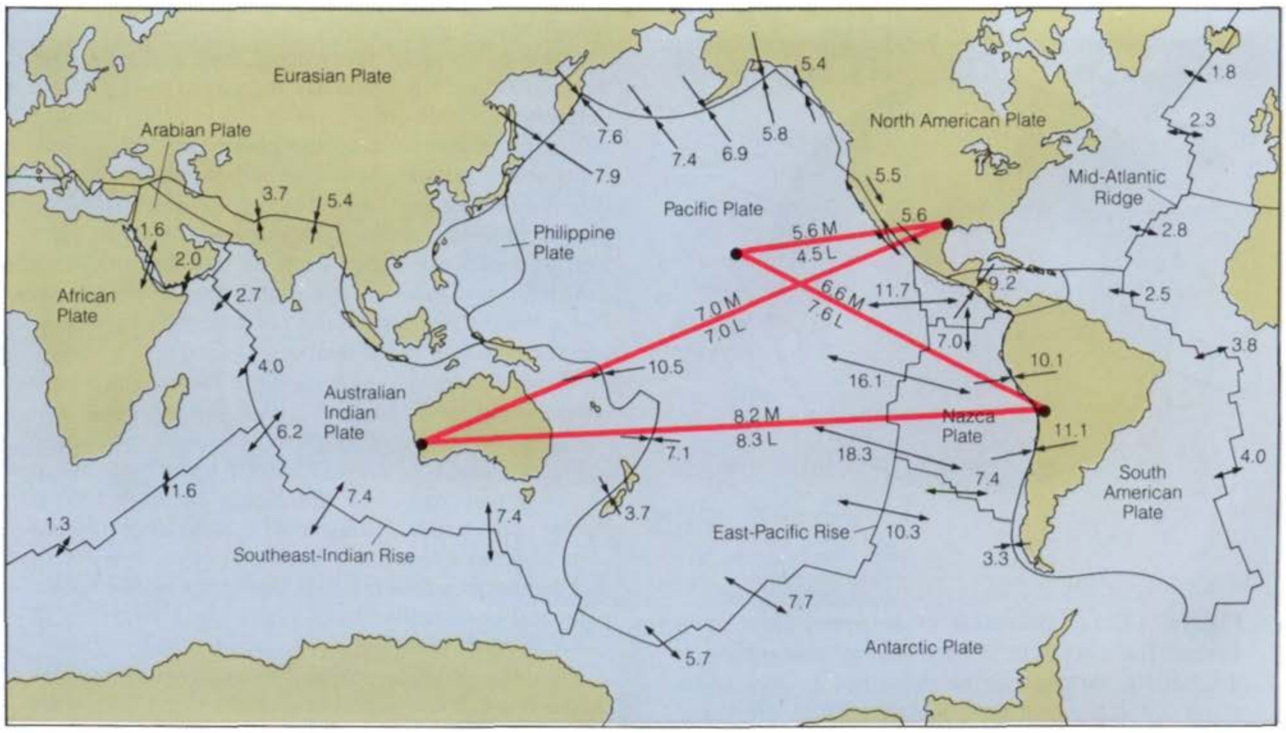


Figure 6.10 Present-day plate speeds in centimeters per year, determined in two ways. Numbers along the midocean ridges are average speeds indicated from paleomagnetic measurements. A speed of 16.1, as shown for the East Pacific Rise, means that the distance between a point on the Nazca plate and a point on the Pacific plate increases, on the average, by 16.1 cm each year in the direction of the arrows. The long red lines connect stations used to determine plate motions by means of satellite laser ranging (L) techniques. The measured speeds between stations are very close to the average speeds estimated from magnetic measurements (M).

A Closer Look

Plate Speeds

You might think, intuitively, that all points on a plate move with the same speed, but that is incorrect. This intuition would be correct only if the plates were flat and moved over a flat asthenosphere (like plywood floating on water). However, tectonic plates are pieces of a shell on a spherical Earth; in other words, they are curved, not flat. Any movement on the surface of a sphere is a rotation about an axis of the sphere. A consequence of such rotation is that different parts of a plate move with different speeds, as shown in Figure C6.1.

The plate in Figure C6.1 moves independently of the Earth's rotation and rotates about an axis of its own, colloquially called a *spreading axis.* In the figure, point P, where the spreading axis reaches the surface, is a *spreading pole.* The motion of each of the Earth's plates can be described in terms of rotation around the plate's own spreading axis, and the speed of each point on the plate depends on how far that point is from the spreading pole—the speed is greatest at the farthest distance from the pole.

One consequence of rotation around a spreading axis is that the width of new oceanic crust bordering a spreading center increases with distance from the spreading pole (Fig. C6.2).

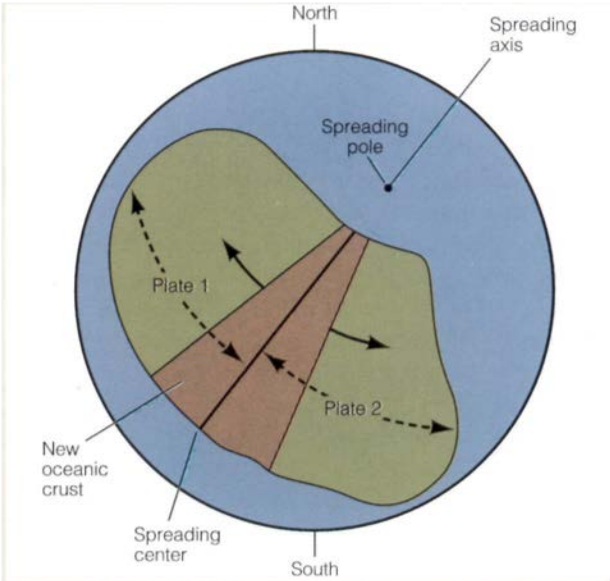


Figure C6.2 The width of new oceanic crust increases away from the spreading pole.

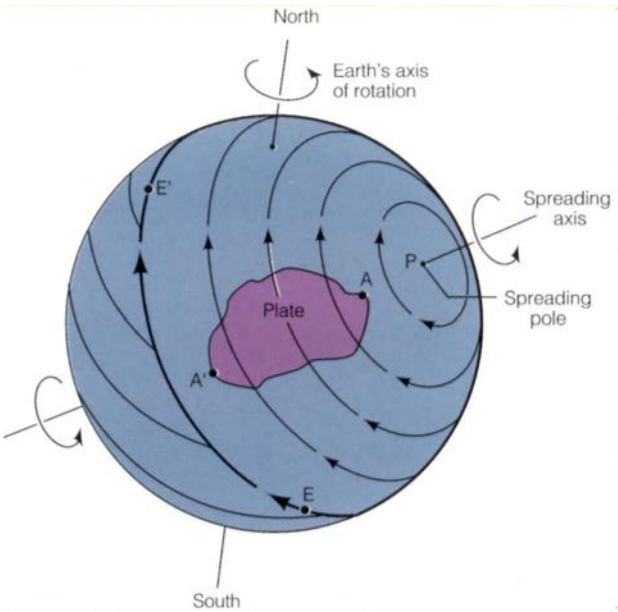


Figure C6.1 Movement of a curved plate on a sphere. The movement of each plate of lithosphere on the Earth's surface can be described as a rotation about the plate's own spreading axis. Point *P* has zero speed because it is the fixed point around which rotation occurs. Point *A'*. at the edge of the plate closest to the equator *EE'*, has a high speed. Point *A*, closest to the pole, has a low speed.

Some plates rotate faster about their spreading axes than others, and the reason has to do with the load on the lithosphere. Continental crust is thicker than oceanic crust, and, consequently, lithosphere capped by continents seems to protrude deeper into the asthenosphere than does lithosphere capped by oceanic crust. The extra protrusion seems to slow things down. Plates that carry lots of continental crust, such as the African, North American, and Eurasian plates, move relatively slowly. Plates without any continental crust, such as the Pacific and Nazca plates, move relatively fast.

The fastest moving plates are the Pacific and Nazca plates. The point on the Pacific plate that is farthest from the spreading axis (the fastest moving point on a plate) moves at a speed of 9 cm/y (3.5 in/yr), about twice as fast as a fingernail grows. The Nazca plate is adjacent to the Pacific plate, and the fastest moving point on the Nazca plate also has a speed of 9 cm/y (3.5 in/yr). This means that the distance between two points, one on the Pacific plate and one on the Nazca plate, grows larger by 18 cm/yr (7 in/yr).

From the symmetrical spacing of magnetic time lines on the two sides of a spreading center (Fig. C6.3), it appears that a spreading center is fixed and that both plates move away from it at equal rates, but this is a case where appearances deceive. The same pattern of magnetic time lines would be observed if the African plate were sta-

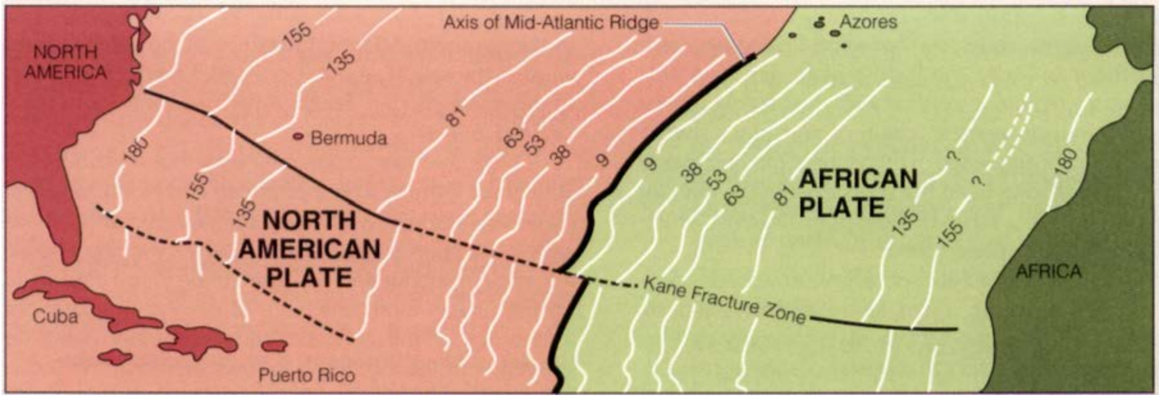


Figure C6.3 Age of the ocean floor in the central North Atlantic, deduced from magnetic striping. Numbers give ages in millions of years before the present.

tionary and both the Mid-Atlantic Ridge and the North American plate were moving westward. In fact, all that can be deduced from magnetic time lines is the *relative speed* of two plates. In order to measure the *absolute speed* of a plate, an external frame of reference is needed.

A familiar example of absolute versus relative motion occurs when one automobile overtakes another. If observers in the two automobiles could see only each other and not any fixed objects outside their cars, they could judge only the *difference* in speed between the two cars. One car could be traveling at 50 km/h (31 mi/h) and the overtaking car at 55 km/h (34 mi/h), but all the observers could determine would be that the *relative speed* difference was 5 km/h (3 mi/h). On the other hand, if the observers measured speed with respect to a stationary reference, such as the ground, they would determine that the *absolute speeds were* 50 and 55 km/h (31 and 34 mi/h), respectively.

We would be constrained to determine only relative plate speeds if a fixed framework did not exist. Fortunately, such a framework does exist. During the nineteenth century, the American geologist James Dwight Dana observed that the age of volcanoes in the Hawaiian chain, some now submerged beneath the sea, increases from southeast to northwest (Fig. C6.4). Apparently, a

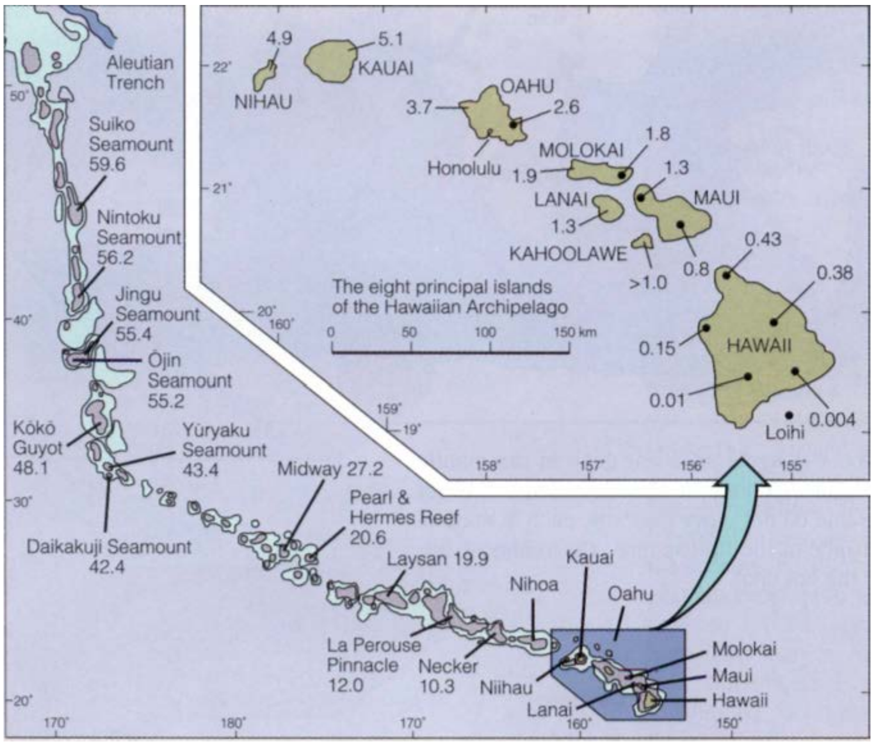


Figure C6.4 Hawaiian chain of volcanoes, showing the oldest reliable ages (in millions of years). The age increases from southeast to northwest.

long-lived magma source (a *hot spot)* lies somewhere deep in the mantle. As the Pacific plate moves, a volcano can only remain in contact with the magma source for about a million years. A chain of volcanoes should therefore be a consequence of lithosphere moving over a fixed hot spot. If long-lived hot spots exist in the mantle, they could provide a series of fixed points against which to measure absolute plate speeds.

More than a hundred hot spots have now been identified (Fig. C6.5). Using them for reference, scientists have determined that the African plate is nearly stationary (evidenced by the fact that volcanoes there seem to be very long lived). Because the African plate is almost completely surrounded by spreading centers, and because the relative speeds of the plates on either side of Africa are the same, we must conclude that the Mid-Atlantic ridge is moving westward and that the mido-cean ridge that runs up the center of the Indian Ocean is moving eastward. Because the absolute motion of the African plate is zero or nearly so, the Mid-Atlantic ridge in the southern Atlantic Ocean must be moving westward at the rate of about 2 cm/y (0.8 in/yr), and the absolute speed of the South American plate must be 4 cm/yr (1.6 in/yr).

A plate that has a spreading edge but no subduction zone must grow in size; the African and North American plates are examples. To keep things in balance, plates with subduction zones must be slowly shrinking. Most of the modern subduction zones are to be found around the Pacific Ocean along the edge of the Pacific plate, and thus much of the oceanic lithosphere now being destroyed is in the Pacific. It follows then that the Indian Ocean, the Atlantic Ocean, and most other oceans must be growing larger, while the Pacific Ocean must be steadily getting smaller. It is estimated that about 200 million years in the future the Pacific Ocean will have disappeared and Asia and North America will have collided as a result. The assembly of a new Pangaea will be underway.

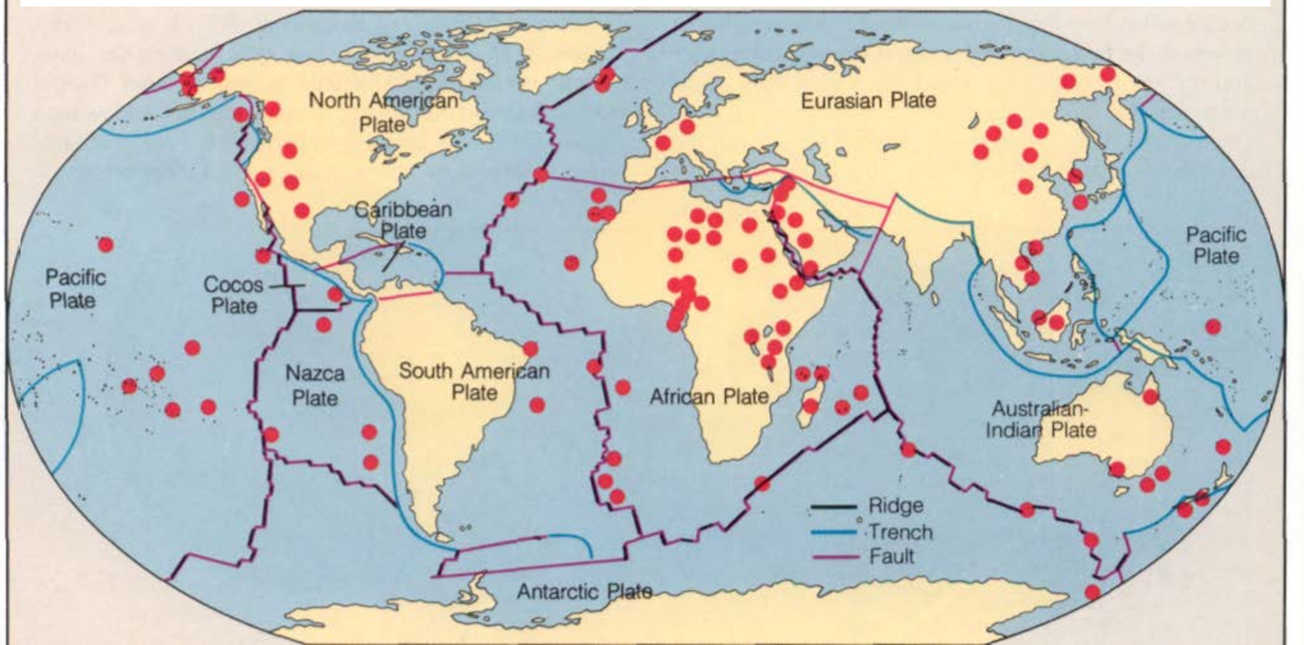


Figure C6.5 Long-lived hot spots (magma sources) deep in the mantle can be used to determine the absolute motions of plates. Because the hot spots lie far below the lithosphere and do not move laterally, each is marked by a chain of volcanoes on the surface of the lithosphere. The youngest volcano in a chain lies directly above the hot spot.

Plate Margins

Plates move as individual units, and interactions between plates occur along their edges (Fig. 6.11). Plate interactions are distinctively expressed by submarine volcanism and by stratovolcanoes, but it has been through studies of earthquakes that scientists have deciphered most of what we know about plate margins:

1. **Divergent margins,** also called **spreading centers** because they are fractures in the lithosphere where two plates move apart, are characterized by earthquakes of low magnitude and shallow foci—no more than 10 km (6 mi) below the surface.
2. **Convergent margins** occur where two plates are moving toward each other. Along convergent margins, either one plate must sink beneath the other, in which case we refer to the margin as a **subduction zone**, or else continental crust on the two plates collides as a result of convergence in which case we refer to the margin as a **collision margin.** Convergent margins have earthquakes with foci that are sometimes very deep—down to 700 km (435 mi)—and frequently of high magnitude.
3. **Transform fault margins** are fractures in the lithosphere where two plates slide past each other, grinding and abrading their edges as they do so. Earthquakes with foci between 10 and 100 km (6 and 62 mi) deep are frequent along most transform fault margins, and many are of high magnitude.

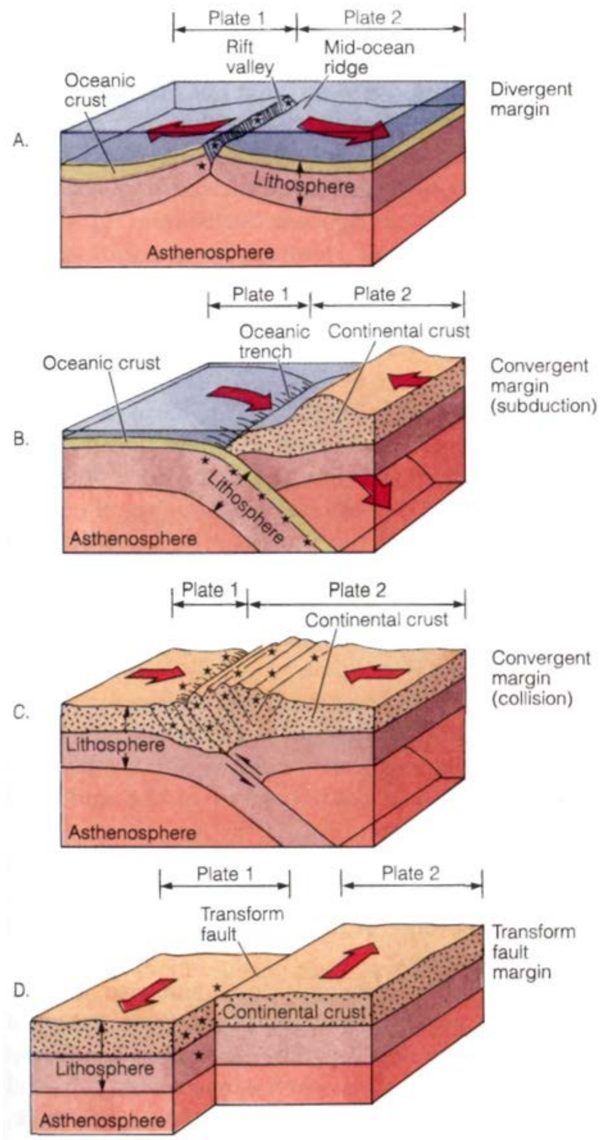


Figure 6.11 The various kinds of plate margins. Stars indicate earthquake foci. A. Divergent margin (also called spreading center). The topographic expression is a mid-ocean ridge formed as a result of volcanism. Earthquakes have shallow foci and low magnitude. B. Convergent margin—subduction. The topographic expression is a deep trench. Earthquake foci down to 700 km, often of high magnitude. C. Convergent margin—collision. The topographic expression is a mountain range. Earthquake foci down to 300 km, sometimes of high magnitude. D. Transform fault margin. No characteristic topographic expression, but margin is often marked by a long, thin valley. Earthquake foci down to 100 km and often of high magnitude.

PLATE TECTONICS AND THE EXTERNAL STRUCTURE OF THE EARTH

The beauty of plate tectonics is that it provides explanations for all the major features we see at the Earth's surface. These features are most easily visualized by considering the different kinds of plate margins: divergent, convergent, and transform fault.

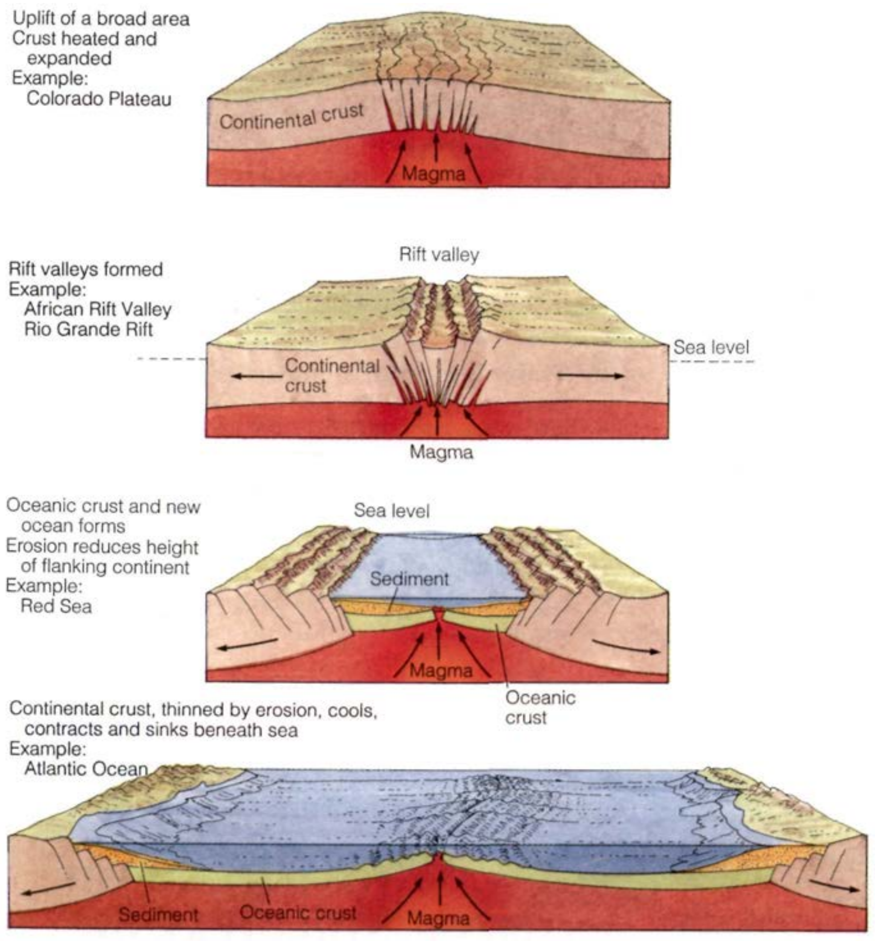


Figure 6.12 The rifting of continental crust to form a new ocean basin. The rifting can cease at any stage. It is not necessarily correct to conclude that the Rio Grande or the African Rift Valley, for example, will someday open to form new oceans.

Divergent Margins

Curious as it may seem, divergent plate margins start life on a continent and become an ocean. The sequence of events is illustrated in Figure 6.12.

The reasons a continent splits and a new ocean forms have to do with heat escaping from the Earth. Huge continental masses are thermal blankets that slow down the escape of heat from the interior. A plate capped by a large continent, such as the African plate, slowly heats up from below, expands, and eventually splits to start a cycle of spreading.

The structure of ocean basins is now understandable in terms of divergent margins. Modern shorelines don't coincide exactly with the boundaries between continental crust and oceanic crust. This is so because some ocean water spills out of the ocean basin onto the continent (Fig. 6.13). The boundaries between continental and oceanic crust are therefore covered by water, and today's shorelines are actually on the continents. As a result, each continent is surrounded by a flooded margin of continental crust that is of variable width and is known as the **continental shelf.** The geological edge of the ocean basin is not the shoreline; rather, it is the place where oceanic crust joins the continental crust. This is the edge of the rift that broke the old continent. The geological edge of an ocean basin is at the bottom of the **continental slope,** a pronounced slope beyond the seaward margin of the continental shelf.

The **continental rise** lies at the base of the continental slope. It is a region of gently changing slope where the oceanic crust meets the continental crust.

Some continental margins coincide with the edges of tectonic plates; the west coast of South America is an example. Other continents sit in the middle of plates, and their margins are far from plate edges; North America and Africa are examples. Regardless of today's configurations, the margins of all continents have, at some time in the geological past, coincided with plate margins.

Beyond the continental slope and rise lies the rarely seen world of the deep ocean floor. Teams of oceanographers and seagoing geologists used deep diving submarines and other new devices to sound and sample the ocean bottom. As a result of this work,

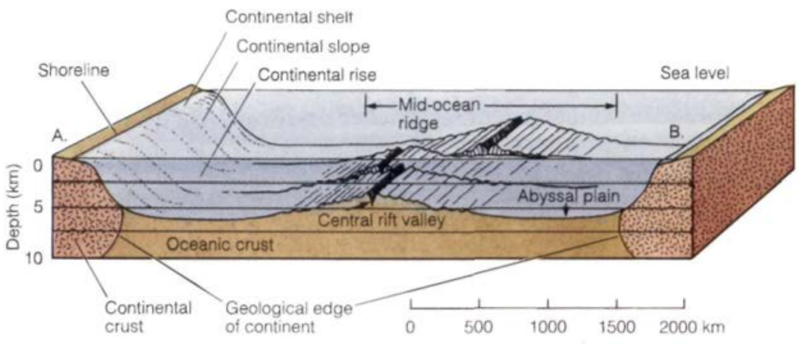


Figure 6.13 Portion of the Atlantic Ocean, showing the major topographic features.

we now know almost as much about the seafloor as we do about the land surface.

Large, flat areas known as **abyssal plains** are a major topographic feature of the seafloor and lie adjacent to the continental rise (Fig. 6.13). These plains generally are found at depths of 3 to 6 km (2 to 4 mi) below sea level and range in width from about 200 to 2000 km (124 to 1243 mi). They are most common in the Atlantic and Indian oceans, which have large, mud-laden rivers entering them. Abyssal plains form when the mud settles through the ocean water and buries the original seafloor topography beneath a blanket of fine debris.

Convergent Margins

Convergent margins are the places where lithosphere capped by oceanic crust sinks into the asthenosphere; the process is called subduction. The geographic features of a subduction zone are shown in Figure 6.14. Particularly striking is the deep-sea trench that marks the place where oceanic-capped lithosphere bends and sinks into the asthenosphere, and an arc-shaped chain of volcanic islands (called an *island arc* or a *magmatic arc)* formed above the sinking lithosphere. Trenches are the deepest parts of the ocean. Besides the prominent trench and the island arc, three less

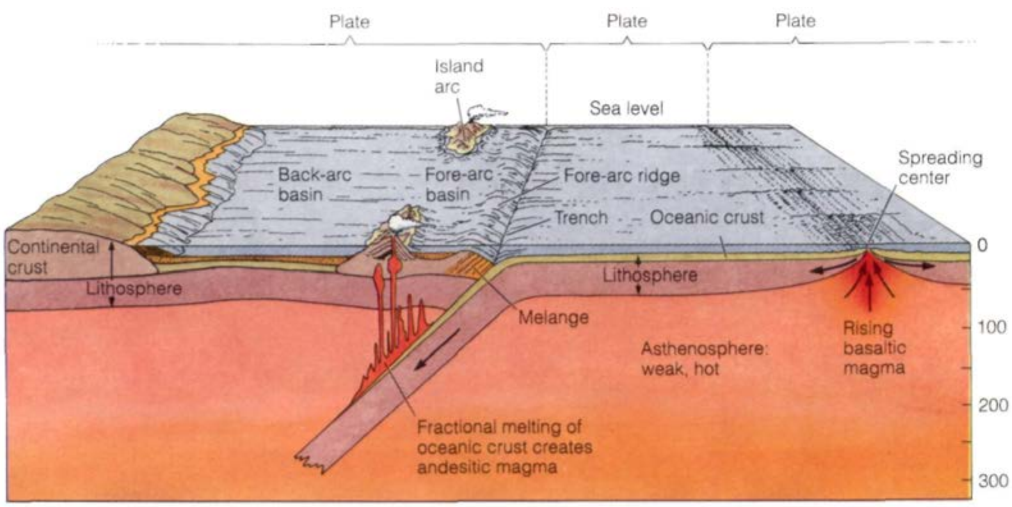


Figure 6.14 Structure of tectonic plates at a convergent margin. Along the line of subduction, an oceanic trench is formed, and sediment deposited in the trench, plus sediment from the sinking plate, is compressed and deformed to create a melange of shattered and crushed rock shaped as a fore-arc ridge. The sinking oceanic crust eventually reaches the temperature where melting commences and forms andesitic magma, which then rises to form an island arc of stratovolcanoes on the overriding plate. On the side of the island arc away from the trench, tensional forces lead to the development of a back-arc basin.

prominent features are present above subduction zones: the fore-arc ridge, fore-arc basin, and back-arc basin. The fore-arc ridge and fore-arc basin lie between the trench and the island arc. A *fore-arc ridge* is commonly underlain by a zone of smashed and shattered rock called a melange that causes a local thickening of the crust. *A fore-arc basin* is a low-lying region between the fore-arc ridge and the island arc. Behind the island arc is another shallow basin, the *back-arc basin.* The islands of Sumatra and Java in Indonesia are an example of a present day magmatic arc that is flanked by a fore-arc ridge and a back-arc basin (Fig. 6.15).

In order to understand how and why a subduction zone develops and thus how a convergent margin begins, it is necessary to consider what happens to a plate of lithosphere as it moves away from a spreading center.

Near a spreading center, the lithosphere is thin and its boundary with the asthenosphere is close to the surface (Fig. 6.16). (We have seen this figure before, as Figure 4.11, when we defined a spreading center. This time, as you look at it, notice that the lithosphere near a spreading center is thin and that it thickens as it moves away.) This thinning happens because magma rising toward the spreading center heats the lithosphere so that only a thin layer near the top retains the strength properties of the lithosphere.

As the lithosphere moves away from the spreading center, it cools and becomes denser. In addition, the boundary between the lithosphere and the asthenosphere becomes deeper, and as a result, the lithosphere becomes thicker and the asthenosphere thinner. Finally, about 1000 km (600 mi) from the spreading center, the lithosphere reaches a constant thickness and is so cool that it is denser than the hot, weak asthenosphere below. Eventually, the cool lithosphere breaks and starts to sink downward. Like a conveyor belt, old lithosphere with its capping of oceanic crust sinks into the asthenosphere and eventually into the mesosphere.

As the moving strip of lithosphere sinks slowly through the asthenosphere, it passes beyond the region where geologists can study it directly. Consequently, what happens next is conjecture. The thin layer of oceanic crust on top of the sinking lithosphere melts at a depth of about 100 km (62 mi) and becomes magma; some of this magma reaches the surface to form volcanoes of the magmatic arc. Figure 6.17 is an example of a present-day magmatic arc of andesitic stratovolcanoes parallel to, but about 150 km (93 mi) from, the trench that marks the place where the Nazca plate sinks below the South American plate. The Sierra Nevada of California is an example of an old magmatic arc, evidence that the western margin of North America was once a subduction zone like the western margin of South America today.

When the sinking rate of a subducting plate is

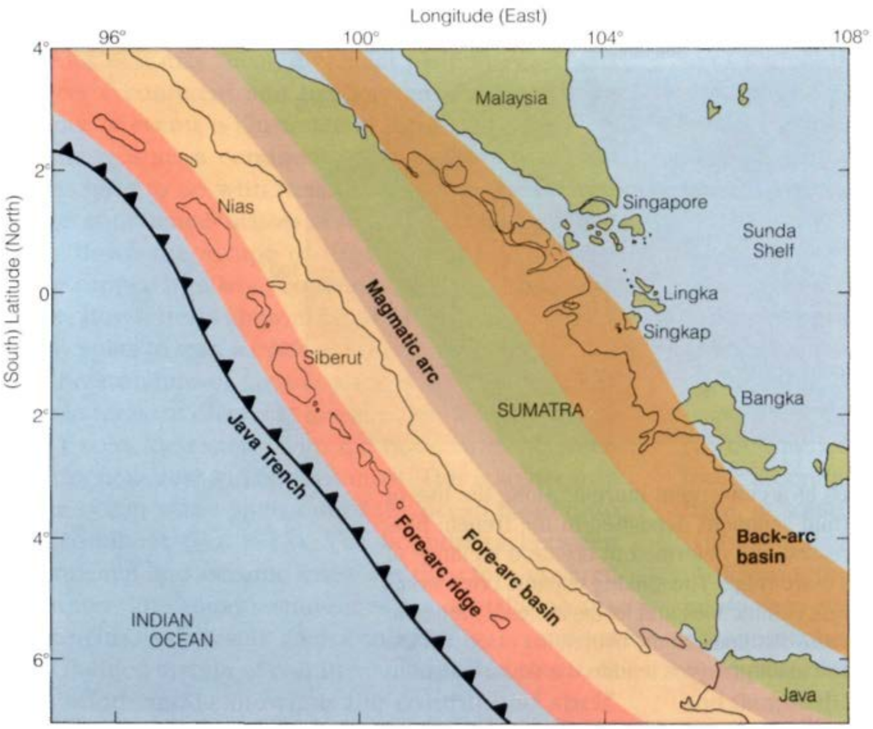


Figure 6.15 Map of portion of Indonesia showing the positions of the major topographic features in a present-day convergent margin.

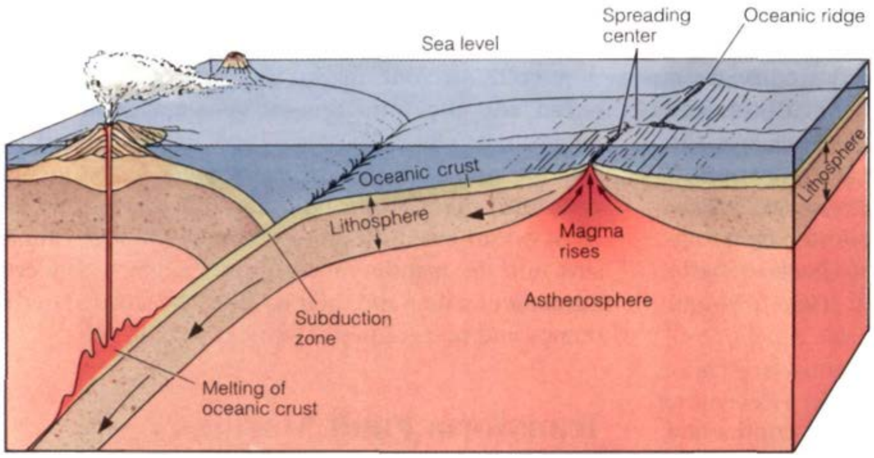


Figure 6.16 Schematic diagram showing the major features of a plate. Near the spreading center, where the temperature is high because of rising magma, the lithosphere is thin. Away from the spreading center, the lithosphere cools, becomes denser and also thicker, and so the lithosphere-asthenosphere boundary is deeper. When the lithosphere sinks into the asthenosphere at the subduction zone, it is reheated. At a depth of about 100 km, the oceanic crust starts to melt, and the magma rises and forms an arcuate belt of andesitic stratovolcanoes parallel to the subduction zone.

Figure 6.17 An example of a present-day convergent margin. This chain of andesitic stratovolcanoes in Ecuador sits above the subduction zone where the Nazca plate sinks below the western edge of the South American plate. Several snow-capped volcanoes are visible in this aerial photograph.

faster than the forward motion of the overriding plate, part of the overriding plate can be subjected to tensional (pulling) stress. If the leading edge of the overriding plate did not remain in contact with the subduction edge, a huge void would open. Therefore, the overriding plate grows slowly larger at a rate equal to the difference in velocities between the two plates. Most commonly, this process is manifested by a thinning of the crust and the formation of a back-arc basin behind and parallel to the island arc (Figs. 6.14 and 6.15).

Collision Zones

Because it is less dense than the mantle, continental crust is too buoyant to be dragged down into a subduction zone. Therefore, when the two members of a converging pair of plates are capped by continental crust, the eventual result is a collision, as Figure 6.18 shows.

A collision zone marks the disappearance of an ocean and the formation of spectacular mountain ranges in its place. The Alps, the Urals, the Himalaya, and the Appalachians are the results of continental collisions; therefore, each is the graveyard of an ancient ocean basin. Because continental crust cannot sink into the mantle, much of the evidence concerning ancient plates and their motions is recorded in the bumps and scars of past continental collisions.

Transform Fault Margins

Transform faults are great vertical fractures that cut right down through the entire lithosphere. One trans-

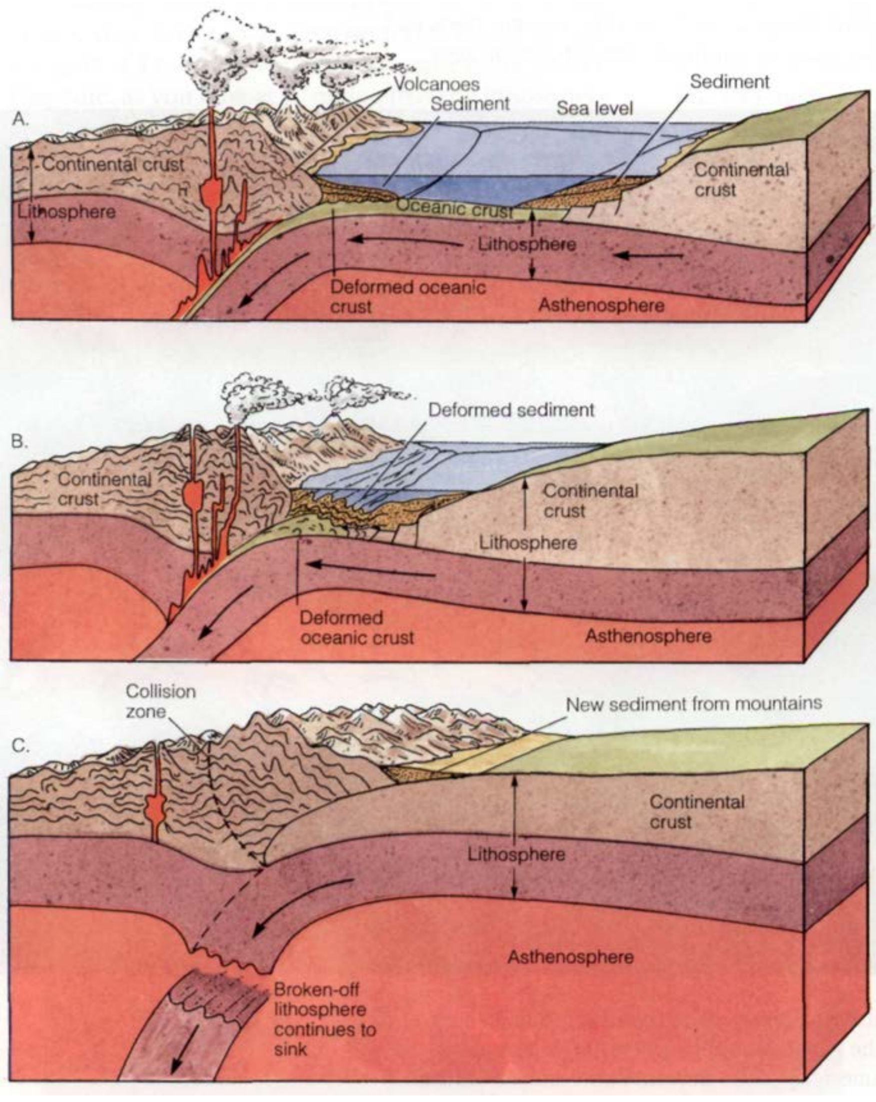


Figure 6.18 Mountains form when two masses of continental crust collide. A. Subducting oceanic lithosphere compresses and deforms sediments at the edge of continent on overriding plate (left). Sediments at the edge of continent on subducting plate (right) are undeformed. B. Collision. Sediment at the edge of continent on subducting plate is deformed and welded onto already deformed continental crust on overriding plate. C. After collision. The leading edge of the subducting plate breaks off and continues to sink. The two continental masses are welded together, and a mountain range stands where once there was ocean.

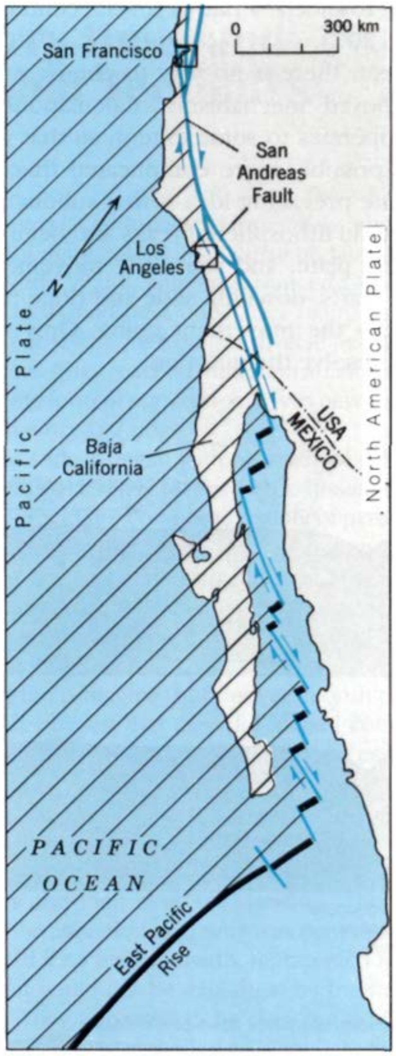


Figure 6.19 The San Andreas Fault is a transform fault margin that separates the Pacific plate from the North American plate. Directions of motion are shown by the arrows. Los Angeles, on the Pacific plate, is moving north, while San Francisco is moving south, bringing the two cities closer together at a speed of 5.5 cm/v; they will be adjacent about 10 million years in the future.

form fault margin that is much in the public eye today because of the threat of earthquakes along it is the San Andreas Fault in California (Fig. 6.19). This fault, which runs approximately north-south, separates the North American plate on the east, on which San Francisco sits, from the Pacific plate on the west, on which Los Angeles sits. The Pacific plate is moving in a northerly direction, and the North American plate is moving in a southerly direction. As the two plates grind and scrape past each other, Los Angeles is slowly moving north and San Francisco is moving south. At times the plate edges grab and lock, and as they do the rocks on both sides flex and bend. When the locked section breaks free, the flexed rock snaps suddenly and an earthquake occurs.

Eventually, many millions of years in the future, Los Angeles and San Francisco will be adjacent. Then, as the two plates continue to move, the fragment of continental crust on which Los Angeles sits will become a long thin island. The trip will end when the future "Los Angeles Island" reaches the subduction zone along the northern edge of the Pacific plate and the island collides with Alaska and the Aleutian islands.

CAUSE OF PLATE TECTONICS

Just as Wegener felt sure that continents had drifted but he could not explain how, so today we are sure that plates move and that convection currents play a role, but we are still unable to say exactly what role. The situation is analogous to knowing the shape, color, size, and speed capability of an automobile and knowing that gasoline supplies the energy needed for movement but not knowing how the gasoline makes the engine work. Until the driving mechanism is explained, plate tectonics must remain only an approximate description. Meanwhile, we can hypothesize about the causes of the motion and test the hypotheses by making detailed calculations based on the laws of nature.

The lithosphere and asthenosphere are closely bound together. If the asthenosphere moves, the lithosphere must move too, just as the movement of sticky molasses moves a piece of wood floating on the surface of the molasses. Conversely, movement of the lithosphere causes movement in the asthenosphere. Such is our state of uncertainty, however, that we cannot yet separate the relative importance of the two effects. Nevertheless, on two points we are quite certain: (1) the lithosphere has energy of motion and (2) the source of this energy is the Earth's internal heat. We know, too, as discussed in the Introduction, that the heat energy reaches the Earth's surface by convection in the mantle. What has not yet been figured out is the precise way the heat energy brought up by convection causes plates to move. However, all scientists who have studied the problem agree that convection keeps the asthenosphere hot and weak by bringing up heat from deep in the mantle and the core. In this sense at least, convection is essential for plate tectonics.

Movement of the Lithosphere

Three forces might play a role in moving the lithosphere. The first is a push away from a spreading center. Rising magma at a spreading center creates new lithosphere and in the process pushes the plates away from each other (Fig. 6.20A). Once the process is started, it tends to keep itself going. The problem with this hypothesis is that pushing involves compression, but the existence of rifts along a midocean ridge indicates a state of *tension* (the opposite of compression).

A second way lithosphere could be made to move is by dragging. Proponents of the dragging idea point out that lithosphere breaks and starts to sink through the asthenosphere because the cold lithosphere is denser than the hot asthenosphere. Because rock is a poor conductor of heat, they argue, the temperature at the center of a tongue of lithosphere can be as much as 1000°C (1832°F) cooler than the mantle at depths of 400 to 500 km (250 to 310 mi). This means that a sinking tongue of lithosphere will continue to be denser than the rock it is sinking through and it must exert a pull on the entire plate. This is like a heavy weight that hangs over the side of a bed and is tied to the edge of a sheet. The weight falls and pulls the sheet across the bed. To compensate for the descending lithosphere, rock in the asthenosphere must flow slowly back toward the spreading center (Fig. 6.20B).

Both the pushing and the dragging mechanism have problems, however. Plates of lithosphere are brittle, and they are much too weak to transmit large-scale pushing and pulling forces without major deformation occurring in their middle. Deformation is not present, however, and midplate seismicity, which would be expected for a plate undergoing deformation, is infrequent.

The third possible mechanism for lithosphere movement is for the whole plate to slide downhill away from the spreading center. The lithosphere grows cooler and thicker away from a spreading center. As a consequence, the boundary between lithosphere and asthenosphere slopes away from the spreading center. If the slope is as little as 1 part in 3000, the lithosphere's own weight could cause the lithosphere to slide at a rate of several centimeters per year (Fig. 6.20C).

At present, there is no way to choose among the three proposed mechanisms. Calculations suggest that each operates to some extent, so that the entire process is possibly more complicated than we now imagine. The prevailing idea is that subduction starts when old, cold lithosphere breaks and begins to sink, pulls on the plate, and starts the movement. Once movement starts, downhill slide and ridge push combine to keep the movement going. Only future research will resolve the question.

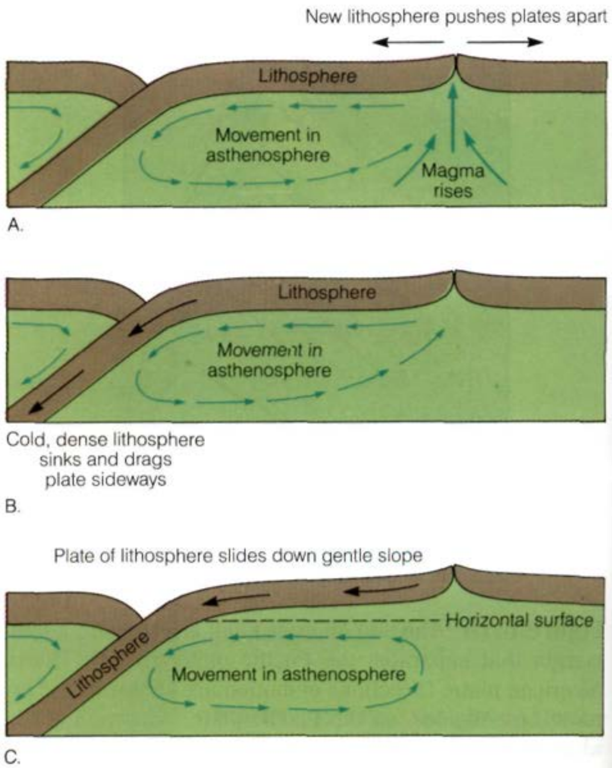


Figure 6.20 Three suggested mechanisms by which lithosphere might move over the asthenosphere. A. Magma rising at a spreading center exerts enough pressure to push the plates of lithosphere apart. B. A tongue of cold, dense lithosphere sinks into the mantle and drags the rest of the plate behind it. C. A plate of lithosphere slides down a gently inclined surface of asthenosphere.

Guest Essay

Harry Hess and Global Tectonics

Charles L. Drake was a slow learner, entering Princeton in 1941 and graduating in 1948. He obtained his Ph.D at Columbia in 1958 and remained on the Columbia faculty until 1969, when he moved to Dartmouth.

I was introduced to Harry Hess's ideas about the nature of global deformation in a course he taught at Princeton after World War II, entitled Advanced General Geology. These ideas had developed over many years, based in part on observations from his diverse field and naval experiences. They reached fruition in an influential premise that some form of convection drives new crust formation along the ocean's ridge.

Hess's ideas began with his study of ultramafic rocks in the Appalachians for his Ph.D. dissertation. Because igneous rocks are composed mainly of pyroxene, amphibole, and olivine, ultramafics arise from magnetic origins. At the time, however, the origin of the ultramafic rocks associated with mountain systems was an enigma. The chemistry of these rocks and the very high temperatures required to melt them made it unlikely that they came from a magma, but field relationships led investigators to suggest that they did indeed come from magmas. A possible solution to the problem came from early submarine gravity expeditions in which Hess participated.

The data from these expeditions found that large negative gravity anomalies were associated with island arc systems and their deep-sea trenches. These anomalies, where the attraction of gravity was considerably less than expected for a uniform earth, represented mass deficiencies that could not be explained by the trenches themselves. They appeared to be due to a downbuckling of the less dense outer shell of the Earth into the more dense rocks of the interior, a feature that Hess named a "geotectocline." He further noted that the bands of ultramafic rocks and negative gravity anomalies of the island arcs were matched by similar features associated with mountain systems, and he concluded that these were produced by similar processes. The surficial rocks were drawn into the downbuckle and deformed and metamorphosed to form the new mountains. The mechanism suggested as a cause of the downbulges was convection resulting from irregular distribution of heat in the Earth's interior. The ultramafic rocks would be squeezed up to shallow depths and, during the deformation process, hydrated to serpentine, a rock composed largely of the mineral serpentine. Thus, serpentines in this model are tectonically emplaced.

Early in the post-war years, expeditions to the Mid-Atlantic ridge led by Maurice Ewing recovered ultramafic rocks, particularly serpentines. Shortly afterward the first seismic refraction measurements at sea showed that the ocean crust was thin and remarkably uniform in its seimic properties, with little difference between the crust of the ocean basins and that of the ridges. The elevation of the ridges could not be explained in terms of crustal thickness, but Hess suggested that it might be accounted for by serpentinization of the ultramafic mantle rocks. If water were leaking outward slowly from the interior of the Earth, when it ascended to a depth where the temperature reached 500°C it would react with the olivine in the rocks to form serpentinite. A considerable volume increase would occur as the result of this reaction, and the Earth's surface over the area would rise. The location of the ridge might be explained as being over the rising limb of a convection cell in the Earth, just as the "geotectocline," or tectogene as it was later called, was produced over the descending limb.

Hess sent samples of ultramafic rocks ranging from zero to 100 percent serpentinite to Francis Birch at Harvard, who measured the velocity of sound in them and noted that the range of seismic velocities reported for the ocean crust and upper mantle could be duplicated by the degree of serpentization of these ultramafics. This, together with Birch's earlier predilection for convection, led him to develop a model for the ocean crust and upper mantle. In this model, the crust and upper mantle consist of serpentinized and unaltered ultramafic rocks. The formation of this crust involved rising limbs of convection cells at the axes of the ridges. Along the margins of some oceans, particularly the Pacific, the ocean crust descends back into the Earth's interior, and the serpentinite, upon reaching the 500° isotherm, is converted back into mantle periodite with a concomitant release of large quantities of water. With some modifications this basic model became known as seafloor spreading.

It is fair to say that Hess's interest in ultramafic rocks associated with mountain systems was almost destined to lead him into concepts of global tectonics. These rocks, part of a suite of mafic and ultramafic rocks called ophiolites and interpreted as old ocean crust and upper mantle, elevated above the sea surface in subduction zones, are key elements of the plate tectonics model and the only remaining remnants of ocean crust older than 200 million years. Hess's newly formed beliefs may well have led him and his fellow members of the American Miscellaneous Society in the 1950s to conceive of the MOHOLE project designed to test the current models by drilling a hole through the entire ocean crust and into the mantle. This project never came to fruition, but led to the Deep Sea Drilling Project and its successors, which have revolutionized our knowledge of the nature of ocean crust and the history of the ocean basins.

Hess's ideas were not always right in detail, but at the time they were advanced they were based solidly on the available data. As Charles Darwin pointed out many years ago in *Descent of Man* (1871), "False facts are highly injurious to the progress of science, for they often endure long; but false views, if supported by some evidence, do little harm, for everyone takes a salutary pleasure in proving their falseness; and when this is done, one path towards error is closed and the road to truth is often at the same time opened." Harry's views were always based on solid data, and his ideas were imaginative, stimulating, and consistent with the facts of the day. Although many have taken salutary pleasure in testing his models and developing new or modified models based on more advanced and difficult kinds of data, many of these owe their genesis to his original thinking.

Summary

1. Alfred Wegener proposed a hypothesis of continental drift in the early years of the twentieth century, but because he could not explain how continents could move, his hypothesis was not widely accepted.
2. Studies of the Earth's magnetism led to a revival of the continental drift hypothesis in the 1950s, and this eventually led to the plate tectonics hypothesis in the 1960s.
3. The source of the Earth's magnetism is the molten outer core.
4. As minerals such as magnetite, which can become permanent magnets, cool through the Curie temperature, they acquire magnetism with the polarity, inclination, and direction of the Earth's field.
5. Paleomagnetism of lavas and sedimentary rocks reveals that either the magnetic poles have wandered in the past (apparent polar wandering) or the continents have drifted. Drifting is the correct conclusion.
6. Seafloor spreading is a hypothesis that new oceanic crust is created at midocean ridges by magma rising from deep inside the Earth.
7. The seafloor spreading hypothesis was proved correct by the record of magnetic polarity reversals recorded in the oceanic crust.
8. The lithosphere is broken into six large and many smaller plates, each about 100 km thick and each slowly moving over the top of the weak asthenosphere beneath it.
9. Three kinds of margins are possible between plates. Divergent margins (spreading centers) are those where new lithosphere forms; plates move away from them. Convergent margins (subduction zones) are lines along which lithosphere capped by oceanic crust is subducted back into the mantle and where continental crust on adjacent plates collides to form a collision margin. Transform fault margins are lines where two plates slide past each other.
10. Divergent margins form when a continent splits because of thermal stresses. As the split grows wider, an ocean forms.
11. Collision zones are the places where oceans disappear.
12. The continental shelf is the flooded margin of a continent. The geological edge of a continent is the bottom of the continental slope, where oceanic crust meets continental crust.
13. A deep-sea trench forms where lithosphere sinks into the asthenosphere as a result of subduction. Parallel to the trench and offset by about 150 km, a line of andesitic volcanoes forms an island arc. The volcanoes form because subducted oceanic crust starts melting at a depth of about 100 km.
14. Plates move as a combination of pull exerted by the sinking, cold tongue of lithosphere, push exerted by magma rising at the spreading center, and by lithosphere sliding down the gently sloping lithosphere-asthenosphere boundary.

Important Terms to Remember

abyssal plain (p. 149)
collision margin (p. 147)
continental rise (p. 148)
continental shelf (p. 148)
continental slope (p. 148)
convergent margin (p. 147)
Curie point (p. 138)
divergent margin (p. 147)
paleomagnetism (p. 139)
seafloor spreading (p. 140)
spreading center (p. 147)
subduction zone (p. 147)
transform fault margin (p. 147)

Questions for Review

1. Who was Alfred Wegener and what revolutionary idea did he suggest? Why were scientists reluctant to accept Wegener's idea when it was first proposed?
2. Explain how the apparent wandering of magnetic poles throughout geologic history can be used to help prove continental drift.
3. How does lava carry a record of the Earth's magnetic field?
4. What are the main features of seafloor spreading? What critical test proved that the seafloor does move?
5. What is a "plate" in plate tectonics?
6. Briefly describe the three kinds of plate margins.
7. Describe what happens when two plates topped by oceanic crust converge. Compare your description with what happens when the converging is between two plates capped by continental crust.
8. How do satellite measurements help confirm the theory of plate tectonics?
9. Identify the major topographic features of the ocean floor, and state how they are related to tectonic plates.
10. How can earthquakes be used to outline the shape of plates and locate plate margins?
11. Draw a cross section through the lithosphere at a convergent plate margin and mark the positions of the fore-arc basin, the back-arc basin, and the magmatic arc.
12. The Himalaya and the Alps are said to be "graveyards" of ancient oceans. Why?
13. What causes earthquakes along a transform fault?
14. Where does the energy come from to cause the movement of tectonic plates?
15. What are the current theories of the causes of plate tectonics?
16. Why do andesitic volcanoes occur around the Pacific rim?

Questions for A Closer Look

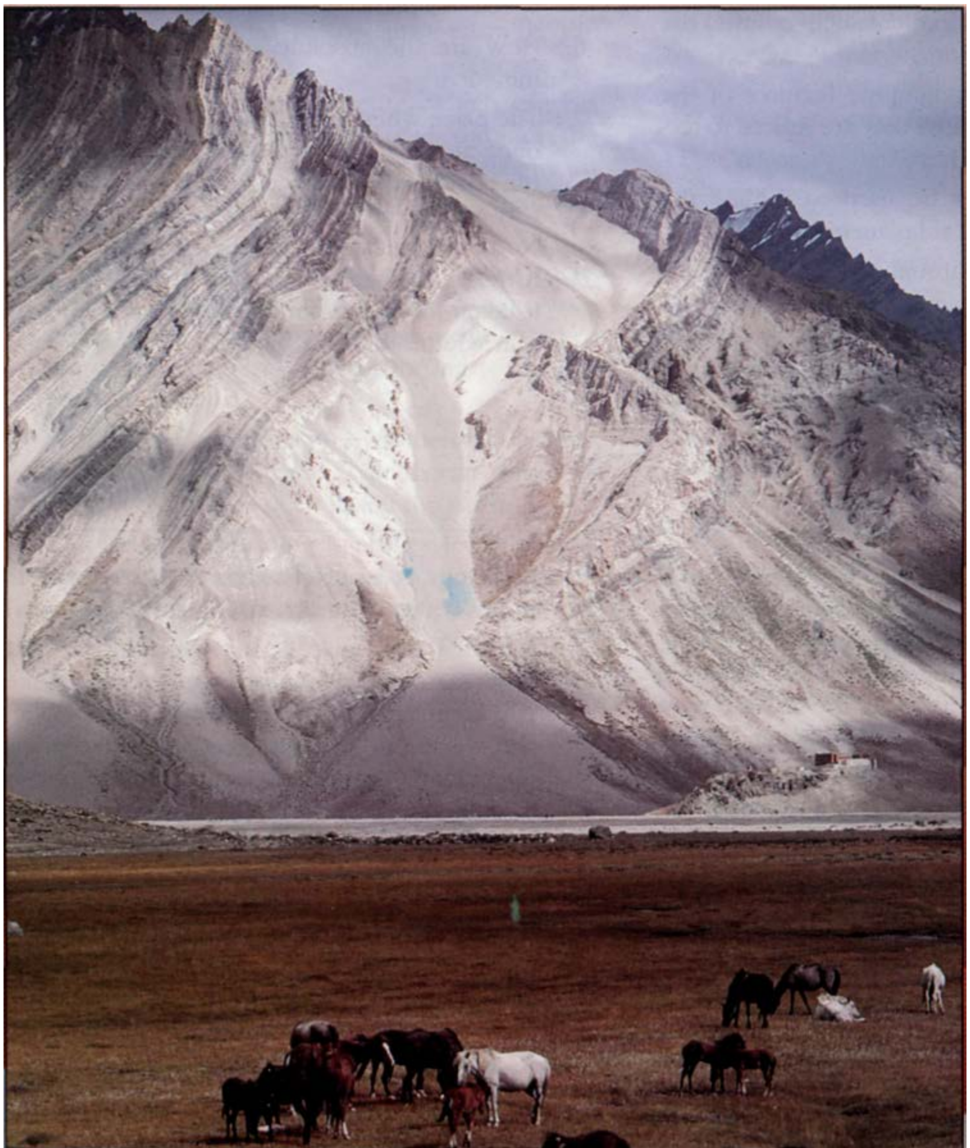
1. What is a spreading pole?
2. Why does the speed differ from place to place on a plate?
3. Where on a plate is the speed a maximum? Where a minimum?
4. How do we know that the Pacific Ocean must be getting smaller?
5. What is the difference between the relative speed of a plate and its absolute speed?
6. How are the absolute speeds of plates determined?
7. The place where New York City now stands was once attached to North Africa at approximately the position of Marrakech in Morocco. New York City and Marrakech are now 5700 km apart. The North Atlantic plate is moving away from the African plate at a speed of 2 cm/y. How long has it taken for the Atlantic Ocean to reach its present width?

Questions for Discussion

1. In the vicinity of Los Angeles, the Pacific plate is moving northerly relative to the North American plate at a speed of 5.5 cm/y. Determine how long it will be before Los Angeles and San Francisco are side by side. Draw a map of the way the west coast of North America might look when Los Angeles and San Francisco are side by side. Now draw a map 10 million years after the two towns are adjacent.
2. Although plate tectonics is manifested by such disasters as massive earthquakes and volcanic eruptions, it is nevertheless said that plate tectonics is a good thing for our planet. Discuss why this is so.
3. What key observations would you plan to make if you were sending a spaceship to another planet and wished to find out whether plate tectonics operated on the planet? Assume for the purpose of discussion that the spaceship cannot land and so all observations have to be made remotely.

CHAPTER 7

The Earth's Evolving Crust



The Zaskar Range, Ladakh, part of the great Himalayan mountain chain which formed as a result of the collision between India and Asia. These strata, once horizontal layers on the floor of the ancient Tethys Sea, were contorted and tilted as a result of the collision and elevated several thousand meters above the sea level.

Tectonics, Erosion, and the Rock Record

Recently, a team of Greek and British scientists discovered that tectonic forces are stretching Greece and slowly making it grow larger!

A century ago the distances between a series of Greek survey monuments were measured very accurately. In 1988 a scientific team remeasured the distances and found that Greece is now a meter longer. They also discovered that Greece is being twisted so that the southern end, the Peloponnesus, is moving to the southwest relative to the rest of Greece. The reason for the stretching and twisting is that a slice of Mediterranean seafloor is being slowly forced under Greece.

Because Greece is being stretched, we conclude that rock in the Greek crust is being deformed. There is nothing unique or unusual about this conclusion. Evidence that rocks can be deformed is easy to find. If you look at a photograph of the Alps, the Rockies, the Appalachians, or any other mountain range, you will see once horizontal layers of sedimentary rocks that are now tilted and bent.

As we saw in Chapter 6, the source of energy for tectonic forces is the Earth's heat energy. The huge, slow, convective flows of hot rock in the mesosphere and asthenosphere continuously buckle and warp the lithosphere. It is those convective forces that are ultimately the cause of the rock deformation we observe in mountain ranges and that are stretching Greece. Convection is slowly but steadily changing the external face of the Earth reservoir, and in response to those changes the other reservoirs, the hydrosphere, atmosphere, and biosphere, have to adjust and change. Where slow, long-term changes to the Earth systems are concerned, it is the solid Earth reservoir that calls the shots.

The colliding, twisting, grinding, and stretching of the crust caused by tectonic forces has been continuous throughout the Earth's long history. However, just as tectonic forces have caused mountains to be raised, so has erosion worn away the uplifted rocks, and then water, wind, and ice have transported the debris and formed sediment. Sediment becomes sedimentary rock, and, when continents collide, sedimentary rock becomes metamorphic rock. The history of all the tectonic bumps and stretches—indeed, the long history of the Earth itself—is read both from the debris of erosion and from deformed rocks, which is all of the sedimentary rocks and the metamorphic rocks formed from them. We therefore begin this chapter by discussing how the ages of rocks are determined; then we turn to sediments and sedimentary rocks, followed by metamorphism and metamorphic rocks; and we end the chapter with a discussion of the structure of continents, how they have grown and changed. (For an alternative view of how some of the Earth's features might be explained, see the "Guest Essay" at the end of this chapter.)

SEDIMENTARY STRATA

Like a perpetually restless housekeeper, nature is ceaselessly sweeping regolith off the solid rock beneath it, carrying the sweepings away and depositing them as sediment in river valleys, lakes, and innumerable other places. We can see sediment being transported by trickles of water after a rainfall and by every wind that carries dust. The mud on a lake bottom, the sand on a beach, even the dust on a windowsill are all sediment. Because erosion and deposition of rock particles take place almost continuously, we find sediment nearly everywhere. By looking closely, it is possible to see that sediment, regardless of where it is deposited, is piled layer on layer. Studies of the layers reveal a great deal about the way the Earth works.

Sedimentary **stratification** results from the arrangement of sedimentary particles in layers (Fig. 7.1). Each sedimentary *stratum* (pl. **strata)** is a distinct layer of sediment that accumulated at the Earth's surface. The layered arrangement of strata in a body either of sediment or of sedimentary rock is referred to as **bedding.** Each **bed** in a succession of strata can be distinguished from adjacent beds by differences in thickness or character.

Stratigraphy

The historical information that geologists work with is largely in the form of layered sedimentary rocks that crop out at the Earth's surface or that can be penetrated by drilling. Examine the rocks shown in Figure 7.1. You will see distinct differences in the thicknesses and colors of the many layers. The differences arise from changes in the environment as the sediments accumulated. Because sedimentary rocks carry important clues about past environments at the Earth's surface, the sequence and age of strata provide the basis for reconstructing much of the Earth's environmental history.

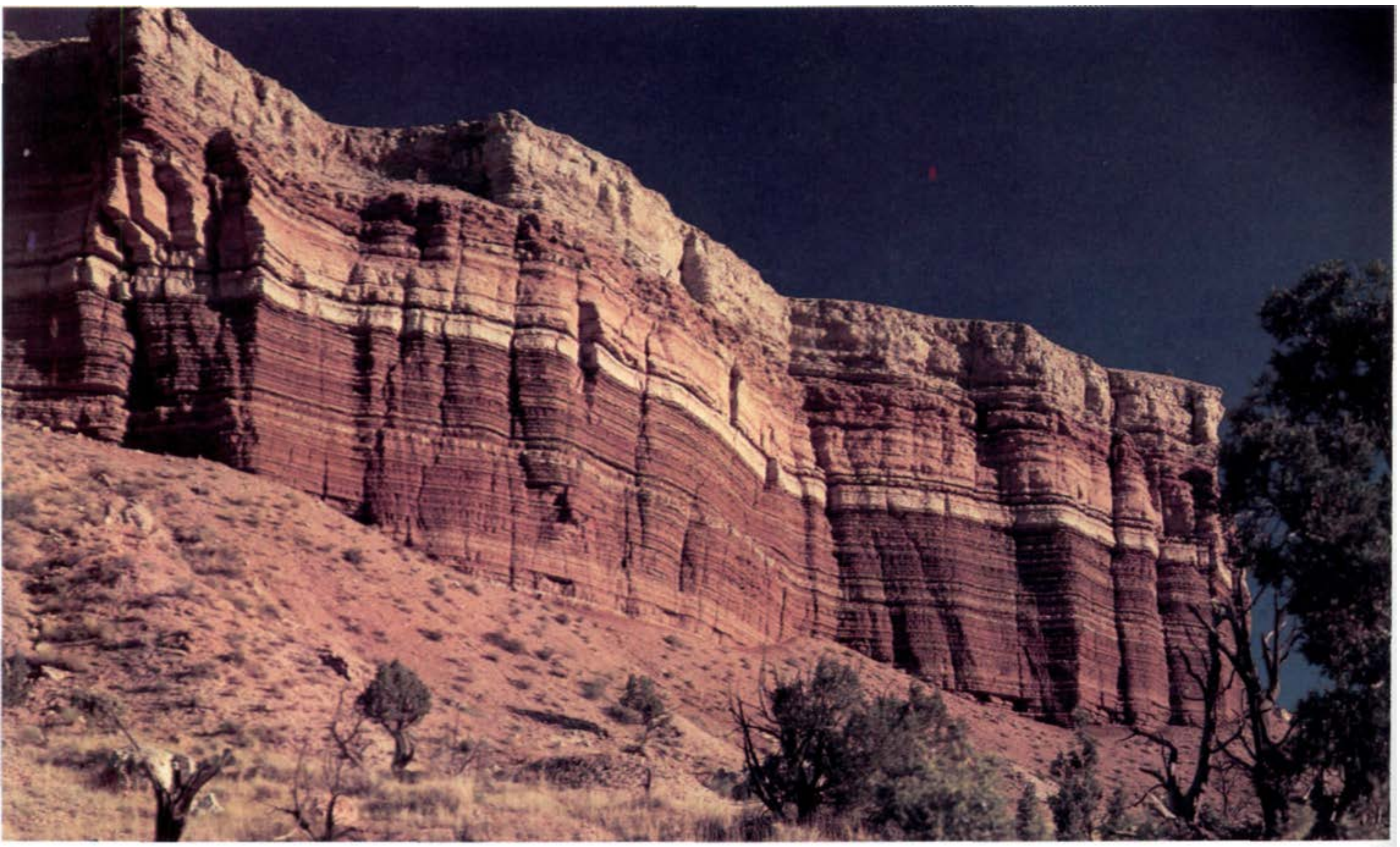


Figure 7.1 Multicolored sedimentary rocks in Capital Reef National Park, Utah. Each layer is a separate stratum.

The study of strata is called **stratigraphy.** Two straightforward and simple, but nevertheless very powerful, laws underlie stratigraphy. The first is the **law of original horizontality,** which states that sediments are deposited in strata that are horizontal or nearly so and parallel to the Earth's surface. From this generalization we can infer that rock layers now inclined, or even buckled and bent, must have been disturbed since the time they were deposited.

The second law is the **principle of stratigraphic superposition,** which states that in any sequence of sedimentary strata the order in which the strata were deposited is from the bottom to the top. The red strata at the bottom of Capital Reef (Fig. 7.1) are older than the light-brown strata at the top of the reef.

Breaks in the Stratigraphic Record

A pile of strata deposited layer after layer without any interruption is said to be *conformable.* Commonly, however, there are substantial breaks or gaps in a sedimentary record. These represent times of nondeposition to which the term **unconformity** is applied. An unconformity records a change in either environmental conditions that caused deposition to cease for a considerable time, or erosion that resulted in loss of part of an earlier-formed depositional record, or a combination of both.

There are three important kinds of unconformities. The first, labeled (1) in Figure 7.2, is a *nonconformity*, where strata overlie igneous or metamorphic rocks. The second and most obvious is the *angular unconformity,* which is marked by angular discontinuity between older and younger strata. It is labeled (2) in Figure 7.2. An angular unconformity implies that the older strata were deformed and then truncated by erosion before the younger layers were deposited across them. The outcrop at Siccar Point, discussed in the Introduction (Fig. 1.4) and used by James Hutton in his hypothesis of the rock cycle, is obviously an angular unconformity. The third kind of unconformity is called a *disconformity;* it is an irregular surface of erosion between parallel strata, and it implies a cessation of sedimentation, as well as erosion, but no tilting. The surface numbered (3) in Figure 7.2 is a disconformity.

A study of unconformities reveals the close relationship between tectonics, erosion, and sedimentation. All of the Earth's land surface is a potential surface of unconformity. Some of today's surface will be destroyed by erosion, but other parts will be covered by sediment and preserved as a record of the present landscape. For example, the Swiss Alps, which were elevated by plate tectonic movements, are now being eroded away. Meanwhile, the eroded material is being carried away by streams and deposited in the Mediterranean Sea. The Mediterranean seafloor was once dry land, but tectonic forces depressed it, just as tectonic forces elevated the Alps. A surface of unconformity separates the young, river-transported sediments and the older rocks of the seafloor on which the sediments are being piled. In a sense, accumulation in one place compensates for destruction in another. As James Hutton recognized, unconformities provide powerful evidence that interactions between the solid Earth on the one hand, and the atmosphere, hydrosphere, or biosphere on the other hand, have been going on throughout the Earth's long history.

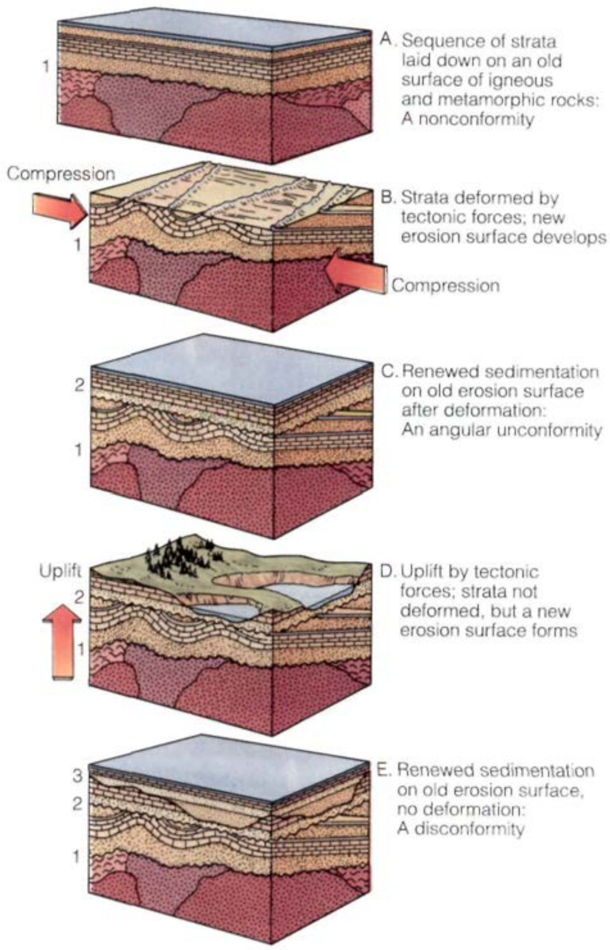


Figure 7.2 Sequence of geologic events leading to the three kinds of unconformity: (1) nonconformity; (2) angular unconformity; and (3) disconformity.

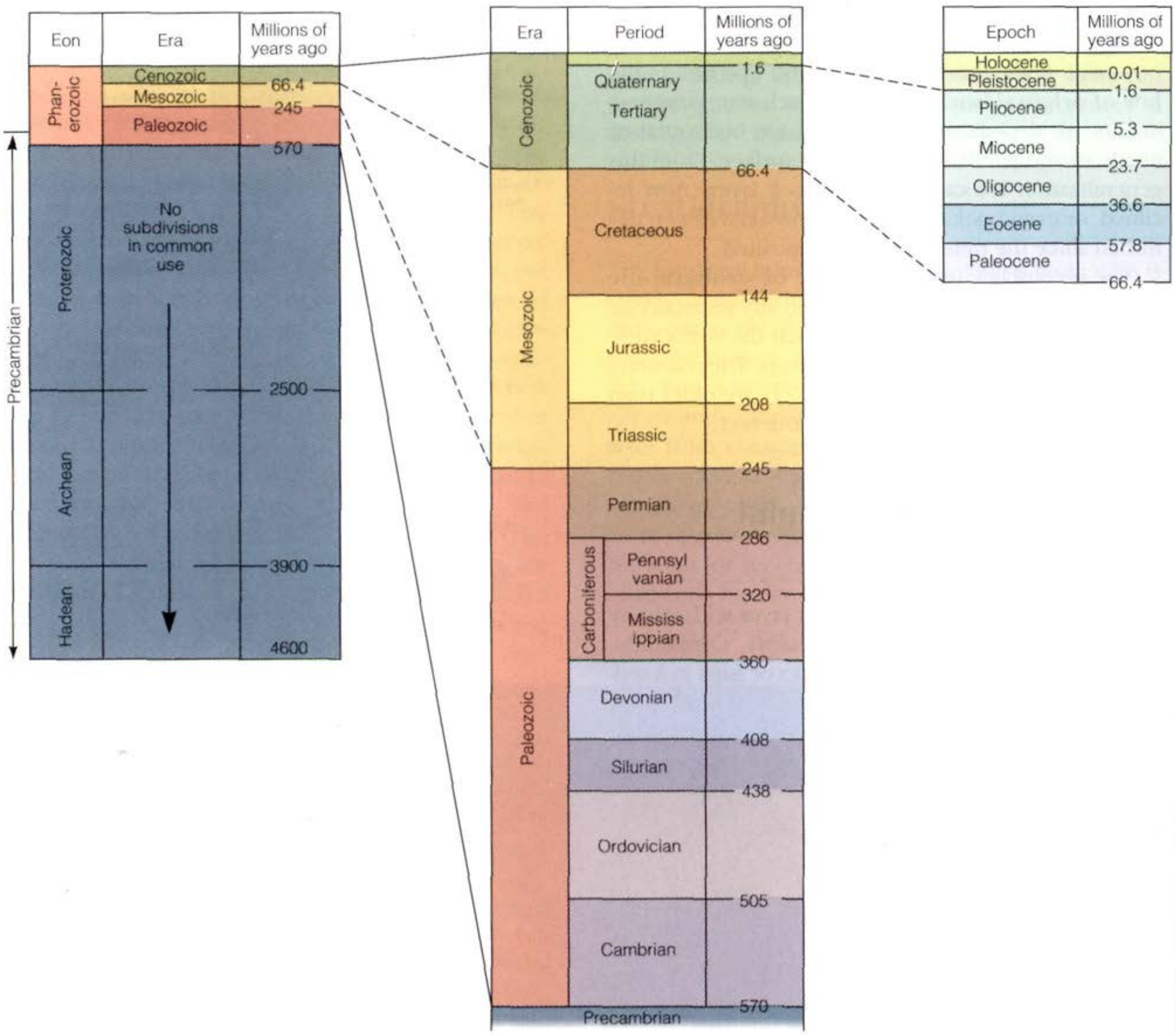


Figure 7.3 The geologic time scale. Absolute ages obtained from radiometric dates. Note that the Pennsylvanian and Mississippian periods are equivalent to the Carboniferous Period of Europe. The time boundary between the Archean and Hadean is uncertain, for no rocks of the Hadean Eon are known on the Earth. Hadean rocks are known to exist on other planets in the solar system.

Stratigraphic Correlation

Any distinctive stratum or group of strata that differ from the strata above and below are given a name. The strata in Figure 7.1, for example, are called the Navajo Sandstone. Early in the nineteenth century an English land surveyor, William Smith, while surveying for the construction of new canals, realized that distinctive sedimentary strata throughout western England lay, as he put it, "like slices of bread and butter" in a definite, unvarying sequence. He became familiar with the characteristics of each layer, especially the fossils each contained, and with the sequence of the layers. By looking at a specimen of sedimentary rock collected from anywhere in southern England, he could name the stratum from which it had come and, of course, the position of the stratum in the sequence.

Smith did not believe that his discovery reflected any particular scientific principle; he thought it was purely practical. Nevertheless, it opened the door to the correlation of sedimentary strata over increasingly wide areas. *Correlation* means the determination of equivalence in age of the succession of strata found in two or more different areas. Smith correlated strata

Table 7.1
Origin of Names for Periods of the Paleozoic, Mesozoic, and Cenozoic Eras, and the Epochs of the Quaternary and Tertiary Periods

Era	Period	Epoch	Origin of Name
Cenozoic	Quaternary[a]	Holocene	Greek for wholly recent
		Pleistocene	Greek for most recent
	Tertiary[a]	Pliocene	Greek for more recent
		Miocene	Greek for less recent
		Oligocene	Greek for slightly recent
		Eocene	Greek for dawn of the recent
		Paleocene	Greek for early dawn of the recent
Mesozoic	Cretaceous	-	Latin for chalk, after chalk cliffs of southern England and France
	Jurassic	Epoch	Jura Mountains, Switzerland, and France
	Triassic	Names	Threefold division of rocks in Germany
Paleozoic	Permian	Used	Province of Perm, Russia
	Pennsylvanian	Only	State of Pennsylvania
	Mississippian	by	Mississippi River
	Devonian	Specialists	Devonshire, county of Southwest England
	Silurian		Silures, ancient Celtic tribe of Wales
	Ordovician		Ordovices, ancient Celtic tribe of Wales
	Cambrian		Cambria, Roman name for Wales

[a]Derived from eighteenth- and nineteenth-century geological time scale that separated crustal rocks into a fourfold division of Primary, Secondary, Tertiary, and Quaternary, based largely on relative degree of lithification and deformation.

initially over distances of several kilometers and later over tens of kilometers. By means of fossils in the sedimentary rocks, it ultimately became possible to correlate through hundreds and then thousands of kilometers.

THE GEOLOGIC COLUMN

Geologists deal with two kinds of time, relative and absolute. Relative time is the order in which a sequence of past events occurred, whereas absolute time is the time in years, when a specific event happened.

One of the great successes of the nineteenth-century geologists was the demonstration, through stratigraphic correlation, that the relative ages of stratigraphic sequences are the same on all continents. Through worldwide correlation, those nineteenth-century geologists assembled a **geologic column,** which is a composite columnar section containing in chronological order the succession of known strata, fitted together on the basis of their fossils or other evidence of relative age.

Standard names have evolved for the subdivisions of the geologic time units corresponding to the rock units of the geologic column. The units of the geologic time scale, which, like the geologic column, can be used worldwide, are eons, eras, periods, and epochs as shown in Figure 7.3 and Table 7.1.

The scientists who worked out the geologic column and time scale were challenged by the question of absolute time. They knew the relative time order in which strata of the geologic column had formed, but they also wished to know whether the sediments in the strata had accumulated during the same length of time. They sought answers to questions such as these: "How much time elapsed between the end of the Cambrian Period and the beginning of the Permian Period?" "How long was the Tertiary Period?" Absolute ages must be determined in order to answer such questions as the age of the Earth, the age of the ocean, how fast mountain ranges rise, and how long humans have inhabited the Earth.

The discovery of radioactivity in 1896 provided a reliable way to measure absolute geologic time. Radioactivity is a process that runs continuously, that is not reversible, that operates the same way and at the same speed everywhere, and that leaves a continuous record without any gaps in it. For a discussion of how radioactivity is used to measure absolute ages, see "A Closer Look: Radioactivity and the Measurement of Absolute Time."

A Closer Look

Radioactivity and the Measurement of Absolute Time

We learned in Chapter 4 that most chemical elements have several naturally occurring isotopes (atoms with the same atomic number and hence the same chemical properties, but different mass numbers). Most isotopes found in the Earth are stable and not subject to change. However, a few, such as carbon-14 (^{14}C), are radioactive because of an instability in the nucleus and will transform spontaneously to either a more stable isotope of the same chemical element or an isotope of a different chemical element—^{14}C, for example, expels an electron from its nucleus and transforms to ^{14}N.

The rate of transformation—*radioactive decay* rate as it is now more commonly called—is different for each isotope. Careful study of radioactive isotopes in the laboratory has shown that decay rates are unaffected by changes in the chemical and physical environment. Thus, the decay rate of a given isotope is the same in the mantle or a magma as it is in a sedimentary rock. This is a particularly important point because it leads to the conclusion that rates of radioactive decay are not changed by geologic processes and therefore can be used to measure absolute time.

All decay rates follow the same basic law that is depicted in Figure C7.1. The law of radioactive decay, stated in words, is that the *proportion* of parent atoms that decay during each unit of time is always the same. The number of decaying parent atoms continuously decreases, while the number of daughter atoms continuously increases.

The rate of radioactive decay is measured by the *half-life,* which is the time needed for the number of parent atoms to be reduced by one-half. For example, if the half-life of a radioactive isotope is 1 hour and we started an experiment with a mineral containing 1000 radioactive atoms, at the end of an hour only 500 parent atoms would remain and 500 daughter atoms would have formed. At the end of a second hour there would be 250 parent and 750 daughter atoms, and after hour 3, 125 parents and 875 daughters. The half-lives of radioactive isotopes used to measure absolute geologic times are

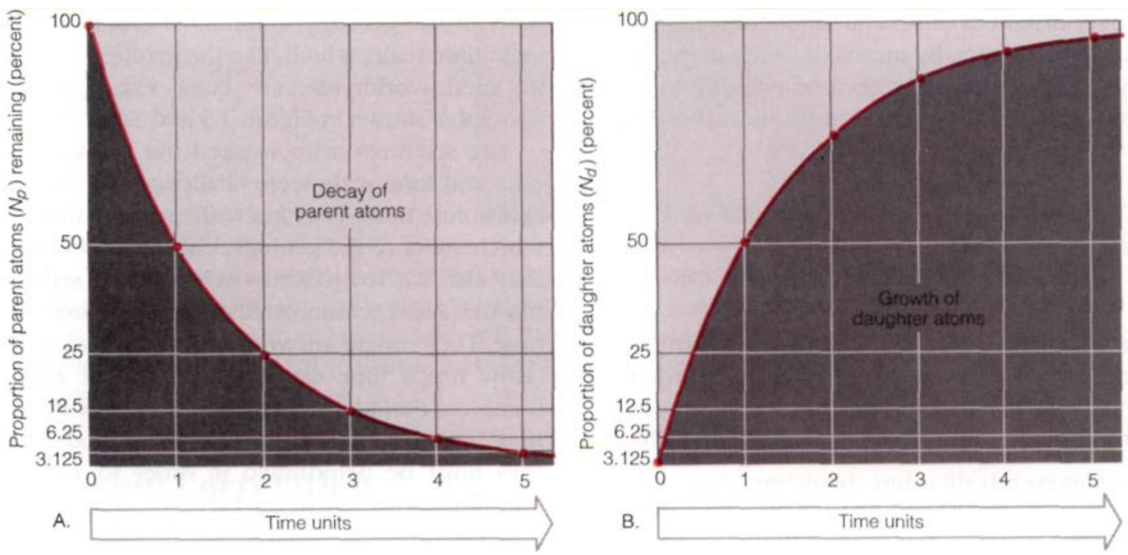


Figure C7.1 Curves illustrating the basic law of radioactivity. A. At time zero, a sample consists of 100 percent radioactive parent atoms. During each time unit, half the atoms remaining decay to daughter atoms. B. At time zero, no daughter atoms are present. After one time unit corresponding to a half-life of the parent atoms, 50 percent of the sample has been converted to daughter atoms. After two time units, 75 percent of the sample is daughter atoms and 25 percent parent atoms. After three time units, the percentages are 87.5 and 12.5, respectively. Note that at any given instant N_p, the number of parent atoms remaining, plus N_d, the number of daughter atoms, equals N_0, the number of parent atoms at time zero.

thousands to millions of years long, but the decay law is the same for all isotopes regardless of the length of the half-life.

In the graphic illustration of radioactive decay in Figure C7.1, the time units marked are half-lives. Of course, the time units are of equal length, but at the end of each unit the number of parent atoms, and therefore the radioactivity of the sample, has decreased by exactly one-half of the value at the beginning of the unit. Figure C7.1 also shows that the growth of daughter atoms just matches the decline of parent atoms. When the number of remaining parent atoms (N_p) is added to the number of daughter atoms (N_d), the result is N_0, the number of parent atoms that a mineral sample started with. That fact is the key to the use of radioactivity as a means of measuring geologic time and determining ages.

Potassium-Argon ($^{40}K/^{40}Ar$) Dating

We have selected one of the naturally radioactive isotopes, potassium-40 (^{40}K), to illustrate how the absolute time of formation of certain minerals can be determined. Potassium has three natural isotopes: ^{39}K, ^{40}K, and ^{41}K. Only one, ^{40}K, is radioactive, and its half-life is 1.3 billion years. The decay of ^{40}K is interesting because two different decay schemes occur. Twelve percent of the ^{40}K atoms decay to ^{40}Ar, an isotope of the gas argon. The remaining 88 percent of the ^{40}K atoms decay to ^{40}Ca. It is important to know that the fraction of ^{40}K atoms decaying to ^{40}Ar is always 12 percent; the percentage is not affected by changes in physical or chemical conditions.

When a potassium-bearing mineral crystallizes from a magma, or grows within a metamorphic rock, it includes some ^{40}K in its crystal structure. As soon as the mineral is formed, ^{40}Ar and ^{40}Ca daughter atoms start accumulating in the mineral, because they are trapped, like the parent ^{40}K atoms, in the crystal structure. Because the ratio of ^{40}Ar to ^{40}Ca daughter atoms is always the same, it is only necessary to measure either ^{40}Ar or ^{40}Ca daughter atoms in order to know how many ^{40}K atoms have decayed. It is more convenient to measure ^{40}Ar because argon is an element that can be measured very accurately.

All that has to be done to determine the absolute time of eruption of an extrusive igneous rock is to select a potassium-bearing mineral in the rock and measure the amount of parent ^{40}K that remains, as well as the amount of trapped ^{40}Ar. With the half-life of ^{40}K known, it is a straightforward matter to calculate the radiometric age—the length of time a mineral has contained its built-in radioactivity clock. What is actually measured, of course, is the time since the mineral formed in a cooling magma. Because the time of mineral formation is effectively the time at which the extrusive igneous rock was formed, the mineral age and the rock age are the same.

Absolute Time and the Geologic Time Scale

Many naturally radioactive isotopes can be used for radiometric dating, but six predominate in geologic studies. These are the two radioactive isotopes of uranium as well as the single radioactive isotopes of thorium, potassium, rubidium, and carbon. These isotopes occur widely in different minerals and rock types, and they have a very wide range of half-lives, so that many geologic materials can be dated radiometrically.

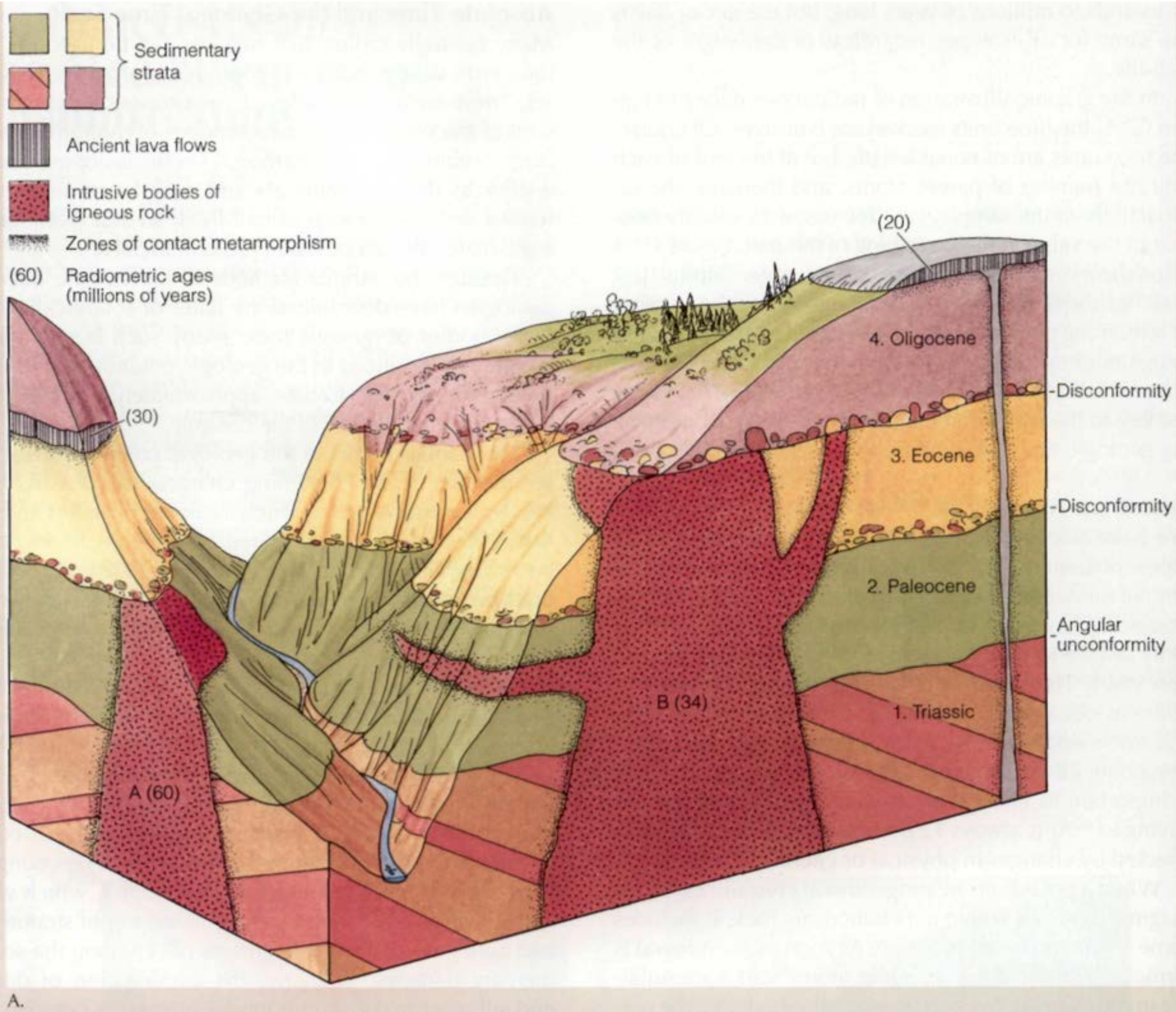
Through the various methods of radiometric dating, geologists have determined the dates of solidification of many bodies of igneous rock. Many such bodies have identifiable positions in the geologic column; as a result, it becomes possible to date, approximately, a number of the sedimentary layers in the column.

The standard units of the geologic column consist of sedimentary strata containing characteristic fossils, but the typical rocks from which radiometric dates (other than ^{14}C dates) are determined are igneous rocks. It is necessary, therefore, to be sure of the relative time relations between an igneous body that is datable and a sedimentary layer whose fossils closely indicate its position in the column.

Figure C7.2A and B shows how ages of sedimentary strata are approximated from the ages of igneous bodies. In Figure C7.2A, a sequence of sedimentary strata containing fossils of known relative ages is separated by an unconformity and two disconformities. Intrusive igneous rock *A* cuts strata 1 and 2 but is truncated by the disconformity at the top of stratum 2. Thus, *A* must be younger than strata 1 and 2 but older than stratum 3, which was laid down on the erosion surface at the top of stratum 2 and contains weathered fragments of *A* among the sedimentary particles. Similarly, the combination of dikes and sills that make up the intrusive igneous complex 6 are truncated by the disconformity at the top of stratum 3, and they must be younger than stratum 3 but older than stratum 4. Lava flow C above the disconformity at the top of stratum 3 must also be younger than stratum 3 and younger than the dike-sill complex *B.* Lava flow C must be older than stratum 4, however, because it is covered by stratum 4, and lava flow *D* must be even younger because it overlies stratum 4.

From the radiometric dates of the igneous bodies and the relative ages of the geologic relations shown in Figure C7.2A, inferences can be drawn about the ages of the sedimentary strata as shown in Figure C7.2B.

Through a combination of geologic relations and radiometric dating methods, twentieth-century scientists have been able to fit a scale of absolute time to the geologic column worked out in the nineteenth century. The scale is being continuously refined, and so the numbers given in Figure 7.3 should be considered the best available now. Further work will make them more accurate.



Stratum	Age (millions of years)	Interpretation
4	<34 (younger than *B*) <30 (younger than *C*) >20 (older than *D*)	Age lies between 20 and 30 million years
3	<60 (younger than *A*) >34 (older than *B*) >30 (older than *C*)	Age lies between 34 and 60 million years
2	>60 (older than *A*)	Age of both is greater than 60 million years
1	>60 (older than *A*)	

B.

Figure C7.2 The application of radiometric dating to the geologic column. For method, see the text discussion. A. Idealized section showing sedimentary strata, angular unconformity, disconformities, lava flows, and igneous intrusions. Numbers in brackets are the radiometric ages of the igneous rocks determined by the potassium-argon method. B. Interpretation of the absolute ages of the four groups of sedimentary strata labeled 1 to 4.

SEDIMENT AND SEDIMENTARY ROCK

There are two families of sediment, clastic and chemical. The principal difference between them is the way the sediment is transported.

> **Clastic sediment** (from the Greek word *klastos,* meaning broken) is simply bits of broken rock and minerals that are moved as solid particles. Any individual particle of clastic sediment is a *clast,* and *clasts* tend to be the rock-forming minerals, such as quartz and feldspar, that are most durable during erosion.
>
> **Chemical sediment** is transported in solution and deposited when the dissolved minerals are precipitated.

The transformation of sediment to sedimentary rock is called *lithification* (from the Greek *lithos,* meaning stone, hence stone-making). As discussed in Chapter 4, lithification happens either by the addition of a cement or by recrystallization of the sediment particles to a firm, coherent mass.

Clastic Sediment and Clastic Sedimentary Rock

Clast size is the primary basis for classifying clastic sediment and clastic sedimentary rock. Clastic sediment can be divided into four main classes, which from coarsest to finest are gravel, sand, silt, and clay (Fig. 7.4). Gravel is further classified on the basis of dominant clast size into boulder gravel, cobble gravel, and pebble gravel (Table 7.2). The names of the clastic sedimentary rocks corresponding to the various clastic sediments are **conglomerate**, **sandstone, siltstone,** and **shale** as listed in Figure 7.4 and Table 7.2.

Clastic sediment is transported in many ways. It may slide or roll down a hillside under the pull of gravity, or it may be carried by a glacier, by the wind, or by flowing water. In each case, when transport ceases, the sediment is deposited in a fashion characteristic of the transporting mechanism. Deposition occurs because of a drop in energy. Sediment transported by wind or water is deposited when the moving air or flowing water slows to a speed at which clasts can no longer be carried. In a general way, grain size in sediment moved by wind or water is related to the speed of the transporting agent: the faster the speed, the larger the clasts that can be moved. When the speed fluctuates, clast sorting occurs. For example, a rapidly flowing river will remove all the fine particles, leaving

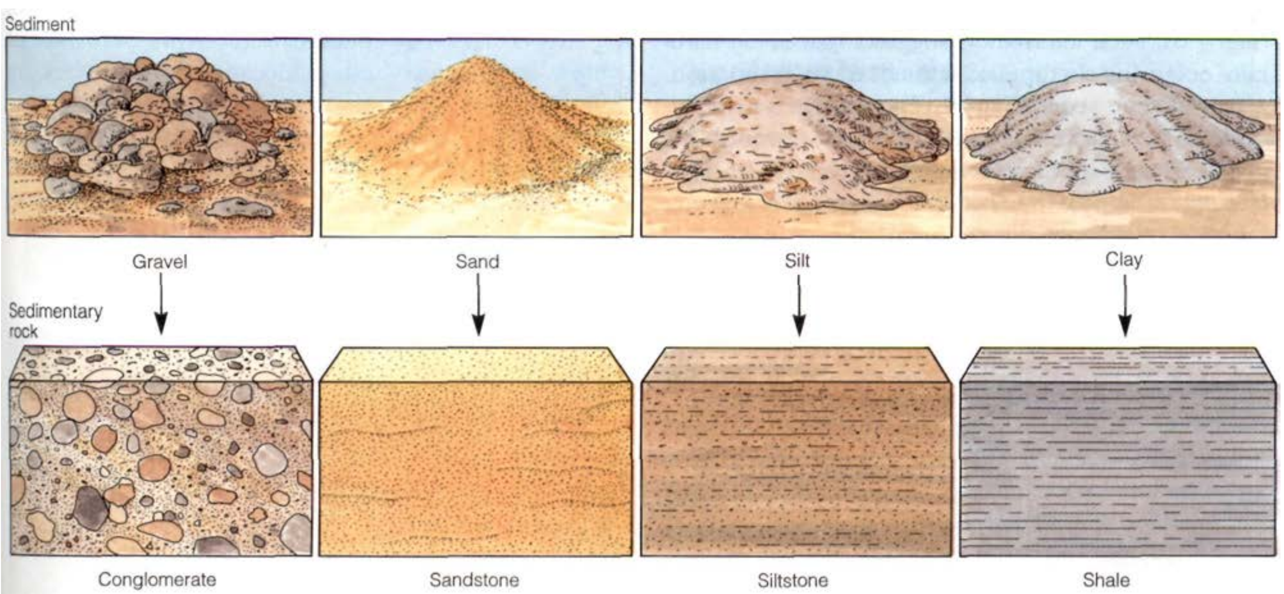


Figure 7.4 Principal kinds of clastic sediment and the sedimentary rocks formed from them. A. Sediment is classified and named for the sizes of the clasts. B. Rock is formed from clastic sediment.

Table 7.2
Definition of Clastic Particles Together with the Sediments and Sedimentary Rocks Formed from Them

Name of Particle	Size Limits of Diameter (mm)[a]	Names of Loose Sediment	Names of Consolidate Rock
Boulder	More than 256	Boulder gravel	Boulder conglomerate[c]
Cobble	64 to 256	Cobble gravel	Cobble conglomerate[c]
Pebble	2 to 64	Pebble gravel	Pebble conglomerate
Sand	1/16 to 2	Sand	Sandstone
Silt	1/256 to 1/16	Silt	Siltstone
Clay[b]	Less than 1/256	Clay	Shale

[a] Note that size limits of sediment classes are powers of 2, just as are memory limits in microcomputers (for example, 2K, 64K, 256K, 512K).

[b] Clay, used in the context of this table, refers to particle size. The term should not be confused with clay minerals, which are definite mineral species.

[c] If the clasts are angular, the rock is called a *breccia* rather than a conglomerate.

behind only the largest clasts. As the speed of the flowing water slows, smaller and smaller clasts are dropped and well-sorted sediment is the result (Fig. 7.5).

A sediment having a wide range of clast size is said to be poorly sorted. Such sediment is created, for example, by rockfalls, by the sliding of debris down hillslopes, by the slumping of loose deposits on the seafloor, by mudflows, and by deposition of debris from glaciers or from floating ice. Some poorly sorted sediments are given specific names (for example, *till* is a poorly sorted sediment of glacial origin, while the corresponding rock is a *tillite*).

Some clastic sediment displays a distinctive *alternation* of parallel layers having different clast sizes (Fig. 7.6). Such alternation suggests that some naturally occurring rhythm has influenced sedimentation. A pair of such sedimentary layers deposited over the cycle of a single year is termed a **varve** (Swedish for cycle). Varves are most common in deposits of high-latitude or high-altitude lakes, where there is a strong contrast in seasonal conditions. In spring, as the ice of nearby glaciers starts to melt, the inflow of sediment-laden water in a lake increases, and coarse sediment is deposited throughout the summer. With the onset of colder conditions in the autumn, streamflow decreases and ice forms over the lake surface. During winter, very fine sediment that has remained suspended in the water column slowly settles to form a thinner, darker layer above the coarse, lighter colored summer layer. Varved lake sediments are common in Scandinavia and New England where they formed beyond the retreating margins of ice age glaciers.

Cross bedding refers to sedimentary beds that are inclined with respect to a thicker stratum within which they occur (Fig. 7.7). Cross beds consist of clasts coarser than silt and are the work of turbulent flow in streams, wind, or ocean waves. As they are move along, the clasts tend to collect in ridges, mounds, or heaps in the form of ripples, waves, or dunes that migrate slowly forward in the direction of the current. Clasts accumulate on the downstream

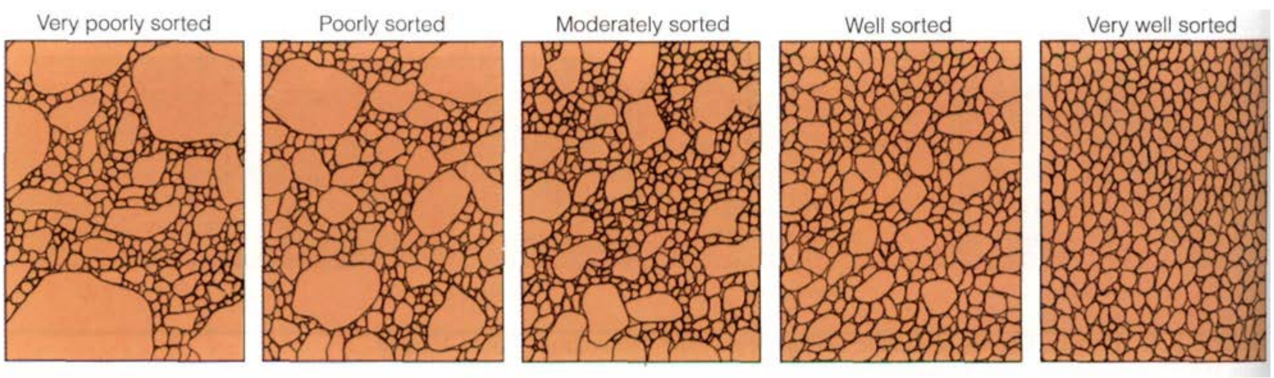


Figure 7.5 Clastic sediment ranges from very poorly sorted to very well sorted, depending on the extent to which the constituent grains are of equal size.

Figure 7.6 Varves deposited in a glacial age lake in southern Connecticut. Each pair of layers in a sequence of varves represents an annual deposit. Light-colored silty layers were deposited in summer, and the dark-colored clayey layers accumulated in winter.

Figure 7.7 Ancient cross-bedded sand dunes that have been converted to sedimentary rock that crops out near Kanab, Utah. The inclination of the cross beds shows that the ancient prevailing winds were blowing from left to right.

Figure 7.8 The fossil remains of a kauri pine lie next to the skeleton of a fossil fish on a bedding plane of a 175-million-year-old shale in Australia.

slope of the pile to produce beds having inclinations as great as 35°. The direction in which cross bedding is inclined tells the direction in which the related current of water or air was flowing at the time of deposition.

Many bodies of sediment contain **fossils,** the remains of plants and animals that died and were incorporated and preserved as the sediment accumulated. Sometimes the form of an original plant or animal is preserved (Fig. 7.8), but more commonly the remains are broken and scattered. As we will see in later chapters, especially chapters 14 to 16, fossils in sedimentary rocks provide important evidence about the history of the biosphere.

Chemical Sediment and Chemical Sedimentary Rocks

Chemical sediment forms when dissolved substances precipitate. One common example of precipitation involves the evaporation of seawater or lake water; as the water evaporates, dissolved matter is concentrated and salts begin to precipitate out as chemical sediment. Chemical sediment formed as the result of evaporation is called an *evaporite.* The most important are halite (NaCl) and gypsum ($CaSO_4 \cdot 2H_2O$); the corresponding rocks are salt and gyprock, respectively.

When chemical sediment forms as a result of biochemical reactions in water, the resulting sediment is said to be *biogenic.* One common example of a biogenic reaction involves tiny plants living in seawater; the plants remove dissolved carbon dioxide and thereby decrease the acidity of the surrounding water. As a result of decreased acidity, calcium carbonate is precipitated; such a precipitate is a biogenic sediment.

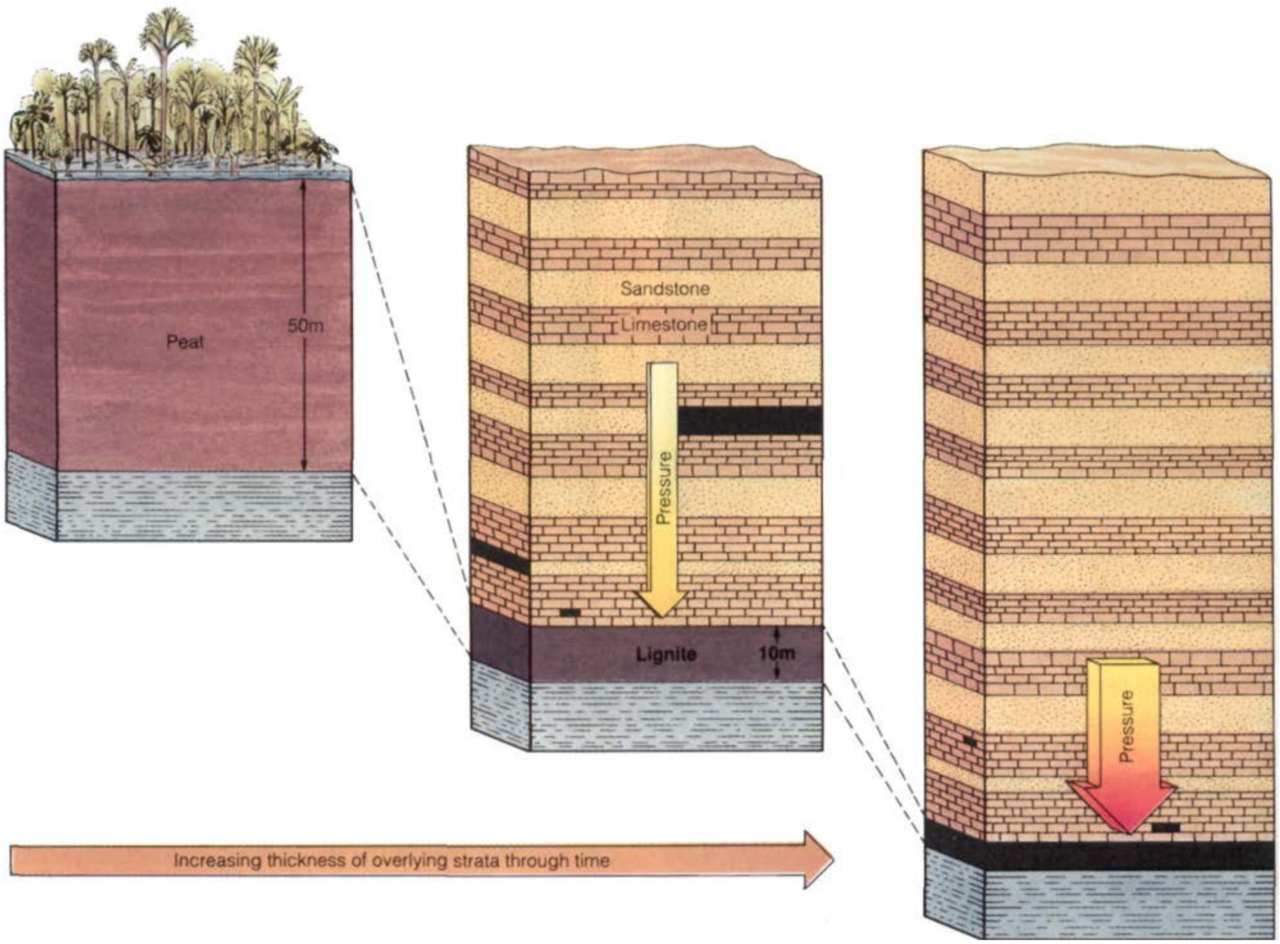


Figure 7.9 Plant matter in peat (a biogenic sediment) is converted to coal (a biogenic sedimentary rock) by decomposition and increasing pressure and temperature as overlying sediments build up. By the time a layer of peat 50 m thick is converted to bituminous coal, its thickness has been reduced by 90 percent. In the process, the proportion of carbon has increased from 60 to 80 percent.

Limestone and **dolostone,** containing the minerals calcite and dolomite, respectively, are the most important biogenic rocks. Limestone composed entirely of shelly debris is called *coquina.* Other limestones consist of cemented reef organisms *(reef limestone),* the compacted carbonate shells of minute floating organisms *(chalk),* and accumulations of tiny round, calcareous accretionary bodies *(ooliths)* that are 0.5 to 1 mm (0.02 to 0.04 in) in diameter *(politic limestone).*

Biogenic sediment can form both in the sea and on land. Plants such as trees, bushes, and grasses contribute most of the biogenic material on land. In water-saturated environments, such as bogs or swamps, plant remains accumulate to form **peat,** a sediment with a carbon content of about 60 percent. Peat is the initial stage in the development of the biogenic sedimentary rock we call **coal.**

As peat is buried beneath more plant matter and accumulating sand, silt, or clay, both temperature and pressure rise. As millions of years pass, the increased temperature and pressure bring about a series of changes. The peat is compressed, water is squeezed out, and the gaseous organic compounds such as methane (CH_4) escape, leaving an increased proportion of carbon. The peat is thereby converted into *lignite* and eventually into *coal* (Fig. 7.9).

METAMORPHISM: NEW ROCKS FROM OLD

Metamorphic rocks are of particular interest because the changes they undergo occur in the solid state. As tectonic plates move and crustal fragments collide, rocks are squeezed, stretched, bent, heated, and changed in complex ways. Even when a rock has been altered two or more times, however, vestiges of its earlier forms are usually preserved because the changes occurred in the solid state. Solids, unlike liquids and gases, tend to retain a "memory" of the events that changed them. In many ways, therefore, metamorphic rocks are the most complex but also the most interesting of the rock families. In them is preserved the story of all the collisions that have happened to the crust. Deciphering the record is an exceptional challenge for geologists. For example, when tectonic plates collide, distinctive kinds of metamorphic rocks form along the plate edges. Therefore, by studying the distribution of metamorphic rocks, it is possible to determine where the boundaries of ancient continents once were. Geologists also use evidence derived from metamorphic rocks to determine how long plate tectonics has been active on the Earth. So far the evidence suggests that plate tectonics has been operating for at least 2 billion years!

The Limits of Metamorphism

Metamorphism describes changes in mineral assemblage and texture in sedimentary and igneous rocks subjected to temperatures above 200°C (392°F) and pressure in excess of about 300 MPa (the pressure caused by a few thousand meters of overlying rock). There is, of course, an upper limit to metamorphism because at sufficiently high temperatures rock will melt. Remember, then: metamorphism refers only to changes in solid rock, not to changes caused by melting. Changes due to melting involve igneous phenomena, as discussed in Chapter 4.

Low-grade metamorphism refers to metamorphic processes that occur at temperatures from about 200°C (392°F) to 320°C (608°F) and at relatively low pressures (Fig. 7.10). High-grade metamorphism refers to metamorphic processes at high temperature (above about 550°C or 1022°F) and high pressure. Intermediate-grade metamorphism lies between low and high grade.

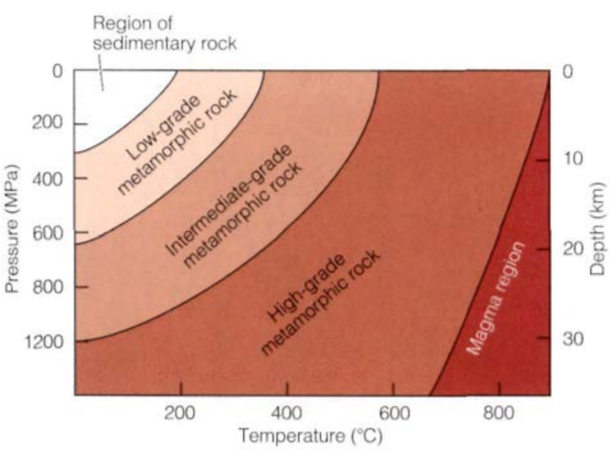


Figure 7.10 Ranges of temperature and pressure (equivalent to depth) under which metamorphism occurs in the crust. At lowest temperatures and pressure, sediment is converted to sedimentary rock. At the highest temperature and pressure, melting commences and the result is magma rather than rock.

Controlling Factors in Metamorphism

In a simplistic way, you can think of metamorphism as cooking. When you cook, what you get to eat depends on what you start with and on the cooking conditions. So too with rocks: the end product is controlled by the initial composition of the rock and by the metamorphic (or cooking) conditions. The chemical composition of a rock undergoing metamorphism plays a controlling role in the new mineral assemblage; so do changes in temperature and pressure. The ways in which temperature and pressure control metamorphism are not entirely straightforward, however, because they are strongly influenced by such factors as the presence or absence of fluids, how long a rock is subjected to high pressure or high temperature, and whether the rock is simply compressed or is twisted and broken as well as compressed during metamorphism.

Chemical Reactivity Induced by Fluids

The innumerable open spaces between grains in a sedimentary rock and the tiny fractures in many igneous rocks are called pores, and all pores are filled by a watery fluid. The fluid is never pure water; it always has dissolved in it small amounts of gases such as CO_2 and salts such as NaCl and $CaCl_2$, as well as traces of all the mineral constituents present in the enclosing rock; and at high temperature the fluid is more likely to be a vapor than a liquid. Regardless of its composition or phase, the *intergranular fluid* (for that is its best designation) plays a vital role in metamorphism.

When the temperature of, and pressure on, a rock undergoing metamorphism change, so does the composition of the intergranular fluid. Some of the dissolved constituents move from the fluid to the new minerals growing in the metamorphic rock. Other constituents move in the other direction, from the minerals to the fluid. In this way, the intergranular fluid serves as a transporting medium that speeds up chemical reactions in much the same manner that water in a stew pot speeds up the cooking of a tough piece of meat.

Pressure and Temperature

When a mixture of flour, salt, sugar, yeast, and water is baked, the high temperature causes a series of chemical reactions—new compounds are formed and the result is a loaf of bread. When rocks are heated, new minerals grow and the result is a metamorphic rock. In the case of the rocks, the cooking is brought about by the Earth's internal heat. Rocks can be heated by burial, by a nearby igneous intrusion, or by a thickening of the crust owing to collision. But burial, collision, and intrusion can also cause a change in pressure. Therefore, whatever the cause of the heating, metamorphism can rarely be considered to be entirely the result of a rise in temperature. The effects attributable to changing temperature and pressure must be considered together.

When discussing metamorphic rocks, scientists often use the term *stress* in place of pressure because stress has the connotation of direction. Rocks are solids, and solids can be squeezed more strongly in one direction than another; that is, stress in a solid, unlike stress in a liquid, can be different in different directions. The textures in many metamorphic rocks record **differential stress** (meaning not equal in all directions) during metamorphism. By contrast, most igneous rocks have textures formed under **uniform stress** (meaning equal in all directions) because igneous rocks crystallize from liquids.

The most visible effect of metamorphism in a differential stress field involves the texture of silicate minerals, such as micas and chlorites, that contain polymerized $(Si_4O_{10})^{4-}$ sheets. Compare Figure 7.11A

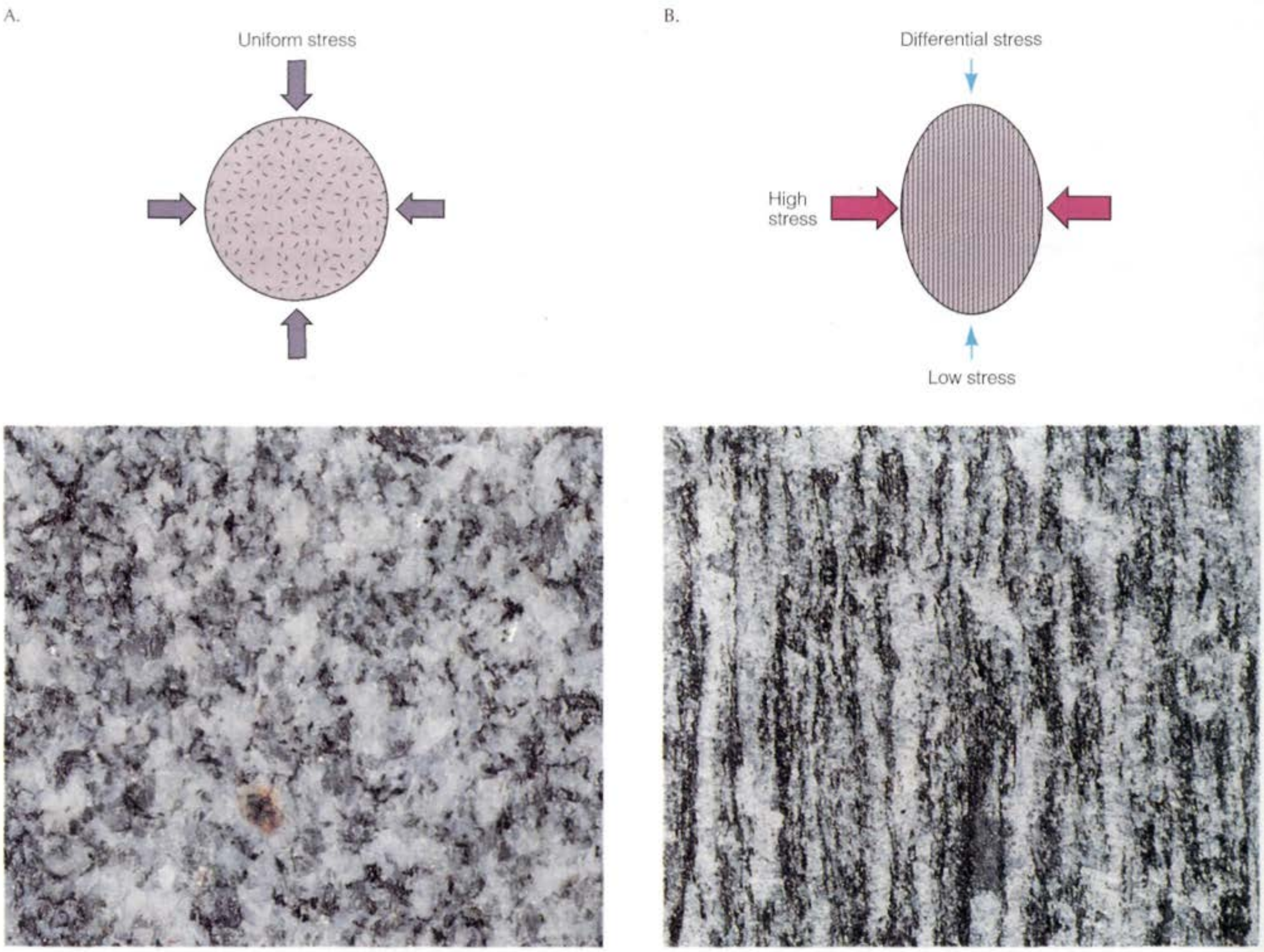


Figure 7.11 Comparison of textures developed in rocks of the same composition under uniform and differential stress. A. Granite, consisting of quartz, feldspar, and mica (the dark mineral) that crystallized under a uniform stress. Note that mica grains are randomly oriented. B. High-grade metamorphic rock, also consisting of quartz, feldspar, and mica, that crystallized under a differential stress. Mica grains are parallel, giving the rock a distinct foliation.

and B. Figure 7.11A is a granite that has a typical texture of randomly oriented mineral grains that grew in a uniform stress field. Figure 7.11B is a metamorphic rock containing the same minerals as those in A, but this rock formed in a differential stress field. Note that in Figure 7.11B all the biotite grains (black) are parallel, giving the rock a distinctive texture.

In a metamorphic rock containing sheet-structure minerals, the sheets are oriented perpendicular to the direction of maximum stress, as shown in Figure 7.11B. The parallel sheets produce a planar texture called **foliation,** named from the Latin word *folium,* meaning leaf. Foliated rock tends to split into thin flakes.

It is important to understand that stress can be high or low. It is the magnitude of the stress that determines a mineral assemblage. Texture, on the other hand, is controlled by differential versus uniform stress. To avoid confusion, geologists often use the term *stress* to discuss texture and *pressure* to discuss mineral assemblages and metamorphic grades.

Metamorphic Mineral Assemblages

Metamorphism produces new mineral assemblages as well as new textures. As temperature and pressure rise, one new mineral assemblage follows another. For any given rock composition, each assemblage is characteristic of a given range of temperature and pressure. A few of these minerals are found rarely (or not at all) in igneous and sedimentary rocks. Therefore, their presence in a rock is usually sufficient evidence that the rock has been metamorphosed. Examples of these metamorphic minerals are chlorite, serpentine, talc, and the three Al_2SiO_5 minerals andalusite, kyanite, and sillimanite. The way mineral assemblages change with grade of metamorphism as a shale is metamorphosed is illustrated in Figure 7.12.

Kinds of Metamorphism

The processes that result from changing temperature and pressure, and that cause the metamorphic changes observed in rocks, can be grouped under the terms *mechanical deformation* and *chemical recrystallization.* Mechanical deformation includes grinding, crushing, and the development of foliation (Fig. 7.13). Chemical recrystallization includes all the changes in mineral composition, in growth of new minerals, and the losses of H_2O and CO_2 that occur as rock is heated and pressurized. Different kinds of metamorphism reflect the different levels of importance of the two processes. The two most important kinds of metamorphism are burial and regional metamorphism.

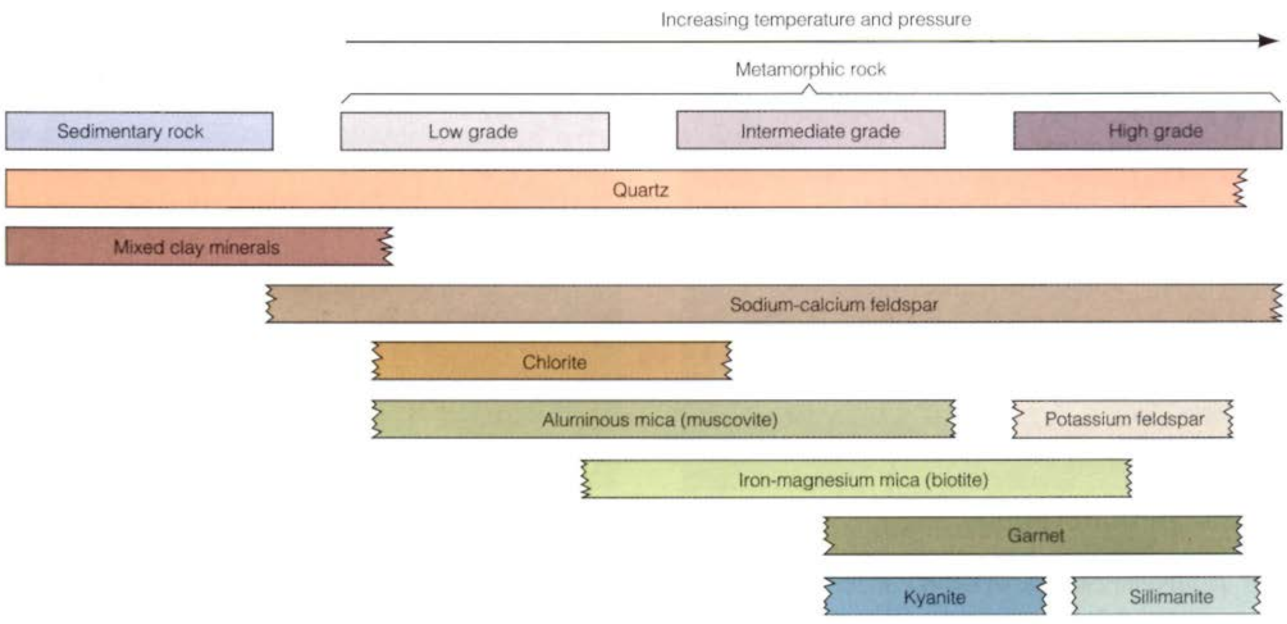


Figure 7.12 Changing mineral assemblages as a shale is metamorphosed from low to high grade. Kyanite and sillimanite have the same composition (Al_2SiO_3) but different crystal structures. They are found only in metamorphic rocks.

A.

B.

Figure 7.13 Development of foliation in a granite by mechanical deformation—the mineral assemblage is unchanged but the texture changes. From Groothoek, South Africa. A. Undeformed granite consisting of quartz, feldspar, and mica. The dark patch in the center of the field of view is a fragment of amphibolite (a metamorphic rock consisting largely of amphibole) that fell into the magma during intrusion. Foliation is not present. B. The original granitic texture has been completely changed, and the granite has been metamorphosed to a gneiss with a distinct foliation. The dark streak above the hammer handle was originally an inclusion of amphibolite like that in A. The amphibolite fragment has been crushed and stretched by the differential stress during the metamorphic processes.

Burial Metamorphism

Sediments, together with interlayered pyroclastics, may attain temperatures in excess of 200°C (392°F) or more when buried deeply in a sedimentary basin and thus subjected to metamorphism. Abundant pore water is present in buried sediment, and this water speeds up chemical recrystallization and helps new minerals to grow. Because water-filled sediment is weak and acts more like a liquid than a solid, however, the stress during **burial metamorphism** tends to be uniform. As a result, little mechanical deformation is involved in burial metamorphism, and the resulting metamorphic rock lacks foliation. The texture looks like that of an essentially unaltered sedimentary rock, even though the mineral assemblages in the two are completely different from one another.

Burial metamorphism is the first stage of metamorphism in deep sedimentary basins, such as deep-sea trenches on the margins of tectonic plates and off the mouths of great rivers. Burial metamorphism is known to be happening today in the great pile of sediments accumulated in the Gulf of Mexico, off the mouth of the Mississippi River.

Regional Metamorphism

The most common metamorphic rocks of the continental crust occur in areas of tens of thousands of square kilometers, and the process that forms them is called **regional metamorphism.** Unlike burial metamorphism, regional metamorphism involves differential stress and a considerable amount of mechanical deformation in addition to chemical recrystallization. As a result, regionally metamorphosed rocks tend to be distinctly foliated. For a discussion of the textures and mineral assemblages of regionally metamorphosed rocks, see "A Closer Look: Kinds of Metamorphic Rock."

Slate, phyllite, and **schist,** the low-, intermediate- and high-grade metamorphic rocks, respectively, produced from shale, are the most common varieties of regionally metamorphosed rocks. **Gneiss** is also a high-grade metamorphic rock produced from shale; it has more quartz and feldspar and less mica than is present in schist. Regionally metamorphosed rocks are usually found in mountain ranges formed as a result of either subduction or collision between fragments of continental crust. During both subduction and collision between continents, sedimentary rock along the margin of a continent is subjected to very intense differential stresses. The foliation that is so characteristic of slates, phyllites, schists, and gneisses is a consequence of those intense stresses. Regional metamorphism is therefore a consequence of plate tectonics.

When segments of ancient oceanic crust of basaltic composition are incorporated into the continental crust as a result of subduction, metamorphism produces two distinctive rocks. Low-grade metamorphism produces **greenschists,** so named because they are rich in chlorite, which is green. Intermediate-grade metamorphism produces **amphibolites,** which are rich in amphiboles.

A Closer Look

Kinds of Metamorphic Rock

Metamorphic rocks are named partly on the basis of texture and partly on mineral assemblage. The most widely used names are those applied to metamorphic derivatives of shales, sandstones, limestones, and basalts. This is so because the first three are the most abundant sedimentary rock types, while basalt is by far the most common igneous rock.

METAMORPHISM OF SHALE

Slate

The low-grade metamorphic product of shale is *slate* (Fig. C7.3). The minerals usually present in shale include quartz, clays, calcite, and possibly feldspar. Slate contains quartz, feldspar, and mica or chlorite. Although a slate may still look like a shale, the tiny mica and chlorite grains give slate a distinctive foliation called slaty cleavage. The presence of slaty cleavage is clear proof that a rock has gone from being a sedimentary rock (shale) to being a metamorphic rock.

Phyllite

Continued metamorphism of a slate to intermediate grade produces both larger grains of mica and a changing mineral assemblage; the rock develops a pronounced foliation (Fig. C7.3), and is called *phyllite* (from the Greek *phyllon,* meaning a leaf). In a slate it is not possible to see the new grains of mica with the unaided eye, but in a phyllite they are just large enough to be visible.

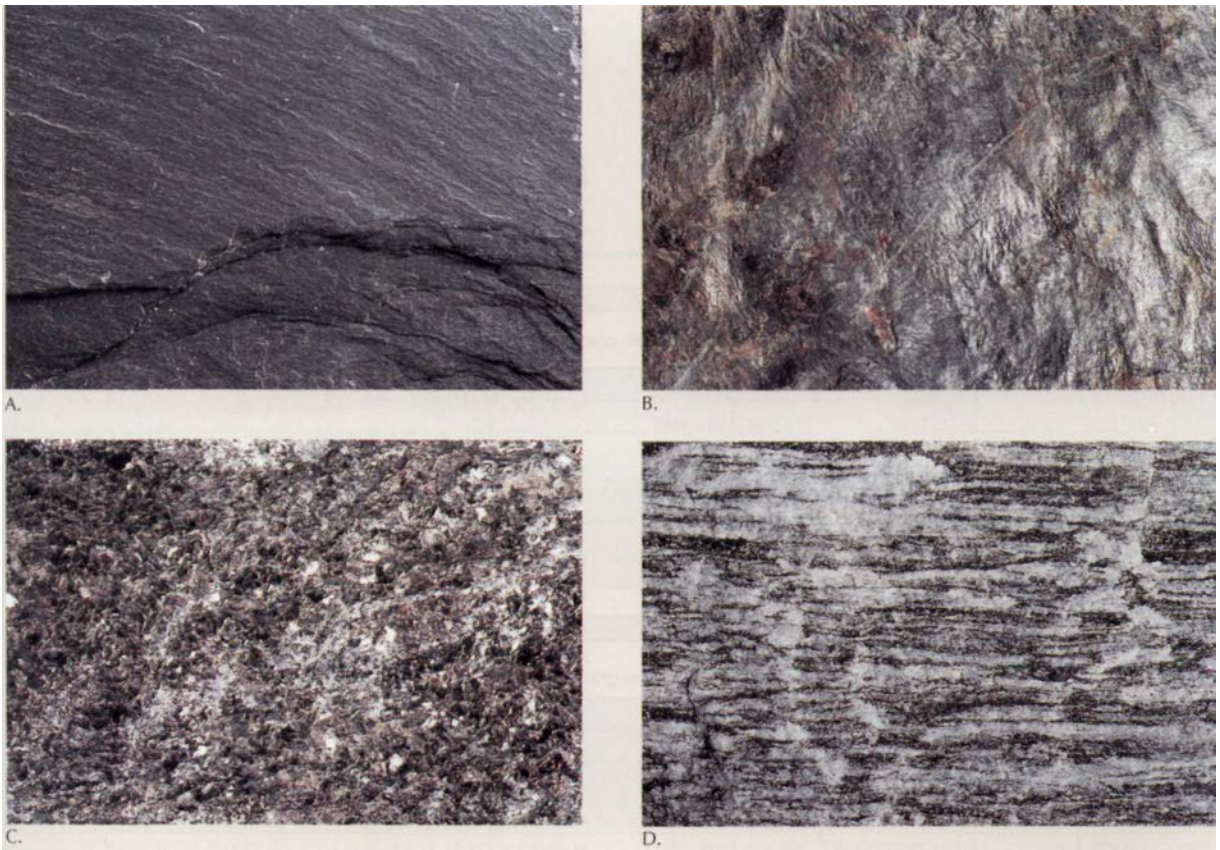


Figure C7.3 Progressive metamorphism of shale and the development of foliation. A. Slate from Bangor, Pennsylvania. Individual mineral grains are too small to be visible. Slaty cleavage records the beginning of metamorphism. B. Phyllite from Woodbridge, Connecticut. Mineral grains are just visible. Foliation is more pronounced. C. Schist, from Manhattan, New York. Mineral grains are now easily visible and foliation is pronounced. D. Gneiss, from Uxbridge, Massachusetts. Quartz and feldspar layers (light) are segregated from mica-rich layers (dark). Foliation is pronounced.

Schist and Gneiss

Still further metamorphism beyond that which produces a phyllite leads to a coarse-grained rock with pronounced schistosity, called *schist* (Fig. C7.3). The most obvious differences between slate, phyllite, and schist are in grain size (Fig. C7.4A). At the high grades of metamorphism characteristic of schists, minerals may start to segregate into separate bands. A high-grade rock with coarse grains and pronounced foliation, but with layers of micaceous minerals segregated from layers of minerals such as quartz and feldspar, is called a *gneiss* (pronounced nice, from the German *gneisto,* meaning to sparkle) (Fig. C7.3).

METAMORPHISM OF BASALT

Greenschist

The main minerals in basalt are olivine, pyroxene, and feldspar, each of which is anhydrous. When a basalt is subjected to metamorphism under conditions where H_2O can enter the rock and form hydrous minerals, distinctive mineral assemblages develop (Fig. C7.4B). At low grades of metamorphism, mineral assemblages such as chlorite + feldspar + epidote + calcite form. The resulting rock is equivalent in metamorphic grade to a slate but has a very different appearance. It has pronounced foliation as a phyllite does, but it also has a very distinctive green color because of its chlorite content: it is termed *greenschist.*

Amphibolite and Granulite

When a greenschist is subjected to an intermediate grade of metamorphism, chlorite is replaced by amphibole; the resulting rock is generally coarse grained and is called an *amphibolite.* Foliation is present in amphibolites but is not pronounced because micas and chlorite are usually absent. At highest grade metamorphism, amphibole is replaced by pyroxene, and an indistinctly foliated rock called a *granulite* develops.

METAMORPHISM OF LIMESTONE AND SANDSTONE

Marble and **quartzite** are the metamorphic derivatives of limestone and sandstone, respectively. Neither limestone nor quartz sandstone (when pure) contains the necessary ingredients to form sheet- or chain-structure minerals. As a result, marble and quartzite commonly lack foliation.

Marble

Marble consists of a coarsely crystalline, interlocking network of calcite grains. During recrystallization of a limestone, the bedding planes, fossils, and other features of sedimentary rocks are largely obliterated. The end result, as shown in Figure C7.5A, is an even-grained rock with a distinctive, somewhat sugary texture. Pure marble is snow white and consists entirely of pure grains of calcite. Such marbles are favored for marble gravestones

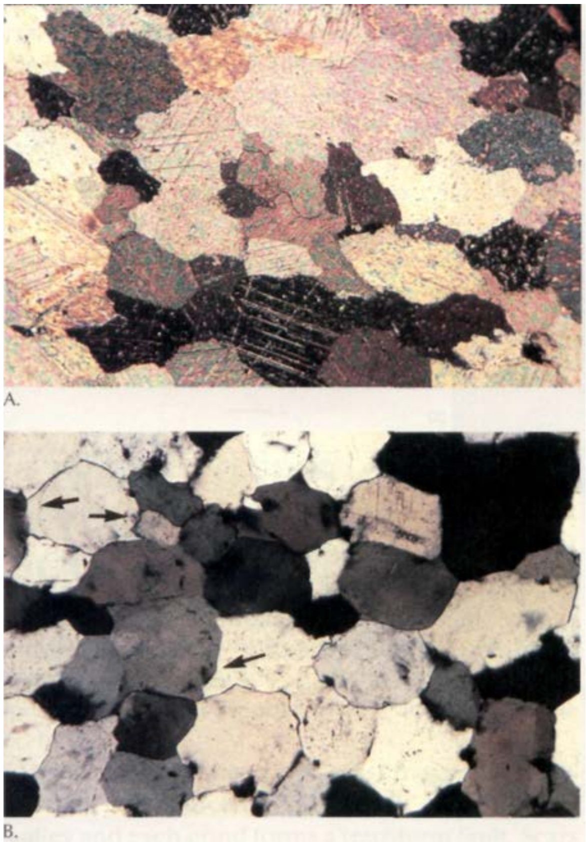
A.

	Intensity of metamorphism →			
	Not metamorphosed	Low grade	Intermediate grade	High grade
Rock name	Shale →	Slate →	Phyllite →	Schist / Gneiss
Foliation	None	Subtle	Distinct; schistosity apparent	Conspicuous; schistosity and compositional layering
Size of mica grains	Microscopic	Microscopic	Just visible with hand lens	Large and obvious

B.

	Intensity of metamorphism →			
	Not metamorphosed	Low grade	Intermediate grade	High grade
Rock name	Basalt $\xrightarrow{+H_2O}$	Greenschist →	Amphibolite →	Granulite
Foliation	None	Distinct	Indistinct;, when present due to parallel grains of amphibole	Indistinct because of absence of micas
Size of mica grains	Visible with hand lens	Visible with hand lens	Obvious if present	Not present

Figure C7.4 Progressive metamorphism of shale and basalt. Both foliation and mica grain size change as a result of increasing temperature and differential stress.



and statues in cemeteries, perhaps because white is considered to be a symbol of purity. Many marbles contain impurities such as organic matter, pyrite, limonite, and small quantities of silicate minerals that impart various colors.

Quartzite

Quartzite is derived from sandstone by the filling in of the spaces between the original grains with silica and by recrystallization of the entire mass (Fig. C7.5B). Sometimes, the ghostlike outlines of the original sedimentary grains can still be seen, even though recrystallization may have completely rearranged the original grain structure.

Figure C7.5 Texture of nonfoliated metamorphic rocks seen in thin section and viewed in polarized light. Notice the interlocking grain structure produced by recrystallization during metamorphism. Each specimen is 2 cm across. A. Marble, composed entirely of calcite. All vestiges of sedimentary structure have disappeared. B. Quartzite. Arrows point to faint traces of the original rounded quartz grains in some of the grains.

Metamorphic Facies

A famous Finnish geologist, Pennti Eskola, pointed out in 1915 that the same metamorphic mineral assemblages are observed again and again. This led him to propose a concept known as **metamorphic facies.** (The term *facies* comes from the Latin for face, or appearance.) According to this concept, for a given range of temperature and pressure, and for a given rock composition, the assemblage of minerals formed during metamorphism is always the same. Based on mineral assemblages, Eskola defined a series of pressure and temperature ranges that he called metamorphic facies.

To help you understand Eskola's idea, another analogy with cooking is appropriate; think of a large roast of beef. When it is carved, you see that the center is rare, the outside is well done, and in between is a region of medium rare meat. The differences occur because the temperature was not uniform throughout. The center, rare-meat facies is a low-temperature facies; the outside, well-done facies is a high-temperature one. The composition of beef varies little, if at all, from roast to roast, and so the facies must depend not on composition but on the temperature. So too with rocks, although in any rock of given composition pressure as well as temperature determines mineral assemblage.

Figure 7.14 illustrates the principal metamorphic facies, together with geothermal gradients to be expected under three different geological conditions and therefore the conditions typical of each facies.

Plate Tectonics and Metamorphism

One of the triumphs of plate tectonics is that it provides, for the first time, an explanation for the distribution of metamorphic facies in regionally metamorphosed rocks. To repeat what we said above, regional metamorphism occurs at a convergent plate boundary, as shown in Figure 7.15.

The temperatures and pressures characteristic of *blueschist* facies in regional metamorphism are reached when crustal rocks are dragged down by a rapidly subducting plate. Under such conditions, pressure increases more rapidly than temperature,

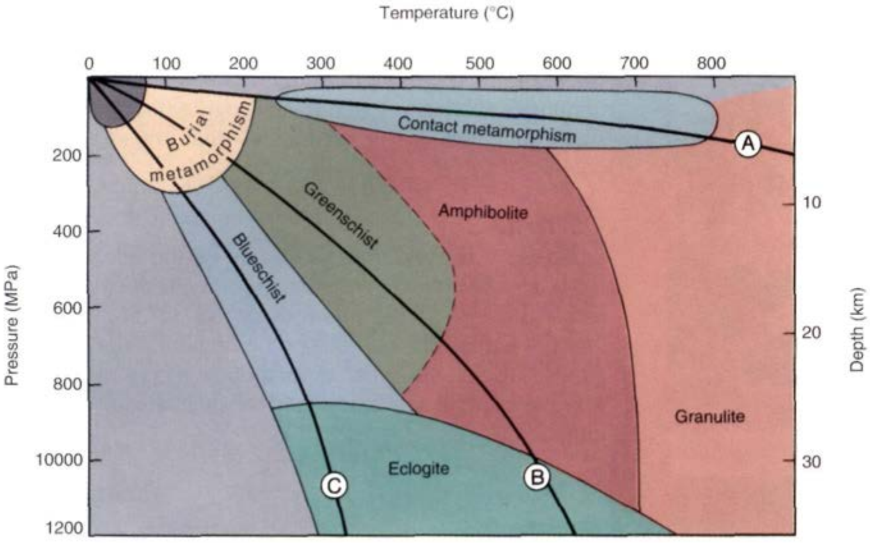


Figure 7.14 Metamorphic facies plotted with respect to temperature and pressure. Curve A is a typical thermal gradient around an intrusive igneous rock that causes low-pressure metamorphism. Curve B is a normal continental geothcrmal gradient. Curve C is the geothermal gradient developed in a subduction zone.

and as a result the rock is subjected to high pressure and relatively low temperature. Rocks subjected to blueschist facies metamorphism are widespread in the Coast Ranges of California. Blueschist metamorphism is probably happening today along the subducting margin where the Pacific plate plunges under the coast of Alaska and the Aleutian Islands,

The metamorphic conditions characteristic of *greenschist* and *amphibolite facies* metamorphism occur where crust is either thickened by continental

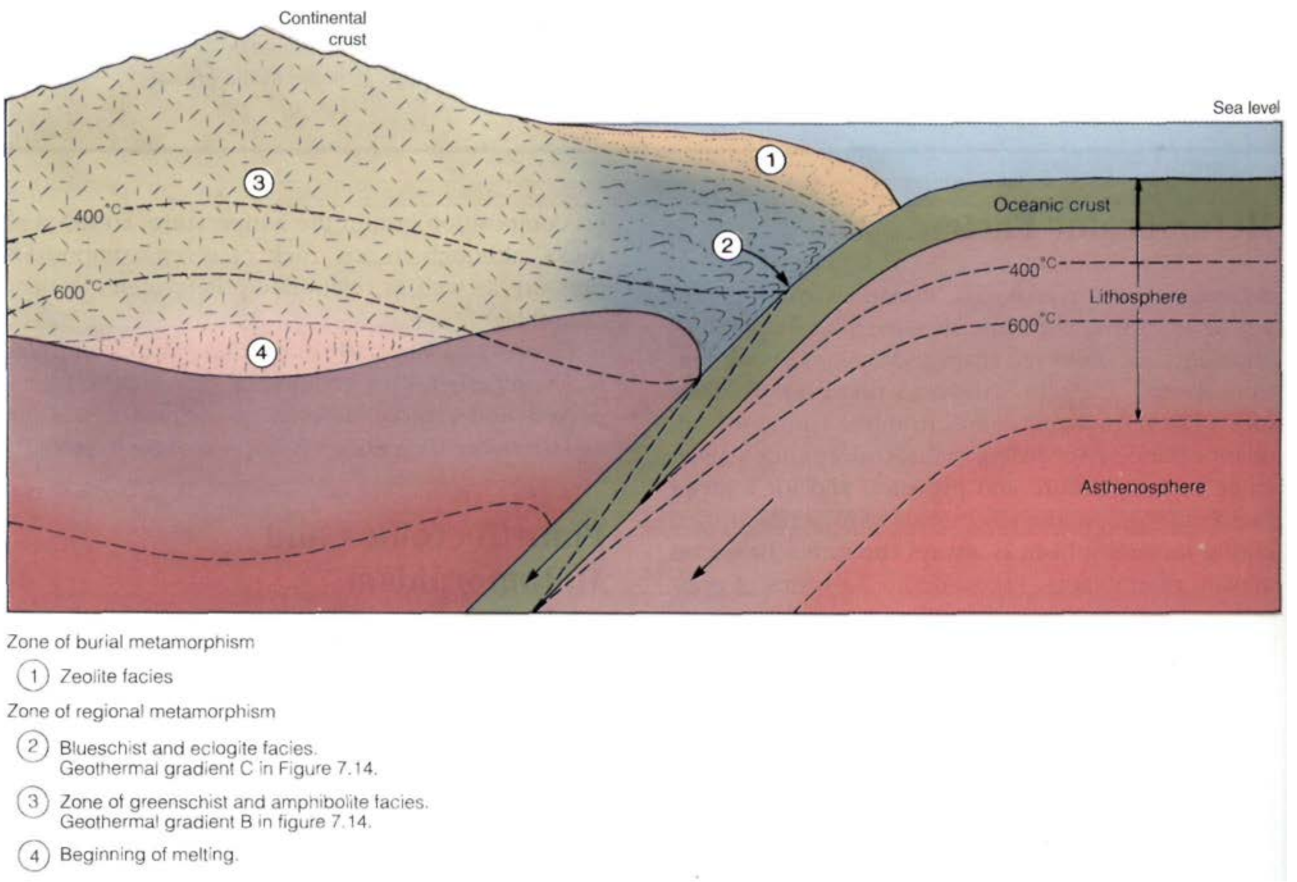


Figure 7.15 Diagram of a convergent plate boundary, showing the different regions of metamorphism. Dashed lines indicate temperature contours.

collision or heated by rising magma. Continental collision is the most common setting for regional metamorphism, and rocks formed in this way are observed throughout the Appalachians and the Alps. Such metamorphism is no doubt occurring today beneath the Himalaya, where the continental crust is thickened by collision, and beneath the Andes, where it is both thickened by subduction and heated by rising magma. If the crust is sufficiently thick, rocks subjected to amphibolite facies or higher grade metamorphism can reach temperatures at which partial melting commences, and metamorphism passes into magma formation.

PLATE TECTONICS, CONTINENTAL CRUST, AND MOUNTAIN BUILDING

Continental crust is simply a passenger being rafted on large plates of lithosphere. It is, however, a passenger that is buffeted, stretched, fractured, and altered by the ride.

Each collision between two crust fragments forms a mountain belt of metamorphic rocks, each stretch a rift valley and each grind forms a transform fault. Scars left in the continental crust by tectonic movements are evidence of former plate motions. That this continental evidence exists is fortunate because the most ancient known crust in the ocean is only about 180 million years old. Thus, the only direct evidence concerning geological events more ancient than 180 million years comes from the continental crust.

Before beginning our discussion on how the continental crust has grown and been shaped by plate tectonics, it is helpful to look first at the large-scale structure of continents.

Regional Structures of Continents

On the scale of a continent, two kinds of structural units can be distinguished in the continental crust: cratons and orogens. A **craton** is a core of very ancient rock (Fig. 7.16). The term is applied to those ancient parts of the Earth's crust that have attained isostatic stability. Rocks within cratons may be deformed, but the deformation is invariably ancient, and how ancient cratons formed is still a matter of intense research.

Draped around cratons are the second kind of crustal building unit, **orogens,** which are elongate regions of crust that have been intensely bent and fractured during continental collisions. Crust in an orogen is commonly thicker than crust in a craton, and many orogens—even some very old ones—have not yet attained isostatic equilibrium. Orogens are the eroded roots of ancient mountain ranges that formed as a result of collisions between cratons. Orogens differ from each other in age, history, size, and details of origin; however, all were once mountainous terrains, and all are younger than the cratons they surround. Only the youngest orogens are mountainous today. Ancient orogens, now deeply eroded, reveal their history through the kinds of metamorphic rock they contain and the way the rocks are twisted and deformed.

An assemblage of cratons and ancient orogens is called a **continental shield.** North America has a huge continental shield at its core, and around the shield are five young orogens: the Grenville, Appalachian, Caledonide, Innuitian, and Cordilleran orogens (Fig. 7.16). Because the North American shield crops out in Canada (especially Ontario and Quebec), but is mostly covered by younger, flat-lying sedimentary rocks in the United States, geologists often refer to it as the Canadian Shield. That portion of a continental shield that is covered by a thin layer of younger sedimentary rocks is called a *stable platform.*

Through careful mapping, geologists have identified several ancient cratons and orogens in the Canadian Shield. All the craton rocks are older than 2.5 billion years. Such rocks can be observed in many places in eastern Canada, but in the United States they crop out only in a small region around Lake Superior. By drilling through the stable platform that covers most of the U.S. portion of the shield, geologists have discovered that cratons and ancient orogens similar to those that surface in eastern Canada lie below much of the central United States and part of western Canada.

The three small cratons, Slave, Wyoming, and Kaminak, and the three larger cratons, Churchill, Superior, and North Atlantic, shown in Figure 7.16, were once minicontinents. By about 1.6 billion years ago, these minicontinents had become welded together to form the assemblage of cratons and ancient orogens (in other words, the Canadian Shield) we see in North America today. Collisions between the three small cratons and the large Churchill craton did not form large orogens, but each time two of the larger cratons collided, an orogen was formed between them. For example, when the Superior and Churchill cratons collided, the Trans-Hudson orogen formed. The existence of ancient collision belts—orogens—between the cratons of the Canadian Shield is the best evidence available to support the idea that plate tec-

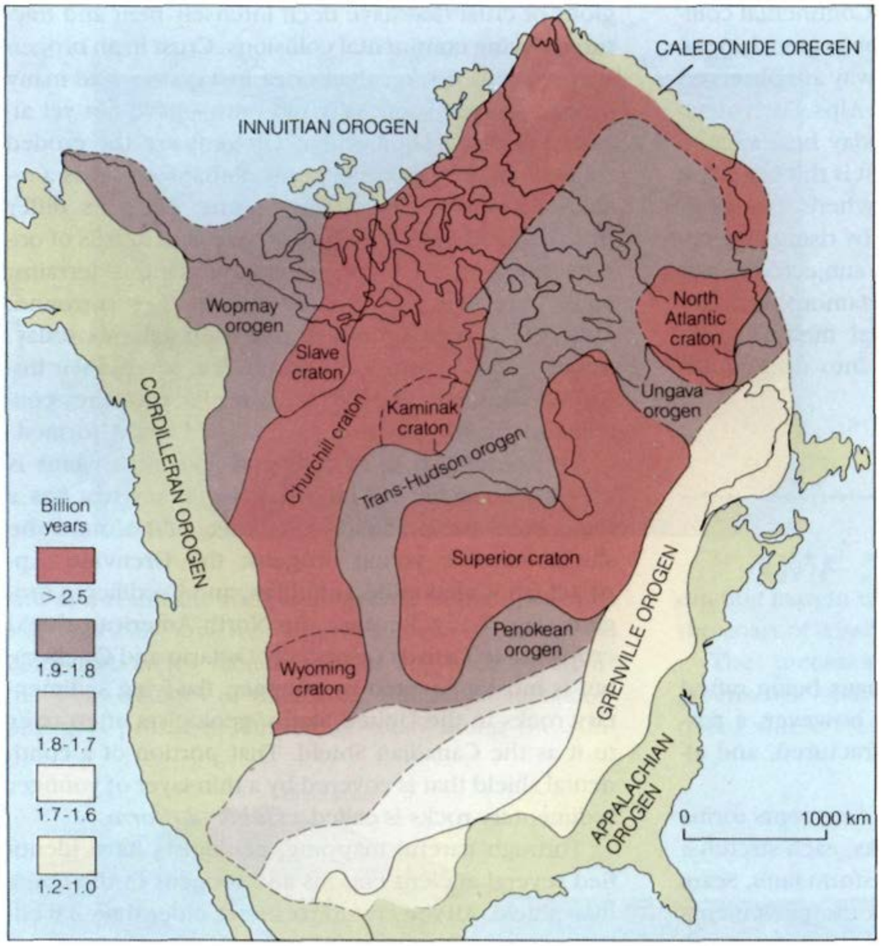


Figure 7.16 The North American cratons and associated orogens. The Grenville orogen is about 1 billion years old, while the Caledonide, Appalachian, Cordilleran, and Innuitian orogens are each younger than 600 million years. The assemblage of cratons and orogens, all older than 1.8 billion years, which are surrounded by the five young orogens, is the Canadian Shield.

tonics operated at least as far back as 1.8 billion years ago. This means that the solid Earth portion of the Earth system must have been working very much as it works today, for *at least* 1.8 billion years.

Continental Margins

The fragmentation, drift, and welding together of pieces of continental crust are direct consequences of plate tectonics. Various combinations of these processes are responsible for the five types of continental margins we know of today: passive, convergent, collision, transform fault, and accreted terrane. Before we discuss additional evidence for plate tectonics, it will be helpful to review briefly the features associated with each of the continental margins.

Passive Continental Margins

A passive continental margin is one that occurs in the stable interior of a plate. The Atlantic Ocean margins of the Americas, Africa, and Europe are passive. The eastern coast of North America, for example, is in the stable interior of the North American Plate, far from the plate margins. Passive continental margins develop when a new ocean basin forms by the rifting of continental crust, as illustrated in Figure 6.12. This process can be seen happening today in the Red Sea, which is a young ocean with an active spreading center running down its axis (Fig. 7.17). New, passive continental margins have formed along both edges of the Red Sea.

Passive continental margins are places where great thicknesses of sediment accumulate. The kinds of sediment deposited are distinctive, and the Red Sea provides an example. Deposition commenced with clastic, nonmarine sediments, followed by chemical sediments (rock salt) and then clastic marine shales. The sequence apparently arises in the following manner. Basaltic magma, associated with the formation of the new spreading center that splits the continent, heats and expands the lithosphere so that a plateau forms with an elevation of as much as 2.5 km (1.6 mi) above sea level (Fig. 6.12). Tensional stress breaks the crust along the plateau and forms a rift, with the result that there is a pronounced topographic relief between the plateau and the floor of the rift. The earliest rifting of what is now the Red Sea must have looked

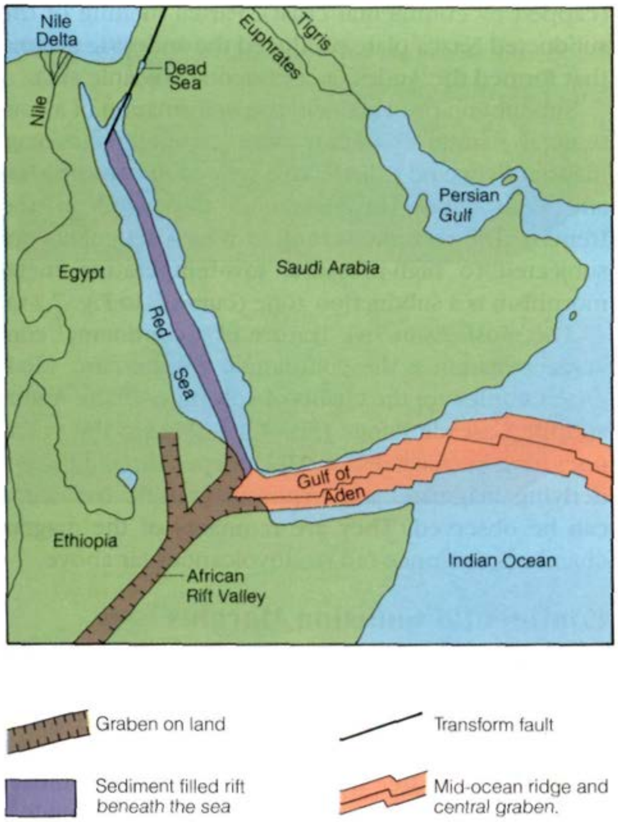


Figure 7.17 Three spreading centers meet at a triple junction. Two, the Gulf of Aden and the Red Sea, are actively spreading, and there are passive continental margins along the adjacent coastlines. The African Rift appears to be a failing rift that will not develop into an open ocean.

very much the way the African Rift Valley looks today. Before the rift floor sank low enough for seawater to enter, clastic nonmarine sediments, such as conglomerates and sandstones, were shed from the steep valley walls and accumulated in the rift. Associated with these sediments are basaltic lavas, dikes, and sills, all formed by magma rising up the fractures. As the rift widened, a point was reached where seawater entered. The early flow was apparently restricted, and the water was shallow, resembling a shallow lake more than an ocean. The rate of evaporation would have been high, and as a result strata of rock salt were laid down on top of the clastic nonmarine sediments. Finally, as rifting continued and the depth of the seawater increased, normal clastic marine sediments were deposited. This is the stage the Red Sea is in today. Eventually, as further rifting exposes new oceanic crust, the Red Sea will evolve into a younger version of the Atlantic Ocean.

Notice in Figure 7.17 that the Gulf of Aden, the Red Sea, and the northern end of the African Rift Valley meet at angles of 120°. Such a meeting point formed by three spreading centers is called a *plate triple junction.* Two of the centers, the Gulf of Aden and the Red Sea, are active and still spreading. The third, the African Rift Valley, is apparently no longer spreading and possibly will not evolve into an ocean. What will remain on the African continent is a long, narrow rift filled primarily with nonmarine sediment. The formation of triple junctions with one arm not developing into an ocean is apparently a characteristic feature of passive continental margins. This can be seen from Figure 7.18, which shows the reassembled positions

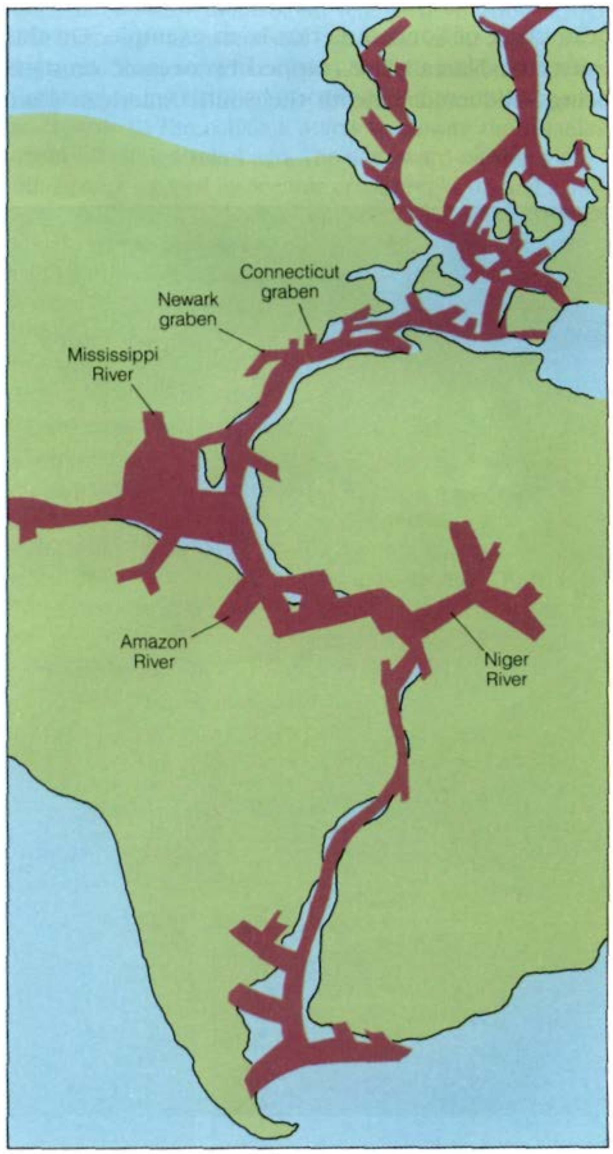


Figure 7.18 Map of a closed Atlantic Ocean showing the rifts that formed when Pangaea was split by a spreading center. The rifts on today's continents are now filled with sediment. Some of them serve as the channelways for large rivers.

of the continents that today flank the Atlantic Ocean. Note that some of the world's largest rivers, such as the Niger, the Amazon, and the Mississippi, flow down valleys formed by undeveloped rifts associated with the opening of the Atlantic Ocean. This is a compelling example of one of the ways the solid Earth portion of the Earth system influences the hydrosphere.

Continental Convergent Margins

A continental convergent margin is one where the edge of a continent coincides with a convergent plate margin along which oceanic lithosphere is being subducted beneath the continental lithosphere. The Andean coast of South America is an example. On this coast, the Nazca plate (capped by oceanic crust) is being subducted beneath the South American plate (capped by continental crust). Partial melting of the subducted Nazca plate produced the andesitic magma that formed the Andes (a continental volcanic arc).

Subduction produces intense deformation of a continental margin (together with characteristic magmatic activity and a distinctive style of metamorphism and deformation in sediments deposited in the trench). The tectonic setting in which sediments are subjected to high-pressure, low-temperature metamorphism is a subduction zone (curve C in Fig. 7.14).

The most distinctive feature of a continental convergent margin is the continental volcanic arc. Modern examples are the chains of volcanoes in the Andes and the Cascade Range (Fig. 7.19). Where the volcanoes have been eroded and the deeper parts of the underlying magmatic arc exposed, granitic batholiths can be observed. They are remnants of the magma chambers that once fed stratovolcanoes far above.

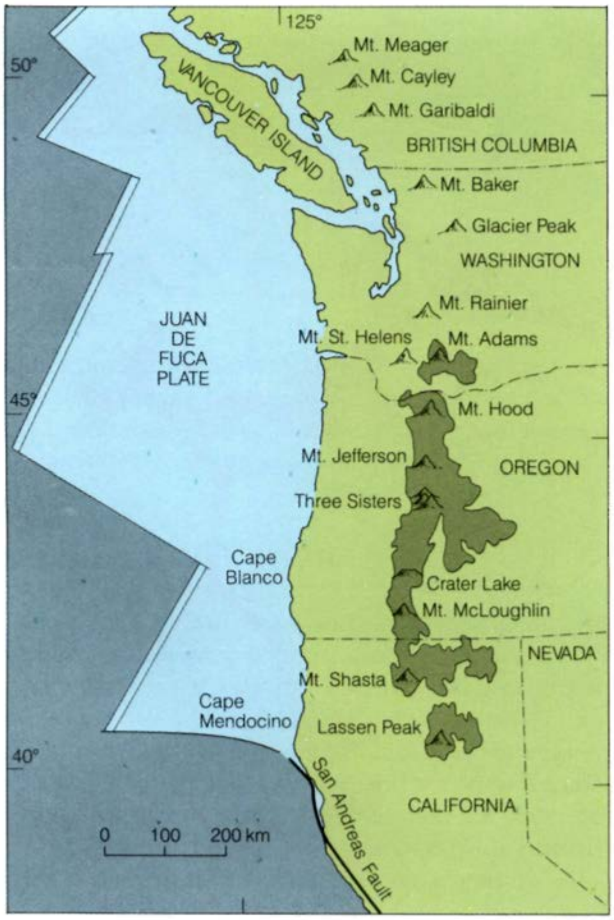


Figure 7.19 Volcanoes of the Cascade Range, a continental volcanic arc above a subduction zone. Each volcano has been active during the last 2 million years. Magma to form the volcanoes comes from partial melting of oceanic crust on the Juan de Fuca Plate as it is subducted beneath the North American Plate.

Continental Collision Margins

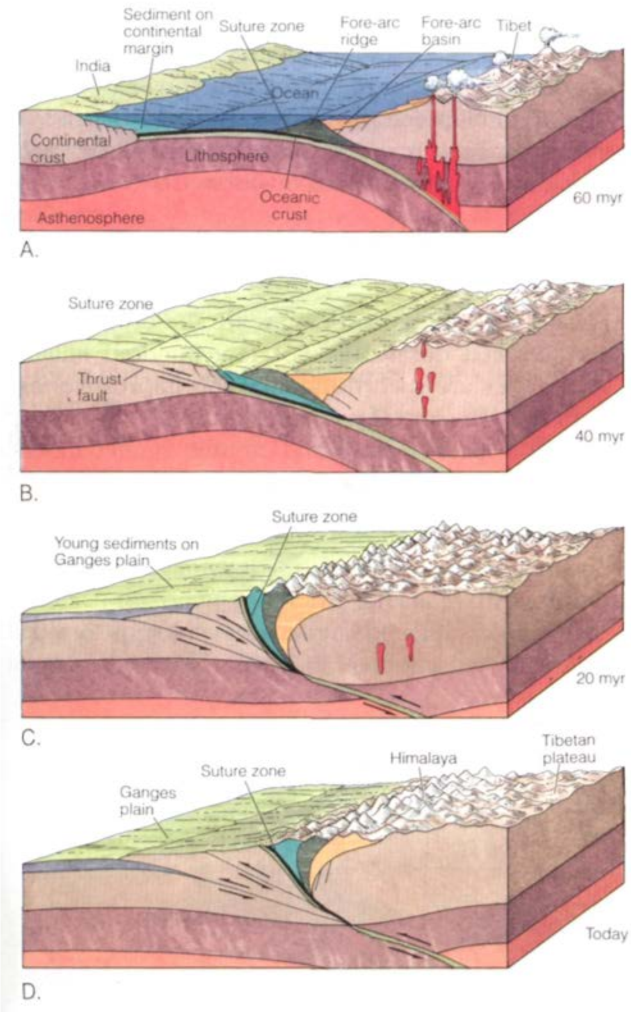
A continental collision margin is one where the edges of two continents, each on a different plate, come into collision (Fig. 7.20). A modern example is the line of collision between the Australian-Indian plate and the Eurasian plate. India, on the Australian-Indian plate, and Asia, on the Eurasian plate, have collided, and the Alpine-Himalayan mountain chain is the result.

When continental crust is carried on a plate that is being subducted beneath a continental convergent margin, the two continental fragments must eventually collide. The collision sweeps up and deforms any sediment that accumulated along the margins of both continents and forms a mountain system characterized by metamorphism of the sediment, together with intense fracturing and folding of strata.

All modern continental collision margins are young orogens characterized by soaring mountain systems. Occurring in great arc-shaped systems a few hundred kilometers wide, these mountain systems commonly reach several thousand kilometers in length. Within a system, strata are compressed, fractured, folded, and crumpled, commonly in an exceedingly complex manner. Metamorphism and igneous activity are always present. Examples are widespread: the Alps, the Himalaya, and the Carpathians are all young mountain systems still being actively uplifted, while older systems that are slowly being eroded down include the Appalachians and the Urals.

Transform Fault Margins

A transform fault continental margin occurs when the margin of a continent coincides with a transform fault boundary of a plate. The most striking example of a modern transform fault continental boundary is the

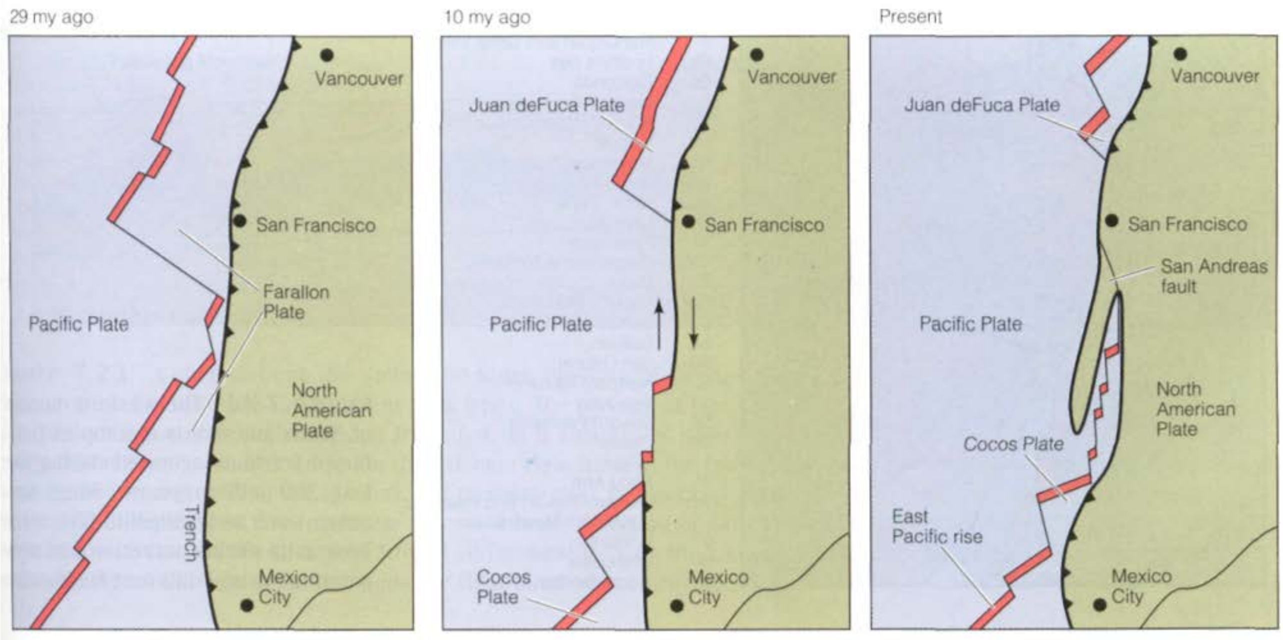


western margin of North America, from the Gulf of California to San Francisco, where it is bounded by the San Andreas Fault.

The San Andreas Fault apparently arose when the westward-moving North American continent overrode part of the spreading center (the East Pacific Rise), as shown in Figure 7.21. The San Andreas is the transform fault that connects the two remaining segments of the old spreading center.

Figure 7.20 Collision between two fragments of continental crust shown schematically for the collision between India and Tibet. A. India, on the left, moves north. Sixty million years ago an ocean still separated India and Tibet. B. India and Tibet start to collide about 40 million years ago. Sediment is buckled and fractured, and the lithosphere is thickened. C. The collision starts to elevate the Himalaya about 20 million years ago. The downward-moving plate of lithosphere capped by oceanic crust breaks off and continues to sink. D. The edge of the remaining segment of the plate on which India sits, and which is capped by buoyant continental crust, is partly thrust under the edge of the overriding plate on which Tibet sits, causing further elevation of the collision zone. The process is continuing and the Himalaya are still rising.

Figure 7.21 Origin of the San Andreas Fault. Twenty-nine million years ago, the edge of North America overrode a portion of the spreading center separating the Pacific Plate from the Farallon Plate, creating two smaller plates in the process, the Cocos Plate and the Juan de Fuca Plate. The San Andreas Fault is the transform fault that connects the remaining pieces of the severed spreading ridge.



Accreted Terrane Margins

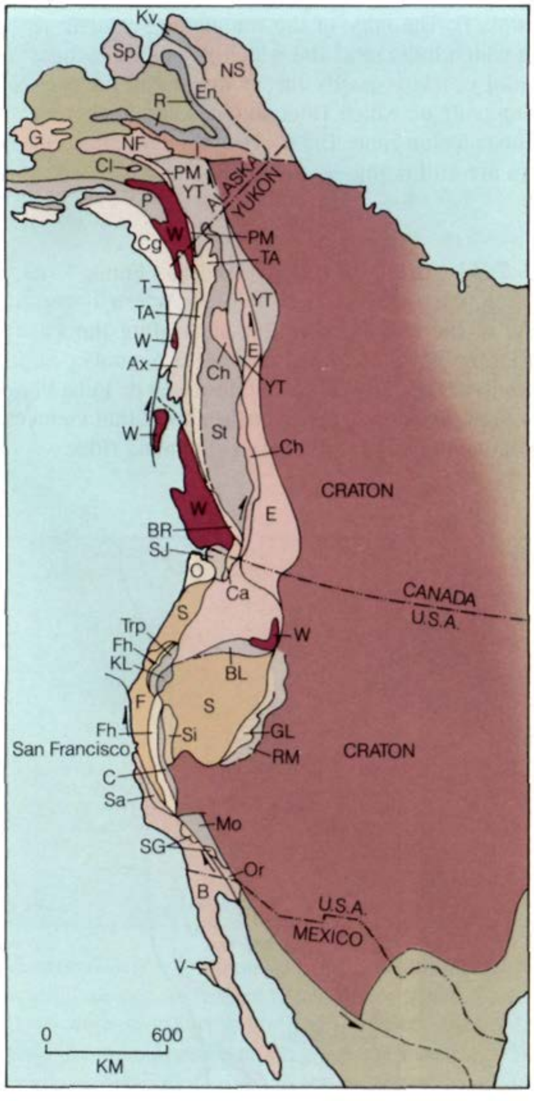
An **accreted terrane** continental margin is a former continental convergent margin or continental transform fault margin that has been further modified by the addition of rafted-in, exotic fragments of crust. They are the most complex of the five kinds of continental margin. The western margin of North America from central California to Alaska is an example (Fig. 7.22).

Plate motion can raft fragments of crust tremendous distances. Eventually, any fragment that has not been consumed by subduction is added (accreted) to a larger continental mass. Some of the fragments form when they are sliced off the margin of a large continent by a transform fault, much as the San Andreas Fault is slicing a fragment off North America today. Other combinations of volcanism, rifting, fracturing, and subduction can also form fragments of crust that are too buoyant to be subducted. In the western Pacific Ocean, there are many such small fragments of continental crust; examples include the island of Taiwan, the Philippine islands, and the many islands of Indonesia. Each fragment, called a *terrane*, is a geological entity characterized by distinctive rocks.

Mountain Building

Today's great mountain ranges are the orogens that have formed during the last few hundred million years. They are such distinctive and impressive features that we close this chapter by briefly describing one of the most beautiful and carefully studied mountain systems, the Appalachians.

The Appalachians are a mountain system 2500 km



Symbol	*Name*
Ax	Alexander
B	Baja
BL	Blue Mountains
BR	Bridge River
C	Calaveras
Cg	Chugach
Ch	Cache Creek
Cl	Chulitna
E	Eastern assemblages
En	Endicott
F	Franciscan and Great Valley
Fh	Foothills belt
GL	Golconda
I	Innoko
KL	Klamath Mountains
Kv	Kagvik
Mo	Mohave
NF	Nixon Fork
NS	North Slope
O	Olympic
P	Peninsular
PM	Pingston and McKinley
R	Ruby
RM	Roberts Mountains
S	Siletzia
Sa	Salinian
SG	San Gabriel
Si	Northern Sierra
SJ	San Juan
Sp	Seward Peninsula
St	Stikine
T	Taku
TA	Tracy Arm
Trp	Western Triassic and Paleozoic of Klamath Mountains
V	Vizcaino
W	Wrangellia
YT	Yukon-Tanana

Figure 7.22 The western margin of North America is a complex jumble ol terranes accreted during the last 200 million years. Some terranes, such as Wrangellia (W), were broken up during accretion and now occur in several different fragments.

(1554 mi) long that borders the eastern and southeastern coasts of North America and continues offshore, as eroded remnants, beneath the sediment of the modern continental shelf. The sedimentary strata in the system contain mud cracks, ripple marks, fossils of shallow-water organisms, and, in places, freshwater materials such as coal. Evidence is strong that sediment was deposited on the continental shelf of an old passive continental margin. The sedimentary strata, which thicken from west to east, are underlain by a basement of metamorphic and igneous rocks (Fig. 7.23).

Most but not all of the old strata of the Appalachians have now been deformed. Today, if we approach the central Appalachians from the west, we first see the former sediment occurring as essentially flat-lying, undisturbed strata. These strata were too far from the line of collision to suffer any deformation. Continuing eastward, we notice that the same strata thicken and we reach the point where the effects of the collision become apparent—the strata become gently bent into wavelike folds. Many of Pennsylvania's oil pools are found in these gently folded strata.

Proceeding east, we see the folds becoming less gentle and the development of gently inclined fractures until, finally, we reach the core of the Appalachians. Here the ancient basement rocks and the deep-water sediments deposited long ago on the old continental rise have been pushed upward and can be examined. The strata are increasingly metamorphosed, and deformation becomes more intense as the line of collision is reached.

The Appalachians have a complex history that started more than 600 million years ago, as demonstrated in Figure 7.24. Notice that three collision events are postulated and that an ancient ocean disappeared as a result of the collision about 350 million years ago. Today the old mountain system is being slowly eroded away, and the sediment is being deposited along the passive continental margin of eastern North America.

In Figure 7.24A you can see that 600 million years ago a passive continental margin bounded proto-North America. Eventually, 400 to 500 million years

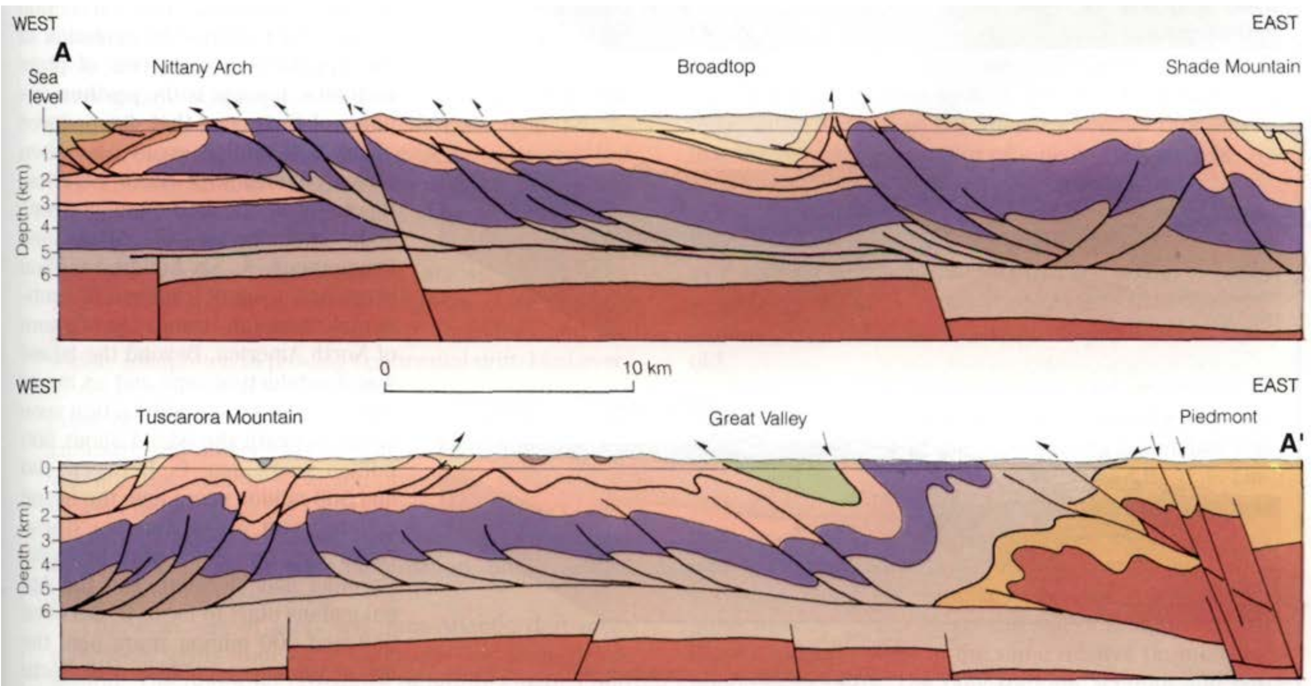


Figure 7.23 A slice through the Valley and Ridge Province of the Appalachians in Pennsylvania. Colors represent different rock types. The prominent purple unit, originally flat-lying but now contorted and fractured, is a stratum of limestone approximately 500 million years old. Rocks on the extreme right hand of the lower half of the diagram (brown and yellow) are igneous and metamorphic. Heavy black lines, including those that are curved, are fractures. Arrows indicate direction of movement along the fractures. Note that the slice runs from *A* on the west to *A'* on the east, so that the left-hand edge of the bottom section joins the right-hand edge of the top section.

ago, the passive margin became a convergent margin as a result of subduction starting. It is now thought that all passive margins become continental convergent margins when old oceanic lithosphere fractures close to the join between oceanic crust and continental crust and subduction commences. If this hypothesis is correct, at some unknown time in the future a new subduction zone will form along the Atlantic margin of North America, the Atlantic will slowly start to close, and eventually a new mountain system will form when Africa and Europe collide with North America.

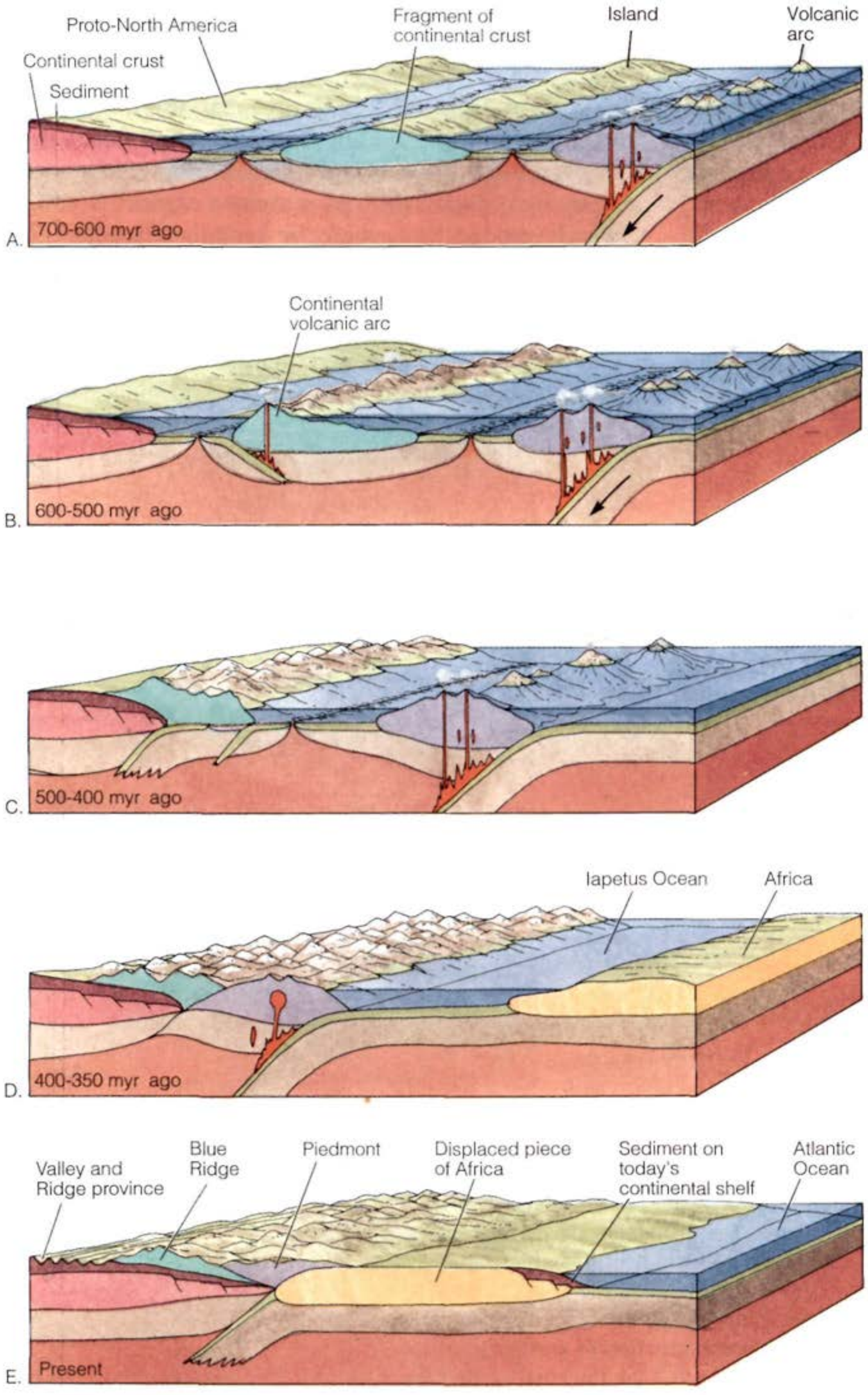


Figure 7.24 A sequence of subduction, collision, and accretion events that explains the evolution of the Appalachians in terms of plate tectonics. Iapetus is the posthumous name of the ocean that disappeared about 350 million years ago when Africa collided with North America. Iapetus was one of the minor Greek gods and father of Atlas and Prometheus. A. Six hundred million years ago, a small fragment of continental crust (an island) lay offshore of North America. Beyond the island was a subduction zone and an arc of volcanoes. B. A new subduction zone starts beneath the island about 500 million years ago. C. Between 400 and 500 million years ago, the island collides with North America (these rocks can be seen today in North Carolina and Virginia), and the Appalachians start to form. D. Between 350 and 400 million years ago, the arc of volcanoes collides with North America. The Iapetus Ocean slowly closes as North Africa approaches and eventually collides. E. When Pangaea broke apart about 200 million years ago, a fragment of Africa remained attached to North America. The passive continental margin so formed (today's margin) bounds the eastern edge of North America.

Guest Essay

Plate Tectonics and Continental Drift: A Skeptic's View

Paul D. Lowman Jr., geologist, B.S., from Rutgers University, Ph.D. from University of Colorado. Hired in 1959 by NASA, Dr. Lowman has carried out research in lunar geology, comparative planetology, remote sensing, and tectonics.

"A master plan into which everything we know about the Earth seems to fit." W. K. Hamblin's 1978 description of plate tectonics is accepted by almost all geologists. I am not one of them; this essay will explain why.

Let me put my views up front. Plate tectonic theory is undeniably a great achievement and an enormous stimulus to geologic education. Sea-floor spreading, transform faulting, and subduction are supported by several independent methods. However, I believe that plate tectonics is basically an explanation of ocean basin geology and that plate tectonics does *not* imply continental drift. Here are some of my reasons.

To begin, drift of the major continents, in particular trans-Atlantic drift, has not been directly demonstrated. The NASA Crustal Dynamics Project has measured baselines several thousand kilometers long, with a precision equal to the length of a fingernail, by space geodesy methods—satellite laser ranging (SLR) and very long baseline interferometry (VLBI). What *has* been directly determined is plate motion in and around the Pacific basin. Islands on the Pacific Plate have been shown to be moving toward Japan at several centimeters per year. The plate has also been shown to be internally rigid. Movement of Australia has also been directly measured, but this "continental drift" occurs as the result of motion of a dominantly oceanic plate. Baja California is similarly moving as part of the Pacific Plate.

Space geodesy measurements across the Atlantic Ocean have shown increasing distances close to those predicted from sea-floor spreading anomalies and are widely interpreted as proving continental drift. However, to demonstrate drift as a corollary of plate tectonics, it must be shown that entire plates—North American, Eurasian, and others—are moving as rigid units and that we are not seeing localized intraplate deformation. When I helped plan the Crustal Dynamics Project baselines in 1979, we recommended many intracontinental baselines to allow for such deformation, only a few of which could be established.

The weakness in supposed trans-Atlantic drift measurements stems from what I consider intraplate deformation. The opposing coastlines of North America and Europe are, in plate theory, "passive" margins. However, it is becoming clear that these margins are affected by contemporary seismicity, horizontal compression, and crustal deformation, expressed as active faulting and folding. Furthermore, continental crust on these margins is now known to be "thin-skinned," underlain by great thrust faults that may decouple surface rocks from the basement. It follows that space geodesy stations on such passive margins cannot truly measure the motion of entire plates until they are tied into extensive intraplate nets, and observations are continued for decades.

The revival of interest in continental drift was triggered by studies in paleomagnetism, specifically, apparent polar wander paths (APWP). The APWP over several hundred million years are different as determined from different continents. But if the continents are restored to their supposed pre-drift positions, the APWP coincide. What can be wrong with such an elegant argument?

First, paleomagnetic techniques can determine only paleolatitude, not paleolongitude. When the individual pole positions, rather than averaged curves, are plotted on world maps as done by A. A. Meyerhoff, the scatter is huge, often wider than the Atlantic. Another problem is that the remnant magnetism of a given rock may be affected by tectonic deformation, metamorphism, chemicals changes, and secular magnetic variation, or variation measured in decades of over 40 degrees in longitude as measured at London in 360 years. It is thus not surprising that paleo-pole positions are very hard to determine, especially for rocks more than a few million years old.

There are more inconsistencies and anomalies in APWP determinations than the student would guess from most diagrams. One of the most striking anomalies was discovered in 1979 by P. W. Schmidt and B. T. T. Embleton. Plotting APWP for two long Proterozoic periods from North America, Africa, Australia, and Greenland, they found that these paths coincided, unlike th ,' different paths found for younger rocks. Such coin :idental paths directly contradict continental drift, implying that these continents were in the same relative positions for hundreds of years. The only solution Schmidt and Embleton could find for this anomaly was that the Earth's radius had increased some 45 percent during the period involved—a balloon-like inflation. When it comes to the expanding Earth hypothesis, I take the majority view, which is that the idea verges on the preposterous. However, many eminent geologists such as S. Warren Carey and L. C. King have vigorously argued for an expanding

Earth. Close examination of these arguments shows that their proponents have generally been forced to them by anomalies impossible to explain with plate tectonic theory. Surely this tells us something about the theory.

Let me now examine another widely cited line of evidence, the distribution of land or freshwater fossils. The general reasoning is that similar fossils are found on continents now separated by wide oceans, implying that these continents were once joined. The problem is one of dispersal: How could land dwellers get across impassable oceans?

This question has several answers. To illustrate one, consider the iguanas, snakes, and tortoises of the Galapagos Islands, 800 km from the nearest continent. One way these creatures could have been carried across the oceans is on driftwood. A television program several years ago showed a South American iguana clinging to a log being washed out into the Pacific. Whether this particular lizard made it to the Galapagos I have no idea, but such accidents must occasionally happen off camera. Large rafts of vegetation, in effect floating islands, have sometimes been sighted in the Atlantic, ripped off by the Amazon River. These can carry all sorts of land animals, some of which probably survive to reach the opposite coast. These accidental ocean crossings are uncommon, but over geologic time spans—hundreds of thousands of years or more—they are almost certain to happen for any particular species.

The fossil record may actually argue directly against continental drift. The usual Pangaea reconstruction, for example, shows India before drift as nestled against what are now Antartica and southern Africa. India supposedly broke away from Gondwana, drifting northward until it collided with Asia. However, this familiar scenario has been strongly disputed by two paleontologists, Chattterjee and Hotton. Extensive documented catalogues of Indian fossils show many affinities with the fauna of Asia, but very few of those with the famous Antartica and Australia. Furthermore, the Indian fossils do not show the endemism, like the well-known marsupial fauna of Australia, that a former island continent should display.

The most obvious evidence for drift is the well-known parallel coastlines around the Atlantic Ocean. I agree that this cannot be coincidence. But there are anomalies in even this glaringly obvious evidence. First, a glance at any map will show that the circumpolar continent margins are not at all parallel. Another Arctic anomaly is the Nares Strait, between Greenland and Ellesmere Island. The opening of Baffin Bay, to produce the roughly parallel coastlines, would produce hundreds of kilometers of lateral offset in the Nares Strait. To test this concept, geologists who had mapped in Greenland and Ellesmere Island met in 1981 to compare notes. They found at least four geologic markers crossing the Strait showing no significant offset, and others at least consistent with no offset. The majority opinion was that there had been little horizontal movement along the Strait, implying that Greenland had not drifted from North America.

One of the main problems faced by Alfred Wegener was the motive force driving the continents. Plate tectonic theory had appeared to have solved this problem by sea-floor spreading and subduction. These well-documented phenomena can in principle produce two plate-driving forces—ridge push and slab pull. In addition, gravity sliding on the asthenosphere might move oceanic lithosphere, which becomes demonstrably colder and denser with increasing distance from spreading centers.

Continental drift still faces Wegener's motive force problem, at least for the circum-Atlantic continents. The difficulty is this. Continents such as North and South America have silica-rich leading edges (the Pacific margins). Such crust is not dense enough to be subducted and is demonstrably not being subducted. Slab pull and gravity sliding will simply not work for plates whose leading edges are continental. That leaves ridge push to do the job alone. But it seems utterly unrealistic to imagine it driving the enormous Eurasian Plate, extending over 120 degrees of longitude from the Mid-Atlantic ridge to Siberia, especially since its leading edge is non-subductable continental crust. Furthermore, the asthenosphere, generally considered a zone of partial melting in the mantle, cannot be detected under stable continental crust (in contrast to the ocean basins). The driving force problem, incidentally, does not apply to Australia, which is part of a largely oceanic plate.

The classic plate tectonic mechanisms have been demonstrated by several independent methods, and plate rigidity and motions in the Pacific Basin measured by two different space geodesy techniques. In contrast, continental drift, by itself, remains controversial and based on uncertain and subjective arguments. This dichotomy points to my own proposal: plate tectonics with fixed continents. I suggest that sea-floor spreading and subduction do not necessarily lead to continental drift, and specifically not to the classic trans-Atlantic drift.

Two questions will occur to the reader. First, sea-floor spreading without continental drift implies subduction under the Atlantic passive margins. Why have we not observed it? I have two answers. The spreading rates from the Mid-Atlantic ridge are very low, implying correspondingly low, subduction rates. Such slow subduction, probrably less than 2 cm/year, could occur without generating deep focus earthquakes, the movement of downgoing slabs occuring with ductile shear. The second answer is that we may have observed passive margin subduction in the form of landward dipping slabs of crust extending scores of kilometers below the Moho, under Scotland and Southern Africa. This discovery was made by two different groups in the 1980s by deep seismic reflection profiling. No one but me has interpreted these as subducting crust, but they have the geometry such zones should have.

A second question is how can I account for the parallei continental margins around the Atlantic without continental drift? My answer is that if passive margin subduction is occuring, it may be causing "tectonic erosion" in D. E. Karig's phrase—the grinding away of the overlying crust by the downgoing slabs. This seems to be happening in the much more rapid circum-Pacific subduction zones, and it could happen, more slowly, around the Atlantic. Sea-floor spreading is in general occuring at the same rate on opposite sides of the Mid-Atlantic ridge. Corresponding subduction zones should thus have roughly similar rates of subduction erosion on opposite continental margins, resulting in the rough parallelism that has excited interest for centuries.

Regardless of whether my particular theory is correct, it should be clear that plate tectonic theory is not the seamless robe it is often considered. Ruling theories have been overthrown before, and a similar fate may await "plate tectonics and continental drift." I urge students to read the original sources; to look at the original data; to keep open minds on alternative concepts; and, most important, to think things through for themselves.

Summary

1. Stratification results from the arrangement of sedimentary particles in layers. Each bed in a succession of strata is distinguished by its distinctive thickness or character.
2. Two basic laws underlie stratigraphy: the law of original horizontality and the principle of stratigraphic superposition.
3. Substantial breaks or gaps in the sedimentary record are called unconformities. They record changes in environmental conditions of deposition or erosion and loss of earlier-formed sediment.
4. Unconformities record the close relationship between tectonics, erosion, and sedimentation.
5. The geologic column, a composite section of strata fitted together on the basis of relative age, has been carried around the world by correlation of strata based on the fossils in sedimentary rocks.
6. The absolute ages of strata in the geologic column have been determined by radiometric dating.
7. There are two families of sediment: clastic and chemical. Clastic sediment is material transported as solid bits of rocks and minerals. Chemical sediment forms when material is transported in solution and then deposited. If the cause of deposition is biochemical, a chemical sediment is called a biogenic sediment.
8. Sediment is lithified to sedimentary rock by cementation or recrystallization of the sediment particles.
9. Various arrangements of the particles in strata are seen in parallel strata, cross strata, paired strata, and poorly sorted layers.
10. Clastic sedimentary rocks, like sediments, are classified on the basis of predominant particle size. Conglomerate, sandstone, siltstone, and shale are the rock equivalents of gravel, sand, silt, and clay, respectively.
11. Metamorphism involves changes in mineral assemblage and rock texture and occurs in the solid state as a result of changes in temperature and pressure.
12. Mechanical deformation and chemical recrystallization are the processes that affect rock during metamorphism.
13. The presence of intergranular fluid greatly speeds up metamorphic reactions.
14. Foliation, as expressed by a direction of easy breakage in a metamorphic rock, arises from parallel growth of minerals formed during metamorphism.
15. Metamorphism can be explained by plate tectonics. Burial metamorphism occurs within the thick piles of sediment at the foot of continental slopes and in the submarine fans off the mouths of the world's great river systems; regional metamorphism is found in regions of subduction and continental collision.
16. Regional metamorphism, which involves both mechanical deformation and chemical recrystallization, is the result of plate tectonics. Regionally metamorphosed rocks are produced along subduction and collision edges of plates.
17. Two major structural units can be discerned in

the continental crust. Cratons are ancient portions of the crust that are tectonically and isostatically stable. Separating and surrounding the cratons are orogens of highly deformed rock, marking the sites of former mountain ranges.

19. An assemblage of cratons and deeply eroded orogens that forms the core of a continent is a continental shield.
20. There are five kinds of continental margins: passive, convergent, collision, transform fault, and accreted terrane.
21. Passive margins develop by rifting of the continental crust. The Red Sea is an example of a young rift, and the Atlantic Ocean is a mature rift.
22. Continental convergent margins are the locale of belts of metamorphic rock, chains of stratovolcanoes (magmatic arc), and belts of granitic batholiths.
23. Collision margins are the locations of mountain systems. Transform fault margins occur where the edge of a continent coincides with the transform fault boundary of a plate.
24. Accreted terrane margins arise from the addition of blocks of crust brought in by subduction and transform fault motions.

Important Terms to Remember

accreted terrane (p. 184)
bed (p. 160)
bedding (p. 160)
burial metamorphism (p. 174)
chemical sediment (p. 167)
clastic sediment (p. 167)
continental shield (p. 179)
craton (p. 179)
cross bedding (p. 168)
differential stress (p. 172)
foliation (p. 173)
fossil (p. 169)
geologic column (p. 163)
metamorphic facies (p. 177)
metamorphism (p. 171)
original horizontality (law of) (p. 161)
orogen (p. 179)
regional metamorphism (p. 174)
stratum, strata (p. 160)
stratification (p. 160)
stratigraphic superposition (principle of) (p. 161)
stratigraphy (p. 161)
unconformity (p. 161)
uniform stress (p. 172)
varve (p. 168)

SEDIMENTARY ROCKS

coal (p. 170)
conglomerate (p. 167)
dolostone (p. 170)
limestone (p. 170)
peat (p. 170)
sandstone (p. 167)
shale (p. 167)
siltstone (p. 167)

METAMORPHIC ROCKS

amphibolite (p. 174)
gneiss (p. 174)
greenschist (p. 174)
marble (p. 176)
phyllite (p. 174)
quartzite (p. 176)
schist (p. 174)
slate (p. 174)

Questions for Review

1. What two laws underlie the study of stratigraphy? What conclusion would you draw if you observed a pile of sedimentary strata in which the strata are vertical?
2. How and why do breaks occur in stratigraphic sequences, and what significance do they have for determining the history of the Earth?
3. How are strata correlated from place to place?
4. What is the difference between the relative and absolute age of a stratum?
5. What is the geologic column? Name the four eons of the column.
6. Starting from the oldest, name in order the periods of the Paleozoic Era.
7. Name the two families of sediments and describe the basic difference between them.
8. On what basis are clastic sediments and sedimentary rocks classified?
9. Describe two chemical reactions that can lead to precipitation of chemical sediments.
10. If you picked up a sedimentary rock containing rounded clasts about 5 cm in diameter, what name would you give to the rock?
11. What role does the biosphere play in the formation of sediment? Name a biogenic sedimentary rock and suggest how it may have formed.

12. What would you call a sedimentary rock composed entirely of broken bits of plant matter?
13. What features in a sediment or sedimentary rock are responsible for stratification?
14. Briefly describe the factors that control metamorphism.
15. What is foliation? The presence of foliation is a sure clue that a rock has been metamorphosed. Why?
16. A distinctly foliated rock contains quartz, potassium feldspar, garnet, and sillimanite; under what grade of metamorphism would it have formed?
17. What is regional metamorphism?
18. What is the geological setting of regional metamorphism? Name two places in the world where regional metamorphism is probably happening today.
19. What is burial metamorphism? Suggest some place on the Earth where it is probably happening today.
20. What is the metamorphic facies concept and how does it help in the study of metamorphic rocks?
21. Name three minerals that are found only in metamorphic rocks.
22. In what way does the presence of an intergranular fluid influence the speed with which metamorphism proceeds?
23. What are cratons and how do they differ from orogens? Name three orogens in North America that are less than a billion years old.
24. What is the origin of the Cascade Range? the Appalachians?
25. What evidence indicates that plate tectonics has been operating for at least the last 1.8 billion years of Earth history?
26. Name the five kinds of continental margins and describe how they form.
27. Describe the sequence of events that leads to the opening of a new ocean basin flanked by two passive continental margins.
28. With what kinds of continental margin is mountain building associated?
29. How does an accreted terrane margin form? Name a continental margin that was modified by terrane accretion.

Questions for A Closer Look

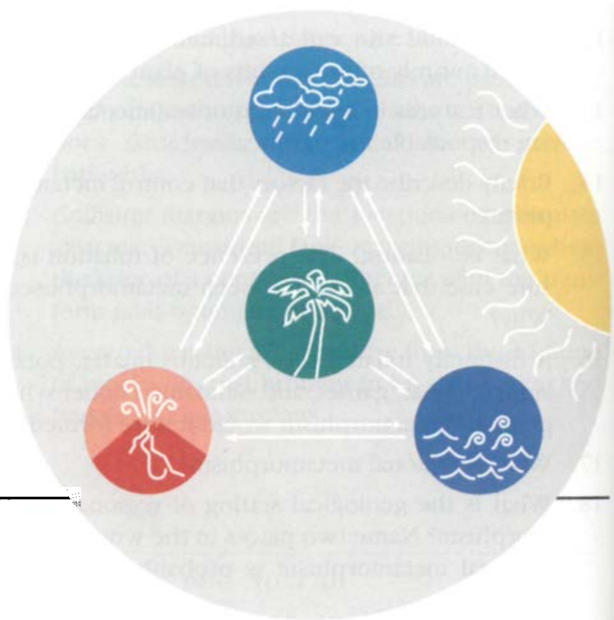
1. What is the law of radioactive decay?
2. If you started with 2000 radioactive atoms and the half-life of the atoms is 1 week, how many radioactive atoms would remain after 4 weeks?
3. How can naturally occurring radioactive atoms be used to determine the absolute ages of minerals.
4. A grain of mica from a granite contains 5000 atoms of ^{40}K and 600 atoms of ^{40}Ar. What is the absolute age of the mica? Note: the half-life of ^{40}K is 1.3 billion years.
5. Describe how radiometric ages are used to determine the absolute age of the geologic column.
6. Describe the changes in a shale as it metamorphosed successively from a slate, to a phyllite, to a schist.
7. What is a greenschist? Why does a greenschist differ from a slate even though both rocks form at the same grade of metamorphism?
8. Why do most marbles and quartzites lack foliation?
9. How does a quartzite differ from a sandstone?

Questions for Discussion

1. What kind of sediment can you identify in the area in which you live? Identify the source of the sediment, how the sediment is transported, where, and why it is being deposited.
2. Discuss the importance of the atmosphere, hydrosphere, and biosphere in the formation of sediment. What kind of sediment would you expect on a planet such as Mars that lacks a hydrosphere and a biosphere?
3. Discuss how the solid Earth reservoir interacts with the other three reservoirs of the Earth system. Why is it that, even though events in the solid Earth, such as the raising of a mountain range, happen very slowly, while events in the atmosphere, hydrosphere, and biosphere happen rapidly, the solid Earth plays a dominant role in long-term changes in the other reservoirs?

PART THREE

The Earth's Blanket of Water and Ice



Water and the Hydrologic Cycle

Water makes the Earth unique. Seen from space, the Earth appears mostly blue and white because of its cover of water, snow, ice, and clouds. Although water has been detected on other bodies of the solar system, it does not appear to be present as a liquid anywhere except on our planet. The surface of Venus is so hot that water exists only as vapor, while very low temperature and pressure at the surface of Mars mean that water can exist there only as vapor or as ice. The surface of Ganymede, the largest of Jupiter's moons, is so frigid that it is covered by a thick "lithosphere" of ice.

In the Introduction, we named the four parts (or reservoirs) that make up the Earth system—lithosphere, hydrosphere, atmosphere, and biosphere—and defined the hydrosphere as the part containing the totality of the Earth's water, exclusive of atmospheric water vapor. Although this definition is straightforward, it is not strictly accurate, for water actually is present in all four parts of the Earth system. In other words, the hydrosphere overlaps to some extent with the lithosphere, atmosphere, and biosphere. It is mainly for convenience of discussion that we distinguish the hydrosphere as the "water sphere," for in it resides the bulk of the Earth's water.

Most of the Earth's water, more than 97 percent, resides in the oceans. Next in importance are the myriad bodies of frozen water (snow and ice) that occupy the high mountains and polar latitudes of our planet. All the remaining water—including that in lakes and streams, in the atmosphere, and in the ground—amounts to only about 1 percent of the total; yet this is the water we are most conscious of and rely on in our daily lives.

The physical state of water is controlled by temperature and pressure. At high temperatures or low pressures, water vapor is the stable state for H_2O, whereas ice forms at low temperatures or high pressures. The air pressure at sea level and that at the top of Mount Everest represent the extremes of air pressures at the Earth's surface. Surface temperatures range from about -100°C to +50°C (-148°F to 122°F). Within these limits of temperature and pressure, water can exist naturally in all three states of matter—solid (ice), liquid (water), or gas (water vapor). In the view across Lemaire Channel in Antarctica, we can witness water in three states: as seawater, as ice in glaciers and floating sea ice, and as clouds that have condensed from water vapor in the air.

The movement of water between the four Earth reservoirs constitutes the *hydrologic cycle.* Both the day-to-day and long-term changes we observe in the hydrosphere are powered by the Sun's heat energy, which evaporates water from the ocean and the land surface. The water vapor thus produced enters the atmosphere and moves as part of the flowing air. Some of the water vapor condenses to form clouds and is precipitated as rain or snow back into the ocean or onto the land. Rain falling on land drains off into streams that flow toward the oceans, seeps into the ground, or is evaporated back into the air, where it is further recycled. Part of the water in the ground is taken up by plants, which return water to the atmosphere through their leaves by a process called transpiration. Most snow remains on the ground for one or several seasons until it melts and the melt water flows away, but snow that nourishes glaciers often remains locked up for hundreds or thousands of years until it, too, eventually melts or evaporates.

Glaciers meet the sea along Lemaire Channel on the Antarctic Peninsula, producing an array of icebergs scattered across the ocean surface.

Although water is always in a state of movement and is continuously being cycled from one reservoir to another, the total volume of water in each reservoir is approximately constant over short time intervals. Over lengthy intervals, however, the volume of water in the different reservoirs can change dramatically. During glacial ages, for example, vast quantities of water are evaporated from the oceans and precipitated on land as snow. The snow slowly accumulates to build ice sheets that are thousands of meters thick and cover vast areas where none exist today. At such times, the amount of water removed from the oceans is so large that the world sea level falls 100 m (110 yd) or more, and the expanded glaciers increase the ice-covered area of the Earth by more than 300 percent.

An important consequence of the hydrologic cycle is the varied landscapes of the Earth. The erosional and depositional effects of streams, waves, and glaciers, coupled with the tectonic movements of crustal rocks, have produced a diversity of landscapes that make the Earth's surface unlike that of any other planets in the solar system. In its effect on erosion and sedimentation, the hydrologic cycle is intimately related to the rock cycle. Furthermore, it is a key component of an array of biogeochemical cycles that control the composition of the atmosphere and influence all living creatures on the Earth.

CHAPTER 8

The World Ocean

Early Polynesian explorers, navigating by stars, winds, and currents, cross the vast expanse of the central Pacific Ocean in search of new islands to colonize.

Polynesian Navigators

More than 1300 years ago, Polynesian explorers set out from Havai'i (an island near Tahiti, now called Raiatea) in great double-hulled canoes on voyages of discovery and conquest in the vast and then-unknown expanse of the North Pacific Ocean. On one such voyage, a kahuna, or priest, named Pa'ao discovered Hawaii and, retracing his route homeward, subsequently assembled a fleet to colonize the islands. Today, we can board a comfortable jet airliner and fly nonstop the 4400 km (2700 mi) from Honolulu to Tahiti in less than six hours, but a voyage by canoe across the uncharted ocean must have taken many weeks and been filled with numerous hazards. The Hawaiian islands were discovered by chance, but subsequent trips between Hawaii and the home islands of the Society and Marquesas groups required a degree of navigational skill that is difficult to imagine. How did the ancient Polynesians repeatedly traverse great ocean distances without the aid of a compass or other modern navigational aids?

Like competent modern sailors, the Polynesian navigators became familiar with the winds and current systems that affected their vessels. In sailing from Havai'i to Hawaii, a canoe had to cross three wind systems and the surface ocean currents related to them (Fig. 8.1). Raiatea lies in the southeast tradewinds belt, where prevailing winds blow from the southeast, as well as in the region of a great westward-flowing ocean current that lies just south of the equator. The initial part of the voyage therefore involved sailing somewhat east of north across this zone until, just north of the equator, a belt of light variable winds and an eastward-flowing current must be passed. Still farther north, a canoe encountered the northeast tradewinds and another westward-flowing current system that helped carry it northwestward toward Hawaii. Because the winds and currents change with latitude, by observing them ancient navigators could get a good idea of their approximate position relative to the equator. A competent sailor could also steer a boat using the major ocean swells generated by the tradewinds. Maintaining a course involves keeping the boat consistently oriented with respect to passing swells.

Whenever land was in sight, a course could be maintained by aligning the vessel with several fixed landmarks such as mountain peaks, rocky promontories, or small islands. At night, a mariner could steer with respect to a succession of stars rising above the horizon; as one star rose too high to steer by, a newly rising star would then be tracked. Polynesian navigators knew the rising and setting points of more than 150 stars, and these points served them well in lieu of a compass. Another astronomical trick they employed involved zenith stars. A star that appears to pass directly over an island will also appear to pass over all points due west and east of the island. Polynesians used such stars to determine when they had reached the same latitude as their island destination, and then they maintained the appropriate course until land was seen.

The ancient Pacific navigators also had ways of an-

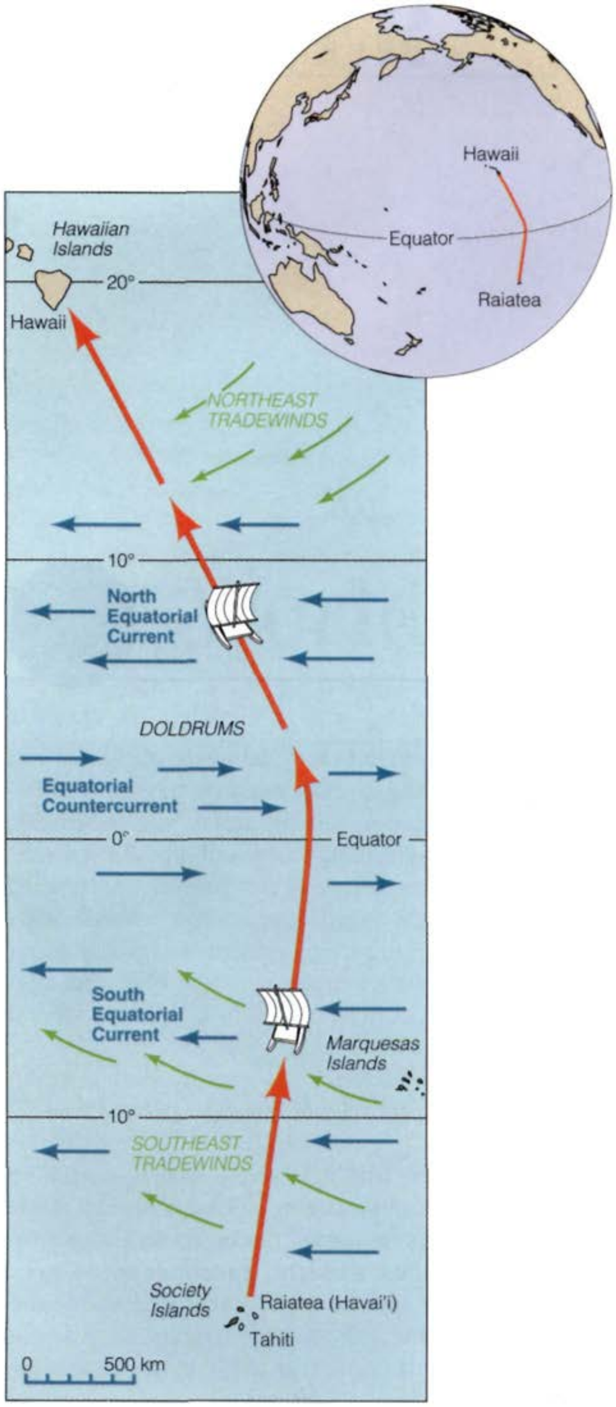


Figure 8.1 Likely route of discovery of the Hawaiian islands. Double-hulled canoes, pushed by the southeast tradewinds, sailed east of north from Havai'i (Raiatea) in the Society islands across the South Equatorial Current. Moving northward across the Equatorial Countercurrent in the Doldrums, the canoes then picked up the North Equatorial Current and the northeast tradewinds, which carried them northwestward toward landfall at the island of Hawaii.

ticipating a landfall. For example, clouds tend to accumulate around high islands, and the base of the clouds often is illuminated by light reflected off bright, sunlit lagoons. The bright glow of the clouds can be visible long before land is seen. Shallow, submerged reefs are indicated by lighter colored water above them and can be used as navigational aids. Plant debris drifting on the ocean surface may be evidence of land to windward. Finally, birds that venture seaward in the morning to fish and then return to land at dusk can point the way toward land over the horizon.

The Polynesians lacked a written language and committed to memory all the information needed to retrace a lengthy ocean voyage. Instead, they relied on a lifetime of personal experience, an intimate knowledge of the visible universe, and a clear understanding of the interrelationship of land, sea, atmosphere, and living organisms.

THE OCEANS

Imagine, if you can, a dry Earth, devoid of water. The Earth's surface would appear far different from the one familiar to us. Viewed from an orbiting spacecraft, a dry Earth would no longer have a bluish color (and we would need another title for this book), the land would lack a vegetation cover, and no clouds would obscure the surface. The Earth would resemble Mars or the rocky moons of Jupiter. We would see the high-standing continents ending where their bordering continental slopes meet a great expanse of empty sea floor. As we learned in Chapter 6, this primary topography reflects the contrasting densities and thicknesses of continental (less dense, thicker) and oceanic (denser, thinner) crust. Circling the Earth, we would see several vast interconnected basins, each floored with oceanic crust and rimmed with continental crust.

If these huge basins were now slowly filled with water, the scene would be transformed. The rising water would initially fill the deepest parts of the basins, creating a number of shallow seas, but as the water level continued to rise, these seas would merge to form a larger and larger ocean that eventually would creep up the continental slopes and spill across the continental shelves. With the ocean basins filled to capacity, more than two-thirds of the Earth's surface would now be covered by water and the Earth would be a unique planet in the solar system, for it would have become the Blue Planet.

Under the ocean, beyond the continental slope, lies the remote world of the deep-ocean floor. With devices for sounding the sea bottom and for sampling its sediment, teams of oceanographers and marine geologists have explored the ocean floor and greatly expanded our knowledge of the submarine regions. Scuba-diving geologists have visited, photographed, and mapped areas of seafloor at depths as great as 70 m (230 ft), and observers in specially designed submersible crafts have descended more than 6 km (3.7 mi) to visit the greatest depths of the ocean floor (see "Guest Essay").

Because of this intensive research involving many nations, we are gradually coming to understand the oceans. The romanticist in each of us may regret that beliefs and legends built up through more than 3000 years of human history—monsters, mermaids, strange and threatening sea gods, fabled cities and castles believed to have sunk into watery deeps—have vanished. These and other poetic visions have faded away as scientific knowledge has steadily increased. In return, however, that knowledge has helped us appreciate the fragile environment of the oceans, which is responsible for a large part of the Earth's biological heritage.

Ocean Geography

Seawater covers 70.8 percent of the Earth's surface. The land comprising the remaining 29.2 percent is unevenly distributed. This uneven distribution is especially striking when we compare two views of the globe: one from a point directly above Great Britain and the other from a point directly above New Zealand (Fig. 8.2). In the first view, more than *46* percent of the viewed hemisphere is land, whereas in the second view it is more than 88 percent water and only about 12 percent land. The uneven distribution of land and water plays an important role in determining the paths along which water circulates in the open ocean and the marginal seas.

Most of the water on our planet is contained in three huge interconnected basins—the Pacific, Atlantic, and Indian oceans. (The Arctic Ocean is generally considered an extension of the North Atlantic.) (Fig. 8.3). All three are connected with the Southern Ocean, a body of water south of 50°S latitude that completely encircles Antarctica. Collectively, these four vast interconnected bodies of water, together with a number of smaller ones, are often referred to as the *world ocean.* The smaller water bodies connected with the Atlantic Ocean include the Mediterranean, Black, North, Baltic, Norwegian, and Caribbean seas, the Gulf of Mexico, and the Baffin and Hudson bays. The Persian Gulf, Red Sea, and Arabian Sea are part of the Indian Ocean, while among the numerous marginal seas of the Pacific Ocean are the Gulf of California, Bering Sea, Sea of Okhotsk, Sea of Japan, and the East China, South China, Coral, and Tasman seas. All these seas and gulfs vary considerably in shape and size; some are almost completely surrounded by land, whereas others are only partly enclosed. Each owes its distinctive geography to plate tectonics, for this ongoing process has led to the creation of numerous small basins both in and adjacent to the major ocean basins.

Depth and Volume of the Oceans

Before the present century, little was known about the depth of the oceans. Water depths were determined from soundings made with either a weighted

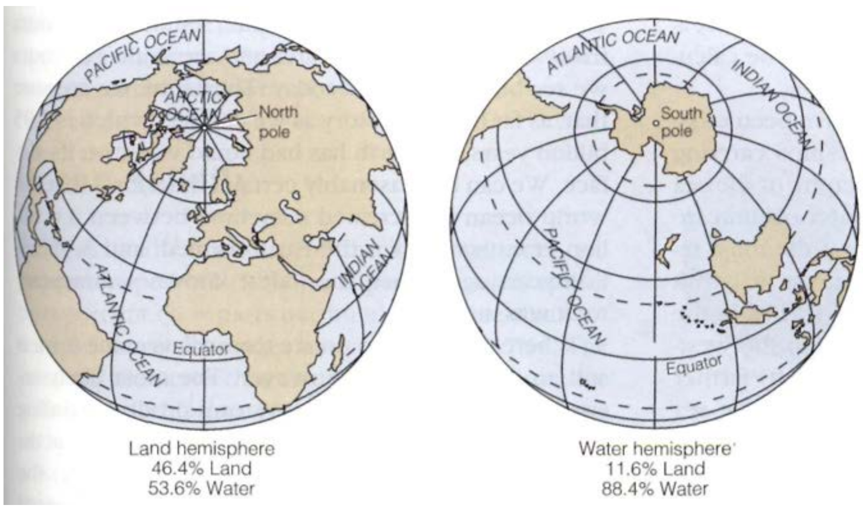


Figure 8.2 The unequal distribution of land and ocean can be seen if we view the Earth from above Britain and above New Zealand. In the first view, land covers nearly half the hemisphere, while in the other nearly 90 percent of the hemisphere is water.

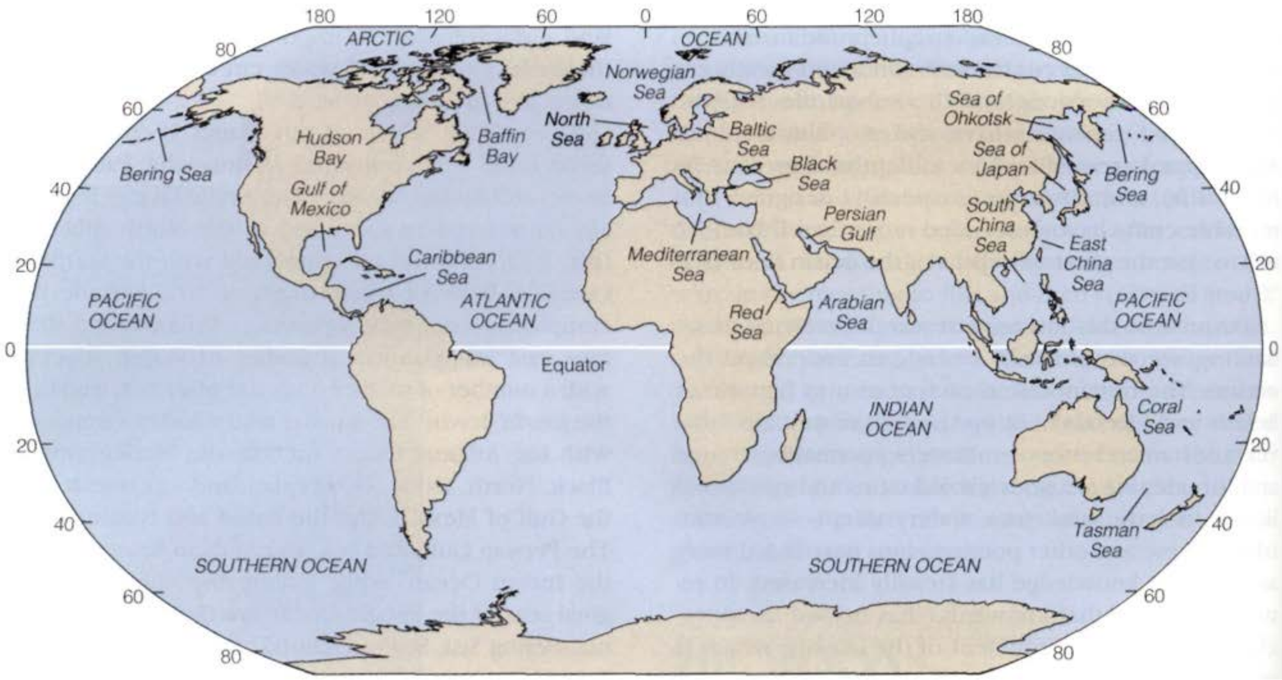


Figure 8.3 Major and minor basins of the world ocean.

hemp line or a strong wire lowered from a ship. Although this technique proved satisfactory and relatively rapid in shallow water, it could take 8 to 10 hours to recover a weighted wire in water thousands of meters deep. By the close of the nineteenth century, about 7000 measurements had been made in water more than 2000 m (6500 ft) deep, and fewer than 600 in water deeper than 9000 m. (29,500 ft) In the 1920s ship-borne acoustical instruments called *echo sounders* were developed to measure ocean depths. An echo sounder generates a pulse of sound and accurately measures the time it takes for the echo bouncing off the seafloor to return to the instrument. Because the speed of sound traveling through water is known, the water depth beneath a ship can be calculated.

Over the past 70 years, the oceans have been crisscrossed many thousands of times by ships carrying echo sounders. As a result, the topography of the sea floor and the depth of the overlying water column are known in considerable detail for all but the most remote parts of the ocean basins. The greatest ocean depth yet measured (11,035 m; 36,205 ft) lies in the Mariana Trench near the island of Guam in the western Pacific. This is more than 2 km (6500 ft) farther below sea level than Mount Everest rises above sea level. The average depth of the sea, however, is about 3.8 km, (12,500 ft) compared to an average height of the land of only 0.75 km (2460 ft).

If we measure the area of the sea and calculate its average depth, we can then calculate that the present volume of seawater is about 1.35 billion cubic kilometers (324 million mi^3) (Fig. 8.4); more than half this volume resides in the Pacific Ocean. We say *present* volume because the amount of water in the ocean fluctuates somewhat over thousands of years, with the growth and melting of continental glaciers (Chapters 10 and 14).

Age and Origin of the Oceans

The Earth's oldest rocks include sedimentary strata that were deposited by water and are similar to strata we see being deposited today. Therefore, we are sure that, as far back in history as we can see, which is 3.95 billion years, the Earth has had liquid water on its surface. We can be reasonably certain, therefore, that the world ocean was created sometime between *4.6* billion years ago, when the Earth formed, and 3.95 billion years ago, when the oldest known sedimentary rock was made.

Where the water to create the oceans came from is still an open question, however. The most likely answer is that it condensed from steam produced during primordial volcanic eruptions. Because volcanic activity has persisted throughout the Earth's history, the volume of the world ocean has probably increased through time.

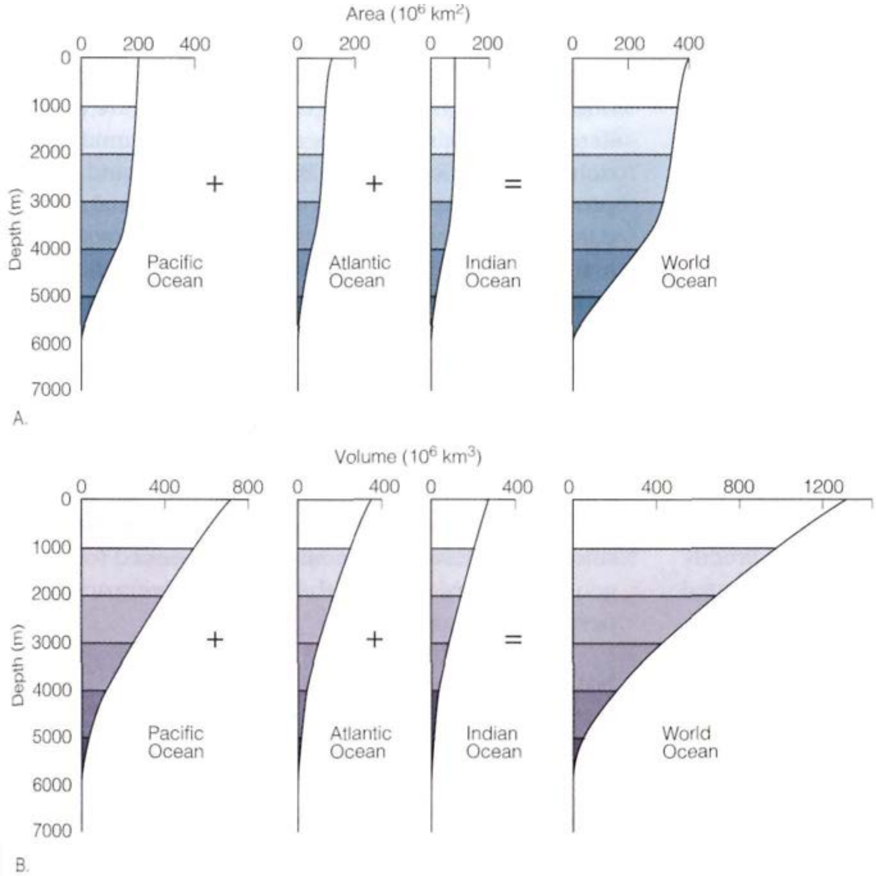


Figure 8.4 Area A. and volume B. of the oceans. The Pacific represents nearly half the area and half the volume of the oceans, with the Atlantic and Indian oceans being comparable to each other in both size and volume. Although the deepest known place in the oceans lies more than 11,000 m below sea level, nearly all the water lies at a depth of less than 6000 m

THE SALTY SEA

About 3.5 percent of average seawater, by weight, consists of dissolved salts, enough to make the water undrinkable. It is enough also, if these salts were precipitated, to form a layer about 56 m (183 ft) thick over the entire seafloor.

Ocean Salinity

Salinity is the measure of the sea's saltiness, expressed in per mil (‰ = parts per thousand) rather than percent (% = parts per hundred). The salinity of seawater normally ranges between 33 and 37‰. The principal elements that contribute to this salinity are sodium and chlorine. Not surprisingly, when seawater is evaporated, more than three quarters of the dissolved matter is precipitated as common salt (NaCL). However, seawater contains most of the other natural elements as well, many of them in such low concentrations that they can be detected only by extremely sensitive analytical instruments. As can be seen in Figure 8.5, more than 99.9 percent of the salinity is caused by only eight ions.

Where do these ions come from? Each year streams carry 2.5 billion tons of dissolved substances to the

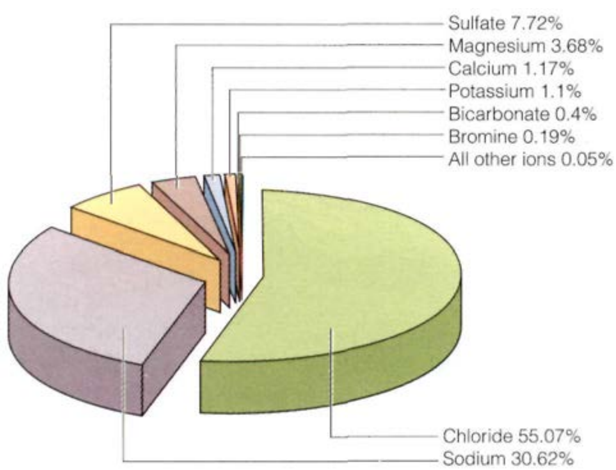


Figure 8.5 Principal ions in seawater. More than 99.9 percent of the salinity of seawater is due to eight ions, the two most important of which (Na^+ and Cl^-) are the constituents of common salt.

sea. As exposed crustal rocks interact with the atmosphere and the hydrosphere (rainwater), cations are leached out and become part of the dissolved load of streams flowing to the sea. The principal anions found in seawater, on the other hand, are believed to have come from the mantle. Chemical analyses of gases released during volcanic eruptions show that the most important volatiles are water vapor (steam), carbon dioxide (CO_2), and chloride (Cl^1) and sulfate (SO_4^{2-}) anions. These two anions dissolve in atmospheric water and return to the Earth in the form of precipitation, much of which falls directly into the ocean. Part of the remainder is carried to the sea dissolved in river waters. Volcanic gases are also released directly into the ocean from submarine eruptions. Other sources of ions include dust eroded from desert regions and blown out to sea, and gaseous, liquid, and solid pollutants released through human activity either directly into the oceans or carried there by streams or polluted air.

The quantity of dissolved ions added by rivers over the billions of years of Earth history far exceeds the amount now dissolved in the sea. Why, then, doesn't the sea have a higher salinity? The reason is that chemical substances are being removed at the same time they are being added. Some elements, such as silicon, calcium, and phosphorus, are withdrawn from seawater by aquatic plants and animals to build their shells or skeletons. Other elements, such as potassium and sodium, are absorbed and removed by clay particles and other minerals as they settle slowly to the sea floor. Still others, such as copper and lead, are precipitated as sulfide minerals in claystones and mudstones rich in organic matter. Because these and other processes of extraction are essentially equal to the combined inputs, the composition of seawater remains virtually unchanged.

Has the ocean always been salty? The best evidence of the sea's past saltiness is the presence, in marine strata, of salts precipitated by the evaporation of seawater. Marine strata containing salts that were concentrated by the evaporation of seawater are common in young sedimentary basins, but they are not known from rocks older than about a billion years. Possibly this is because ancient marine deposits consisting of soluble salts have been completely removed from the geologic record through the slow dissolving action of percolating groundwater.

Salinity of Surface Waters

The salinity of surface waters is related to latitude (Fig. 8.6A). The most important factors affecting salinity are (1) evaporation (which removes water and leaves the remaining water saltier), (2) precipitation (which adds fresh water, thereby diluting the seawa-

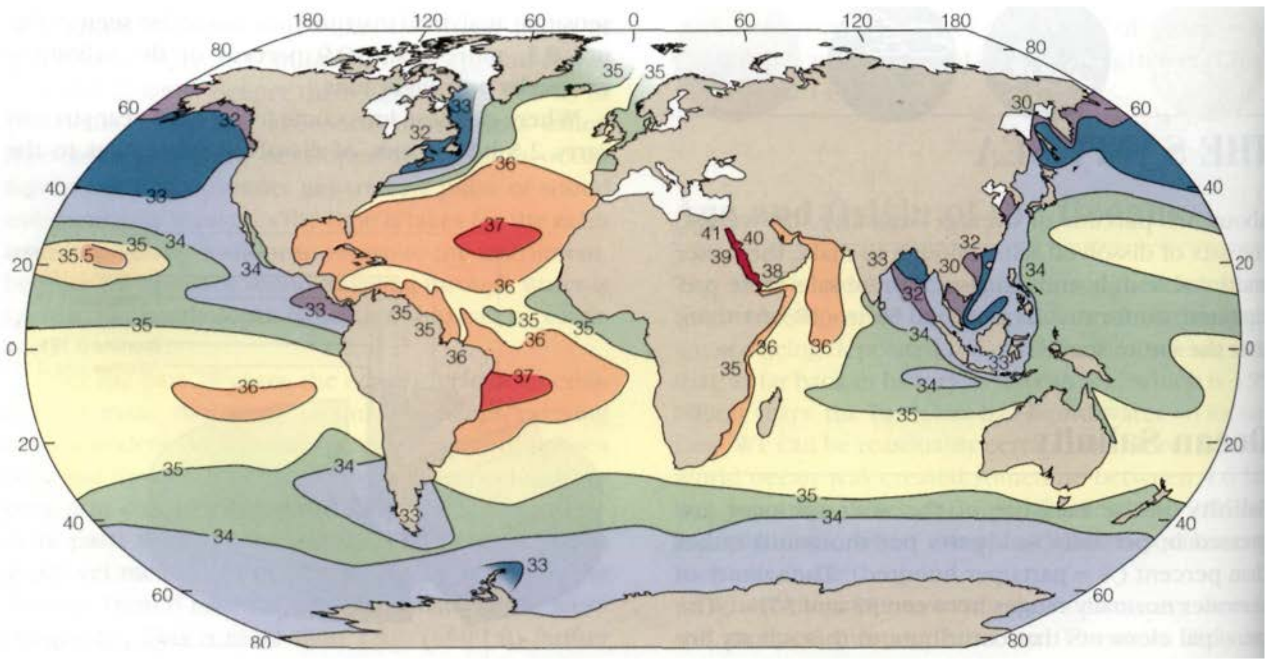


Figure 8.6A Average surface salinity of the oceans. High salinity values are found in tropical and subtropical waters where evaporation exceeds precipitation. The highest salinity has been measured in enclosed seas like the Persian Gulf, the Red Sea, and the Mediterranean Sea. Salinity values generally decrease poleward, both north and south of the equator, but low values also are found off the mouths of large rivers.

ter and making it less salty), (3) inflow of fresh (river) water (which makes the seawater less salty), and (4) the freezing and melting of sea ice (when seawater freezes, salts are excluded from the ice, leaving the unfrozen seawater saltier). As one might expect, salinity is high in the latitudes where the Earth's great deserts lie, for in these zones evaporation exceeds precipitation, both on land and at sea. In a restricted sea, like the Mediterranean, where these is little inflow of fresh water, surface salinity exceeds the normal range; in the Red Sea, which is surrounded by desert, salinity reaches 41‰. Salinity is lower near the equator because precipitation is high and cool water, which rises from the deep sea and sweeps westward in the tropical eastern Pacific and eastern Atlantic oceans, reduces evaporation. It is also low at high latitudes that are rainy and cool. Up to 100 km (62 mi) offshore from the mouths of large rivers, the surface ocean water can be fresh enough to drink.

Minerals from the Sea

Under favorable circumstances, evaporation of seawater can lead to the concentration of salts and their eventual precipitation from solution to form *evaporite deposits.* The most important salts that precipitate from seawater are gypsum, halite, and carnalite. Marine evaporites are widespread. In North America marine evaporite strata underlie an estimated 30 percent of the land area. We use the halite from these deposits for common table salt, gypsum is used for plaster, and carnalite and several other minerals provide potassium for use in commercial fertilizers.

TEMPERATURE AND HEAT CAPACITY OF THE OCEAN

An unsuspecting tourist from Florida who decides to take a swim on the northern coast of Britain quickly learns how varied the surface temperature of the ocean can be. A map of global summer sea-surface temperature displays a pronounced east-west banding, with *isotherms* (lines connecting points of equal temperature) approximately paralleling the equator (Fig. 8.6B). The warmest waters during August, ex-

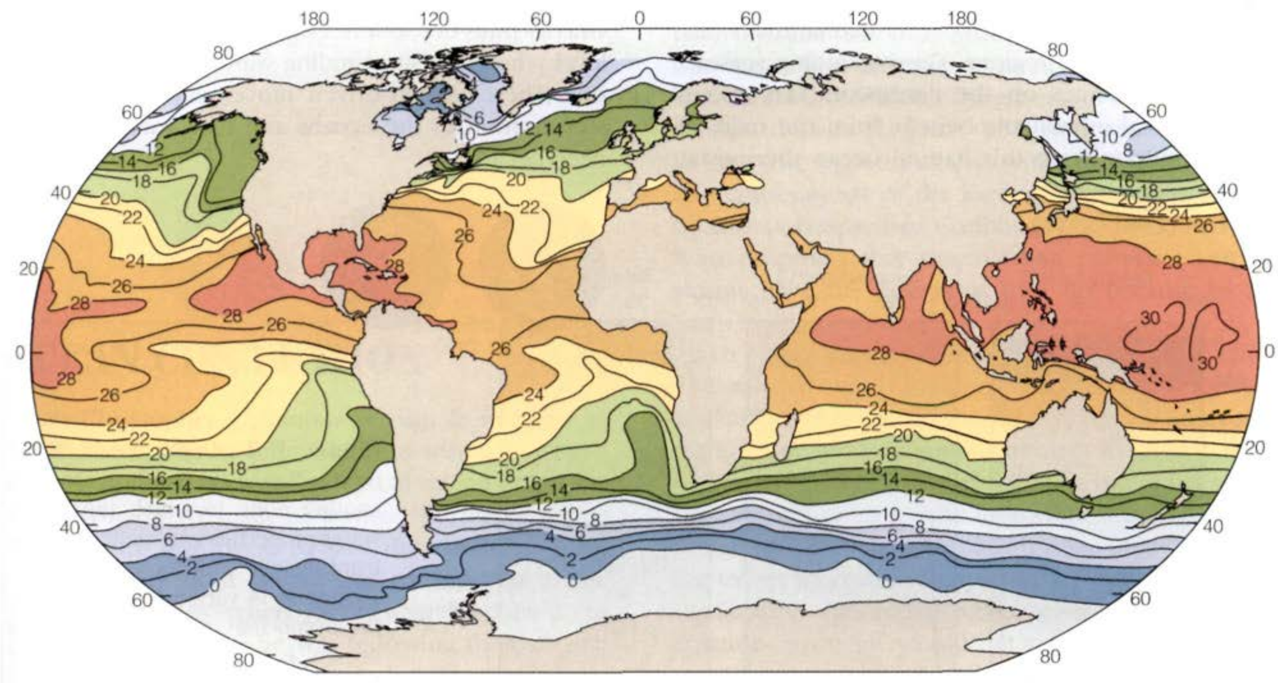


Figure 8.6B Sea-surface temperatures in the world ocean during August. The warmest temperatures (>28°C) are found in the tropical Indian and Pacific oceans. Temperatures decrease poleward from this zone, reaching values close to freezing in the north and south polar seas.

ceeding 28°C (82°F), occur in a discontinuous belt between about 30°N and 10°S latitude in the zone where received solar radiation is at a maximum. In winter, when the belt of maximum incoming solar radiation shifts southward, the belt of warm water also moves south until it is largely below the equator. Waters become progressively cooler both north and south of this belt, and reach temperatures of less than 10°C (50°F) poleward of 50°N and S latitude. The average surface temperature of the oceans is about 17°C (63°F), while the highest temperatures (>30°C or 86°F) have been recorded in restricted tropical seas, such as the Red Sea and the Persian Gulf.

The ocean differs from the land in the amount of heat it can store. For a given amount of heat absorbed, water has a lower rise in temperature than nearly all other substances; that is, it has a high **heat capacity.** Because of water's ability to absorb and release large amounts of heat with very little change in temperature, both the total range and the seasonal changes in ocean temperatures are much less than what we find on land. For example, the highest recorded land temperature is 58°C (136°F), measured in the Libyan Desert, and the lowest, measured at Vostok Station in central Antarctica, is —88°C (—126°F); the range, therefore, is 146°C. By contrast, the highest recorded ocean temperature is 36°C (97°F), measured in the Persian Gulf, and the coldest, measured in the polar seas, is -2°C (28°F), a range of only 38°C.

The annual change in sea-surface temperatures is 0-2°C in the tropics, 5-8°C in middle latitudes, and 2-4°C in the polar regions. Corresponding seasonal temperature ranges on the continents can exceed 50°C. Coastal inhabitants benefit from the mild climates resulting from this natural ocean thermostat. Along the Pacific coast of Washington and British Columbia, for example, winter air temperatures seldom drop to freezing, while east of the coastal mountain ranges they can plunge to — 30°C (—22°F) or lower. In the interior of a continent, summer temperatures may exceed 40°C (104°F), whereas along the ocean margin they typically remain below 25°C (77°F). Here, then, is a good example of the interaction of the hydrosphere, atmosphere, land surface, and biosphere: ocean temperatures affect the climate, both over the ocean and over the land, and climate ultimately is a major factor in controlling the distribution of plants and animals.

VERTICAL STRATIFICATION OF THE OCEANS

The physical properties of seawater vary with depth. To help understand why this is so, think about shaking up a bottle of oil-and-vinegar salad dressing and then set it on a table. After a minute or two, the oil will rise to the top of the bottle and the vinegar will settle to the bottom. The two ingredients become stratified because they have different densities: the less dense oil floats on the denser vinegar. The oceans also are vertically stratified, as a result of variations in the density of seawater. Seawater becomes denser as its temperature decreases and as its salinity increases. Gravity pulls dense water downward until it reaches a level where the surrounding water has the same density. These density-driven movements lead both to stratification of the oceans and to circulation in the deep ocean.

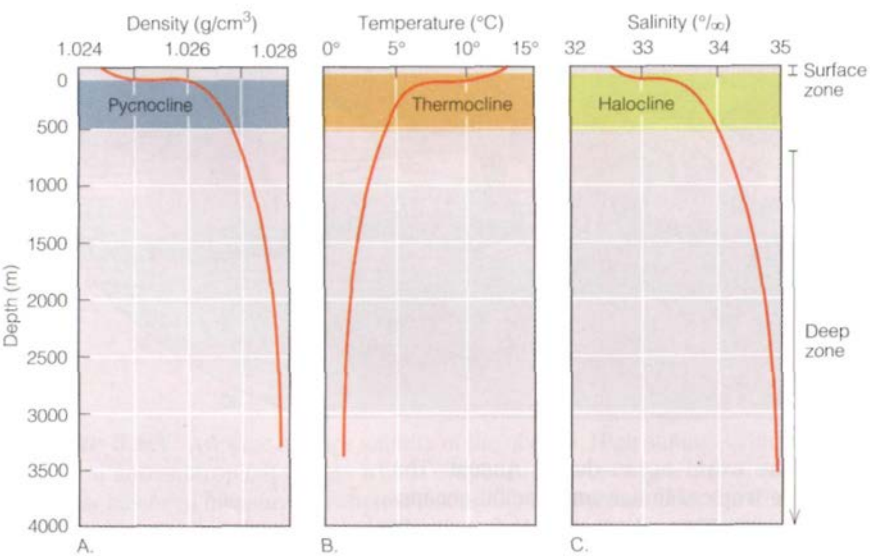


Figure 8.7 Depth zones in the ocean. Below the surface zone lies another zone in which the ocean-water properties experience a significant change with increasing depth. This zone is variously known as A. the *pycnocline*, a zone of increasing density. B. the *thermocline*, a zone of decreasing temperature, C. the *halo-dine*, a zone of increasing salinity. Still lower lies the *deep zone*, where waters are dense as a result of their low temperatures and high salinity.

Oceanographers recognize three major depth zones in the ocean (Fig. 8.7). A *surface zone,* typically extending to a depth of 100 to 500 m (330 to 1640 ft), consists of relatively warm water (except in polar latitudes, where the surface zone is absent). This zone is also referred to as the *mixed layer* because winds, waves, and temperature changes cause extensive mixing in it.

Below the surface zone lies another zone in which the ocean-water properties of temperature, salinity, and density experience a significant change with increasing depth. This zone goes by three different names, one for each property. In the open ocean, temperature commonly decreases markedly, then more slowly, downward through the *thermocline* (Fig. 8.7B). The *halocline,* marked by a substantial increase of salinity with depth, is found over much of the North Pacific Ocean and in other high-latitude waters where solar heating of the ocean surface is diminished and precipitation is relatively high (Fig. 8.7C). The rapid increase in water density that defines the *pycnocline* (Fig. 8.7A) may result from a decrease in temperature, from an increase in salinity, or from both.

Below the zone that encompasses the thermocline, halocline, and pycnocline lies the *deep zone,* which contains the bulk of the ocean's volume (about 80 %$_0$; Fig. 8.4). In low and middle latitudes, the pycnocline effectively isolates water of the deep zone from the atmosphere, but in high latitudes, where the surface zone is absent, water of the deep zone lies in direct contact with the atmosphere.

OCEAN CIRCULATION

When Christopher Columbus set sail from Spain in 1492 to cross the Atlantic Ocean in search of China, he took an indirect route. Instead of sailing due west, which would have made his voyage shorter, he took a longer route southwest toward the Canary Islands, and then west on a course that carried him to the Caribbean islands where he first sighted land. In choosing this course, he was following the path not only of the prevailing winds but also of surface ocean currents. Instead of fighting the westerly winds and currents at 40° N latitude, he drifted with the Canary Current and North Equatorial Current, as the northeast tradewinds filled the sails of his three small ships.

Surface Currents of the Open Ocean

Surface ocean currents, like those Columbus followed, are broad, slow drifts of surface water set in motion by the prevailing surface winds. Air that flows across the sea drags the water slowly forward, creating a current of water as broad as the current of air, but rarely more than 50 to 100 m (165 to 330 ft) deep. The ultimate source of this motion is the Sun, which heats the Earth unequally, thereby setting in motion the planetary wind system. Thus, ocean circulation results from the interplay of several key elements of the Earth system: (1) radiation from the Sun provides heat energy to the atmosphere; (2) nonuniform heating generates winds; and (3) the winds, in turn, drive the motion of the ocean's surface water.

The Coriolis Effect

The direction taken by ocean currents is also influenced by the **Coriolis effect,** a phenomenon by which all moving bodies veer to the right in the northern hemisphere and to the left in the southern hemisphere. The effect is named for the nineteenth-century French scientist who first explained how the Earth's rotation influenced the movement of fluids.

A simple experiment and an insightful hypothesis led to the discovery of the Coriolis effect. The experiment was carried out in 1627 by Joseph Furtenbach, a German scientist, who fired a cannonball vertically into the air and then sat on the muzzle of the cannon and watched the ball hit the ground to the west of the cannon, thus proving that the Earth rotates from west to east. The hypothesis that winds are also deflected as a consequence of the Earth's rotation was offered by Gustav Gaspard de Coriolis (1792-1843) in 1835, who proposed that the swirling pattern taken by storms have the directions they do because of the Earth's rotation.

An object on a rotating body has an angular velocity (velocity due to rotation). The angular velocity is always in the direction of rotation and is a minimum at the equator and a maximum at the poles. It may seem odd to say that angular velocity is a minimum at the equator, but consider the following: Imagine a stone tower built exactly at the north pole; every 24 hours the tower will rotate completely around as a result of the Earth's rotation (Fig. 8.8A). A similar tower on the equator, however, would not rotate at all; rather, it would describe an end-over-end motion. At any latitude between the equator and the pole, some rotation and some end-over-end motion occurs and for this reason the Coriolis effect, which is due to rotation, is latitude-dependent and a maximum at the poles.

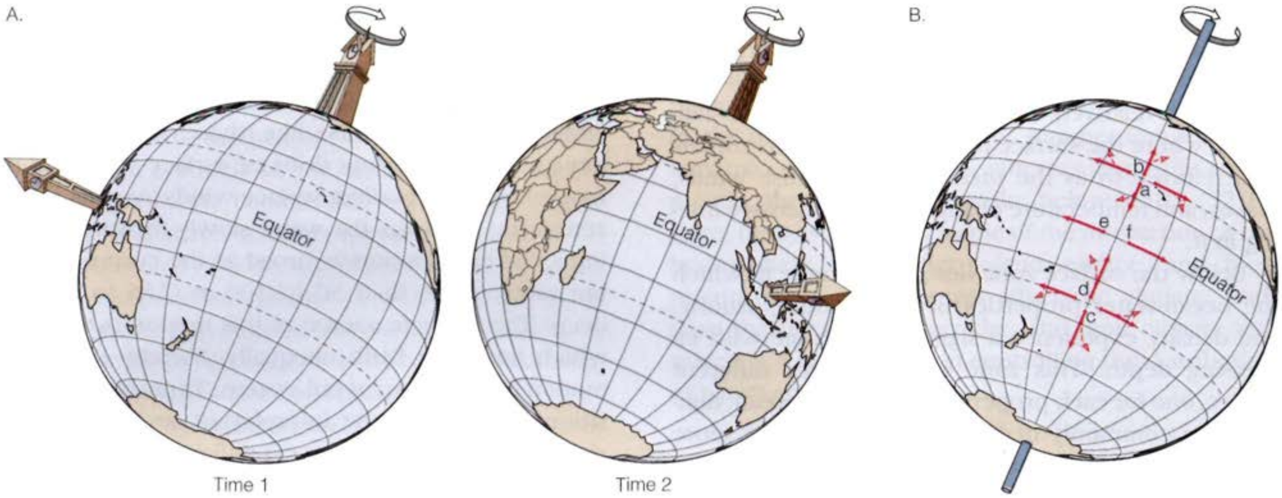


Figure 8.8 Coriolis Effect A. A body at the pole rotates completely around every 24 hours while a body on the equator goes end-over-end but does not rotate. The face on the tower at the pole rotates with respect to an external observer whereas a tower on the equator always presents the same face to an observer. B. On the rotating Earth, an object freely floating on the ocean in the northern hemisphere (*a, b*) is deflected by the Coriolis effect to the right, whereas in the southern hemisphere (*c, d*) it is deflected to the left. A moving object at the equator (*e, f*) is not deflected.

A body that moves on the Earth has two velocity components—the velocity of forward motion and the angular velocity of rotation. In order for a moving body to maintain the same velocity as its location on the Earth, its angular velocity would have to change continually, A change in velocity is an acceleration. The *Coriolis acceleration* is the angular acceleration that would be needed for a moving object to stay on track with respect to the rotating frame of reference, in this case the Earth.

Because the angular acceleration is usually absent or insufficient, the *Coriolis effect* occurs and this, as mentioned above, is a deflection in the path of a moving object toward the right in the northern hemisphere and to the left in the southern hemisphere.

Every moving body is subject to the Coriolis effect. An automobile driven at 100 km/h (62 mi/h) in the northern hemisphere will drift about 3 m (10 ft) to the right every km (0.6 mi) for example. You don't actually see the drift because of the friction of the tires on the road and the hands of the drive using the steering wheel to correct the drift. One of the most dramatic demonstrations of the Coriolis effect happened during World War I when the German army bombarded Paris from a distance of 120 km (75 mi) with a gun called "Big Bertha." The gunners discovered that their shots were falling 1 to 2 km (0.6 to 1.2 mi) to the right of the place they were aiming. The reason for their poor shooting was their failure to account for a deflection in the trajectory due to the Coriolis effect.

The magnitude of the Coriolis effect varies with latitude and with the speed of the moving body. The latitude effect arises from the variation in angular velocity which, as was discussed previously, is a maximum at the poles and a minimum at the equator. The Coriolis effect is therefore a maximum at the poles and zero at the equator.

If a freely floating object on the ocean moves away from the pole (Fig. 8.8B, *a*), its angular velocity about the pole will be slower than that of the water. This causes the object to lag behind the rotation, and so it will be deflected in a clockwise direction (to the right). If such an object were moving toward the pole (*b*), its angular velocity about the pole will be faster than the water, resulting in a counterclockwise deflection (again toward the right). Regardless of the direction of movement, an object in the northern hemisphere will be deflected to the right. In the southern hemisphere, the deflection is to the left (*c,d*), while at the equator the effect disappears (*e, f*). Although the Coriolis effect does not cause ocean currents, it deflects them once they are in motion.

Current Systems

Low-latitude regions in the tradewind belts are dominated by the warm, westward-flowing North and South Equatorial currents (Fig. 8.9). In their midst, and lying in the doldrums belt of light, variable winds, is the eastward-flowing Equatorial Countercurrent.

Each major ocean current is part of a large subcircular current system called a **gyre.** Within each gyre different names are used for different segments of the current system. Figure 8.9 shows the Earth's five major ocean gyres, two each in the Pacific and Atlantic oceans and one in the Indian Ocean. Currents in the northern hemisphere gyres circulate in a clockwise direction; those in the southern hemisphere circulate counterclockwise.

In each major ocean basin, westward-flowing equatorial currents are deflected poleward as they encounter land. Each current thereby is transformed into a *western boundary current* that flows generally poleward, parallel to a continental coastline. In the North Atlantic Ocean this current is called the Gulf Stream, while in the North Pacific it is the Kuroshio Current. In the South Atlantic the Brazil Current follows the South American coast, while in the Pacific and Indian oceans the corresponding currents are the East Australian Current and the Mozambique Current.

On reaching the belts of westerly winds, the Kuroshio Current changes direction to form the North Pacific Currents on the poleward side of the North Pacific gyre, while in the Atlantic, the Gulf Stream passes eastward into the northeast-flowing North Atlantic Current. In the Southern Hemisphere, the poleward moving waters of the Brazil, East Australian, and Mozambique currents enter the Antarctic Circumpolar Current that circles the Earth near latitude 60°S.

At the southeastern ends of the Southern Hemisphere gyres, cool southern waters move northward along western continental coasts forming the West Australian Current in the eastern Indian Ocean, the Humbolt Current along the southwestern coast of South America, and the 16 current off the southwestern coast of Africa.

The northern Indian Ocean exhibits a unique circulation pattern in which the direction of flow changes seasonally with the monsoons (Ch. 13). During the summer, strong and persistent monsoon winds blow the surface water eastward, whereas in winter, winds from Asia blow the water westward.

Ekman Transport

In 1893, Norwegian explorer Fridtjof Nansen (1861-1930) began an epic voyage across the frozen Arctic Ocean in his now-famous vessel, the *Fram.* Frozen in the shifting pack ice, the ship slowly drifted poleward, then southward again, eventually emerging into navigable water nearly three years later. Nansen's observations disclosed something totally unexpected: the floating pack ice moved in a direction 20-40° to the right of the prevailing wind. This phenomenon was subsequently explained mathematically by V. W. Ekman, who postulated that wind blowing across the

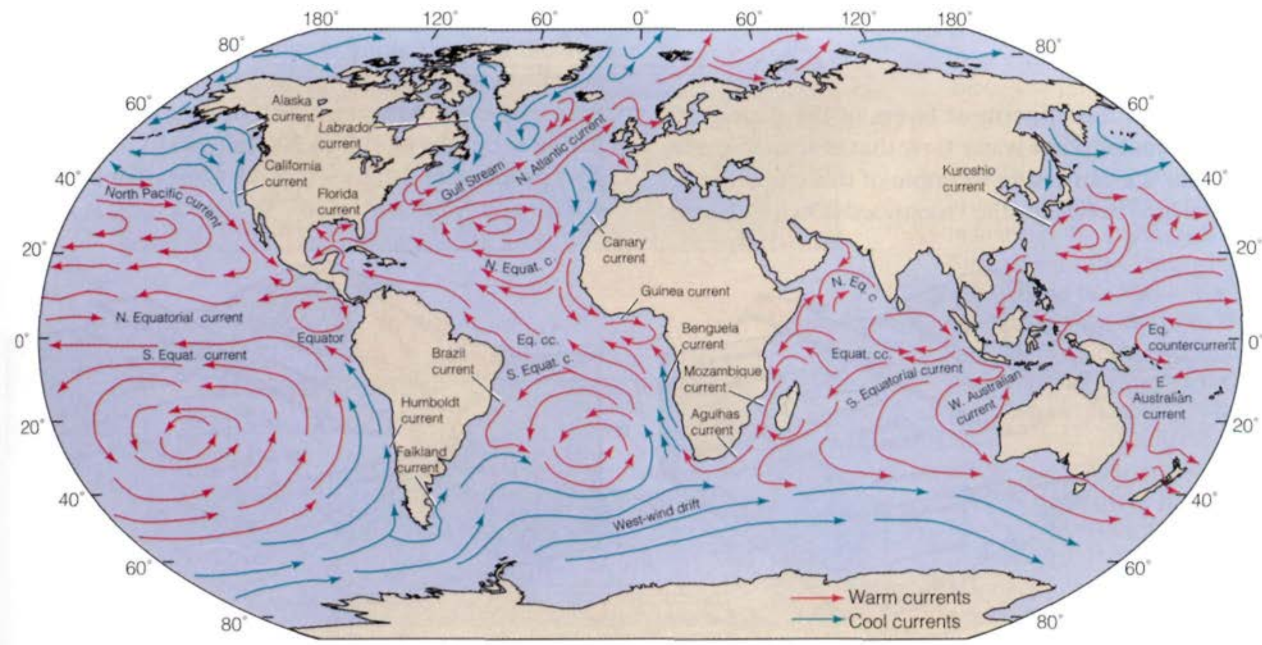


Figure 8.9 Surface ocean currents form a distinctive pattern, curving to the right (clockwise) in the northern hemisphere and to the left (counterclockwise) in the southern hemisphere. The westward flow of tropical Atlantic and Pacific waters is interrupted by continents, which deflect the water poleward. The flow then turns away from the poles and becomes the eastward-moving currents that define the middle-latitude margins of the five great midocean gyres.

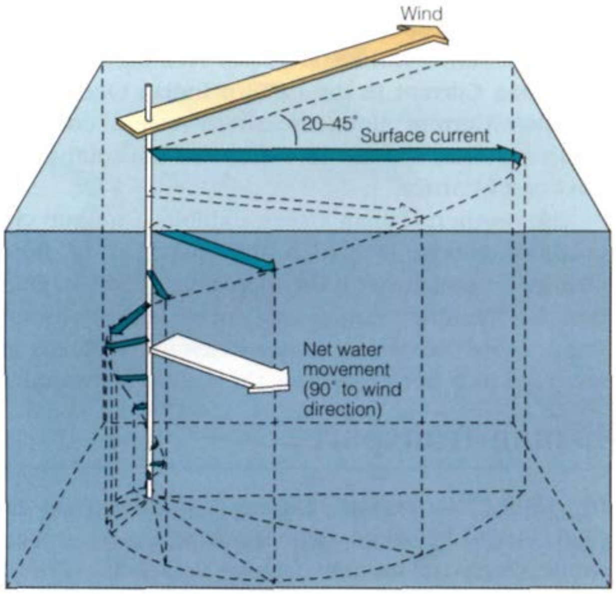


Figure 8.10 Wind blowing across the ocean in the northern hemisphere affects the surface water, producing a net flow 20-45° to the right of the wind direction. The surface water drags on the water immediately beneath, setting it in motion, and so on down the water column. Internal friction steadily reduces the current velocity with depth, and the Coriolis effect shifts each successively slower moving layer farther to the right, thereby producing an *Ekman spiral*. The average flow over the full depth of the spiral is termed the *Ekman transport* and is directed at 90° to the wind direction.

ocean affects the uppermost layers of the water column, producing a net water flow that is at an angle to the wind. We can see an example of this effect in the relationship between the westward-flowing North Equatorial Current and the northeast tradewinds. The surface water layer dragging on the water immediately beneath sets the lower layer in motion, and the process continues downward. Internal friction, however, causes a decrease in current velocity with increasing depth. In addition, the Coriolis effect shifts each successive, slower moving layer farther to the right, producing a spiraling current pattern (called the **Ekman spiral)** when seen from above (Fig. 8.10) At a depth of about 100 m (330 ft), the current has reversed 180° from the direction of the surface wind, and its speed has dropped to only a small percentage of the surface current velocity. The average flow over the full depth of the spiral, called the **Ekman transport**, moves at 90° to the wind direction.

Upwelling and Downwelling

Near coasts, Ekman transport can lead to vertical movement of ocean water. Winds blowing parallel to the coast can drag a layer of surface water tens of meters thick toward or away from land, depending on wind direction (Fig. 8.11). If the net transport is away from land, subsurface waters flow upward and replace the water moving away, a process called **upwelling.** If the net Ekman transport is toward the coast, the surface water thickens and sinks in a process known as **downwelling.** Important areas of upwelling occur along west-facing low-latitude continental coasts (e.g., Oregon/California and Ecuador/Peru), where cold waters, rich in nutrients and originating at depths of 100 to 200 m (330 to 655 ft), support productive fisheries (see "A Closer Look: Understanding El Niño").

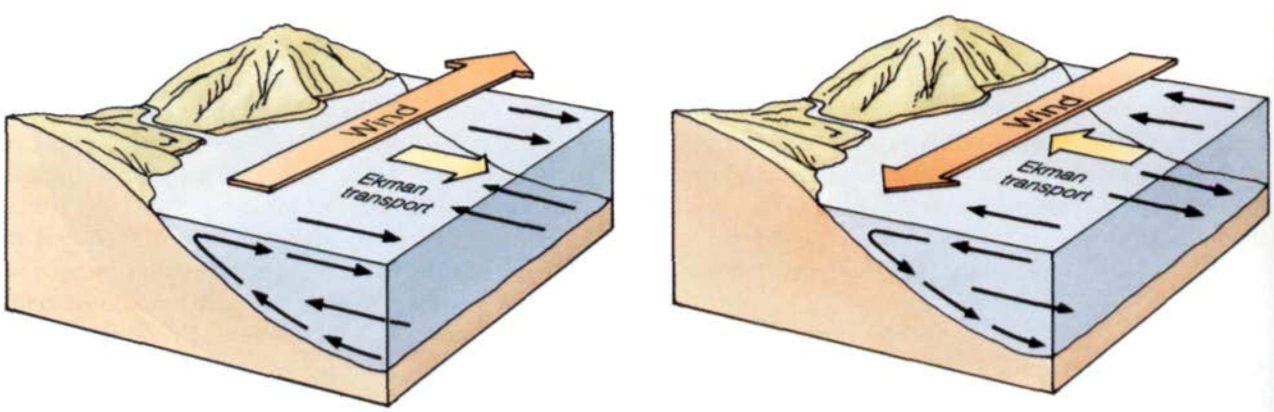


Figure 8.11 Winds blowing parallel to a coast exert a drag on the surface water, forcing it away from or toward the land, depending on wind direction. If the net Ekman transport is away from the land, rising subsurface water replaces water moving offshore, producing upwelling. If the net transport is toward the shore, the surface water thickens and sinks, producing downwelling.

A Closer Look

Understanding El Niño

The fishing grounds off the coast of Peru, among the richest in the world, are sustained by upwelling cold waters filled with nutrients. Periodically, a mass of unusually warm water appears off the coast, an event that Peruvians refer to as *El Niño.* During El Niño years, the tradewinds slacken, upwelling is markedly reduced, and the fish population declines, accompanied by a great die-off of the coastal bird population, which depends on the fish for food. The Peruvian fishery is among the most important in the world, and so the occurrence of an El Niño event constitutes a local economic catastrophe. Coincident with the Peruvian El Niño conditions, very heavy rains fall in normally arid parts of Peru and Ecuador, Australia experiences drought conditions, anomalous cyclones appear in Hawaii and French Polynesia, the seasonal rains of northeast Brazil are disrupted, and the Indian monsoon may fail to appear. During exceptional El Niño years, climates over much of Africa, eastern Asia, and North America are affected. In North America, unusually cold or mild winters can result in the northeastern United States, while the Southeast becomes wetter; in California abnormally high rainfall can produce major flooding and widespread landsliding.

Although El Niño has been experienced by generations of Peruvians, with occurrences often being recorded in ship's logs, its broader significance was recognized only recently. In the late 1960s a link was made between cyclic El Niño events and changing atmospheric pressure anomalies over the equator, anomalies that had earlier been referred to as the *Southern Oscillation.* Today, the *El Niño/Southern Oscillation (ENSO)* is regarded as an extremely important element in the Earth's year-to-year variations in climate. We now recognize that El Niño recurs erratically, but on average about every four years, and that its effect on climates is felt over at least half the Earth. It presents us with an especially instructive example of the close interaction between the Earth's atmosphere, hydrosphere, and biosphere. When

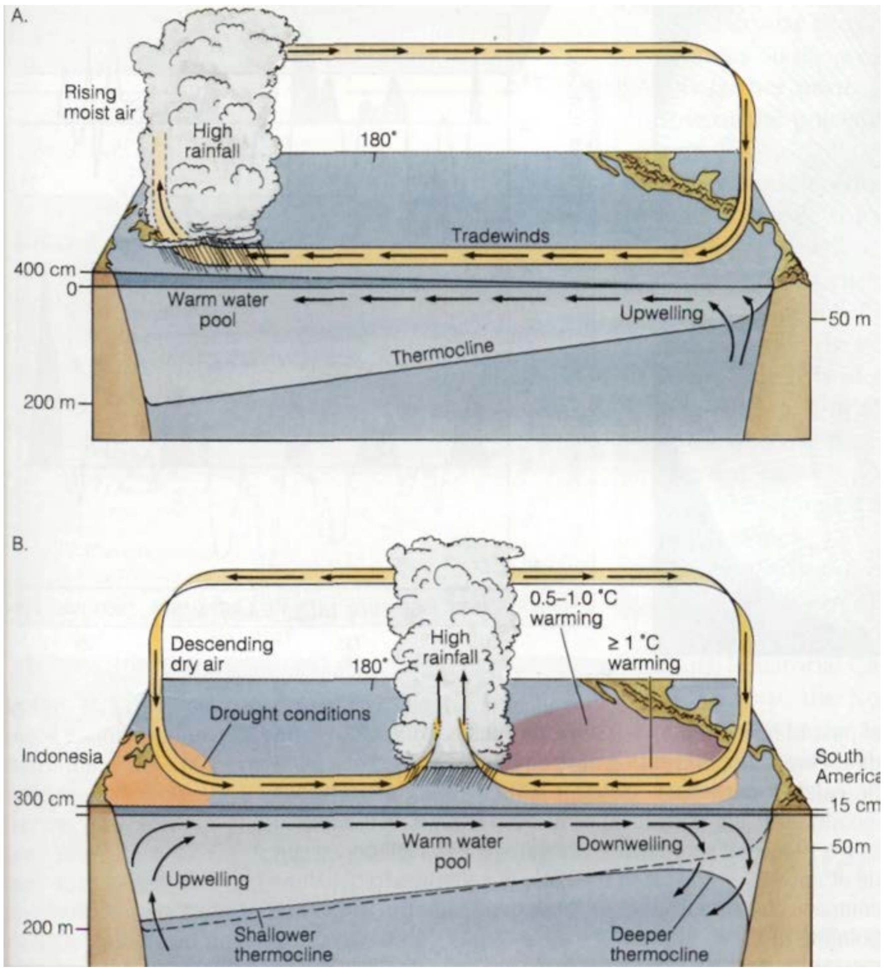


Figure C8.1 El Niño cycle. A. During normal years, persistent tradewinds blow westward across the tropical Pacific from a zone of upwelling water off the coast of Peru. This water warms up as it is transported westward to form a large warm-water pool above the thermocline in the western Pacific. The warm water causes the moist maritime air to rise and cool, bringing abundant rainfall to Indonesia. B. During an El Niño event, the tradewinds slacken and the pool of warm water moves eastward to the central Pacific. Descending cool, dry air brings drought conditions to Indonesia, while rising moist air above the warm-water pool greatly increases rainfall in the mid-Pacific. Surface waters in the eastern Pacific become warmer, and downwelling shuts off the supply of deep-water nutrients, adversely affecting the normally productive fishing ground off the coast of Peru.

an El Niño event occurs, it not only involves the tropical oceans and atmosphere, but also it directly affects precipitation and temperature on major land areas, thereby also impacting plants and animals.

Although many details of the El Niño phenomenon remain under study, the general mechanism is reasonably well understood. During normal years, the tradewinds blowing across the Pacific pile up a large pool of warm water in the west, which contrasts with cooler water that wells up in the eastern tropical Pacific (Fig. C8.1 A). The warm water promotes a large center of heavy rainfall around Indonesia. An El Niño event begins with a slackening of the tradewinds. This leads to expansive warming of surface waters in the central and eastern Pacific as the water that had accumulated in the western Pacific pool sloshes back in the direction of South America. The eastward movement of warm water causes the zone of high rainfall to shift to the central Pacific near the international date line, simultaneously bringing drought conditions to Indonesia (Fig. C8.1B). At the peak of an event, equatorial surface water moves from west to east and also poleward. This flow gradually reduces the

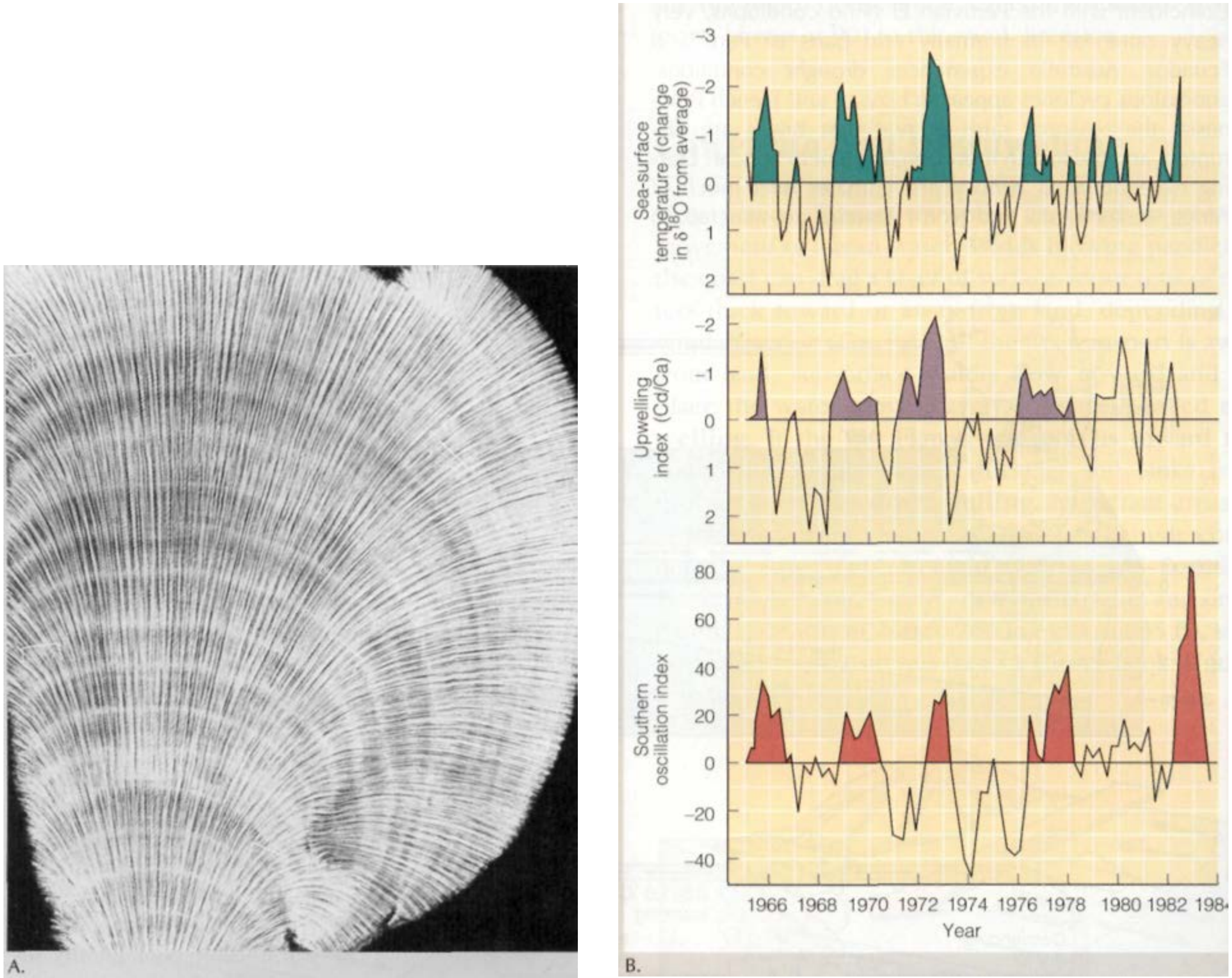


Figure C8.2 Geologic records of past El Niño events. A. A slice through a living coral from the Galapagos islands shows the annual layering (alternating dark and light bands rising from bottom to top of the section) of the calcium-carbonate skeleton. This layering preserves a record of changing surface water conditions, and therefore, of El Niño events. B. Chemical analyses of Galapagos corals produce a record of sea-surface temperature (based on oxygen isotopes) and the relative magnitude of upwelling (based on the ratio of cadmium to calcium) spanning 20 years. These records compare closely with variations of the Southern Oscillation and show that the corals were responding to ENSO cycles.

equatorial pool of warm water, leading to intensification of the tradewinds and an eventual return to normal conditions.

In an attempt to extend the detailed record of El Niño events further back in time and improve our ability to predict future occurrences, scientists have examined the growth rings of living corals, because the rings record annual variations in seawater conditions (Fig. C8.2A). Corals are abundant and widespread throughout the region most strongly affected by El Niño, and individual colonies can live as long as 800 years. Their skeletal chemistry closely reflects surrounding environmental conditions, with the isotope or trace-metal composition of new skeletal material changing as water temperature and water chemistry change. By measuring the chemical composition of annual growth layers, a record of historic and prehistoric El Niño events can be reconstructed for different oceanic sites (Fig. C8.2B) and the dynamics of each cycle can be analyzed. Ultimately, such data may make it possible to predict future El Niño events with considerable confidence.

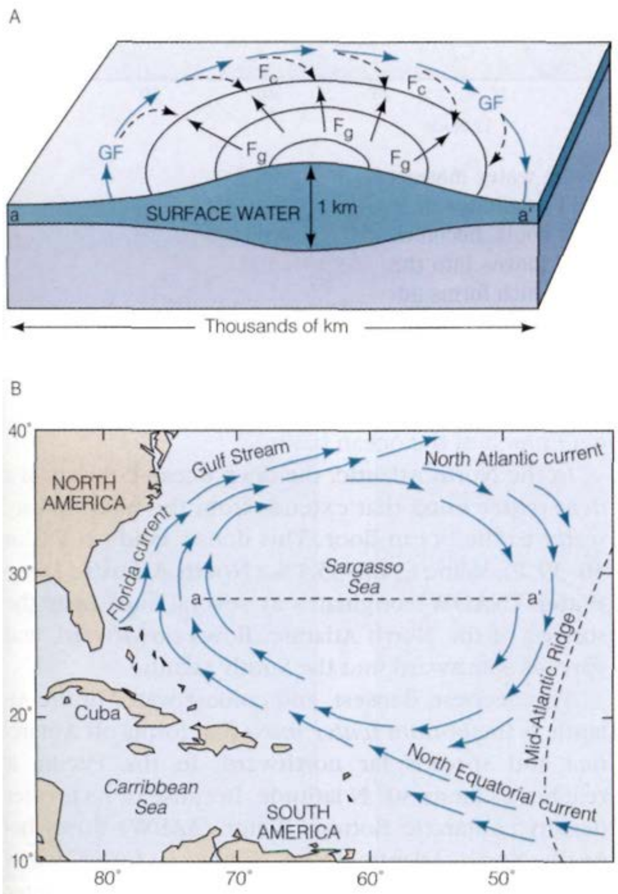


Figure 8.12 Geostrophic flow. A. The Coriolis effect causes major wind-driven surface currents to be deflected toward the middle of a gyre where water piles up to form a gentle mound above the average level of the ocean. In this diagram, the vertical scale is greatly exaggerated. The water piles up until the force of gravity (Fg) pulling the water downslope just balances the Coriolis effect (Fc). The net result is *geostrophic flow* (GF) around the gyre. B. Geostrophic flow in the subtropical North Atlantic traps a broad lens of clear water a kilometer deep, forming the Sargasso Sea.

Geostrophic Flow

The surface of the ocean is not absolutely flat. The persistent tradewinds blowing westward through the tropical Pacific pile up surface water on the western side of the ocean basin. Just as water runs down the slope of a hill on land, seawater will flow downslope where the ocean surface is unusually steep. Therefore, because the sea surface is slightly higher at low latitudes on the western side of the Pacific basin than it is farther north, the resulting slope enhances the flow of the poleward-moving western boundary current.

The Coriolis effect causes wind-driven surface currents to move toward the middle of the five major ocean gyres, where the water piles up to form a gentle mound that rises more than a meter above the average surface level of the ocean. In a gyre, water will pile up until the force of gravity pulling the water downslope balances the Coriolis effect that deflects it (Fig. 8.12A). When this occurs, no further deflection of the surface water takes place and a **geostrophic current** flows smoothly around the gyre. Such currents are strongest along steep slopes and weakest on gentle slopes.

The Sargasso Sea provides an example of geostrophic flow (Fig. 8.12B). This body of water in the western North Atlantic gyre is bounded by the North Equatorial Current to the south, the Gulf Stream to the west, the North Atlantic Current to the north, and the Mid-Atlantic ridge on the east. A lens of clear water a kilometer deep is trapped in the middle of the gyre by geostrophic flow.

Major Water Masses

The water of the oceans is organized vertically into major water masses, stratified according to density. The identity and sources of these masses have been

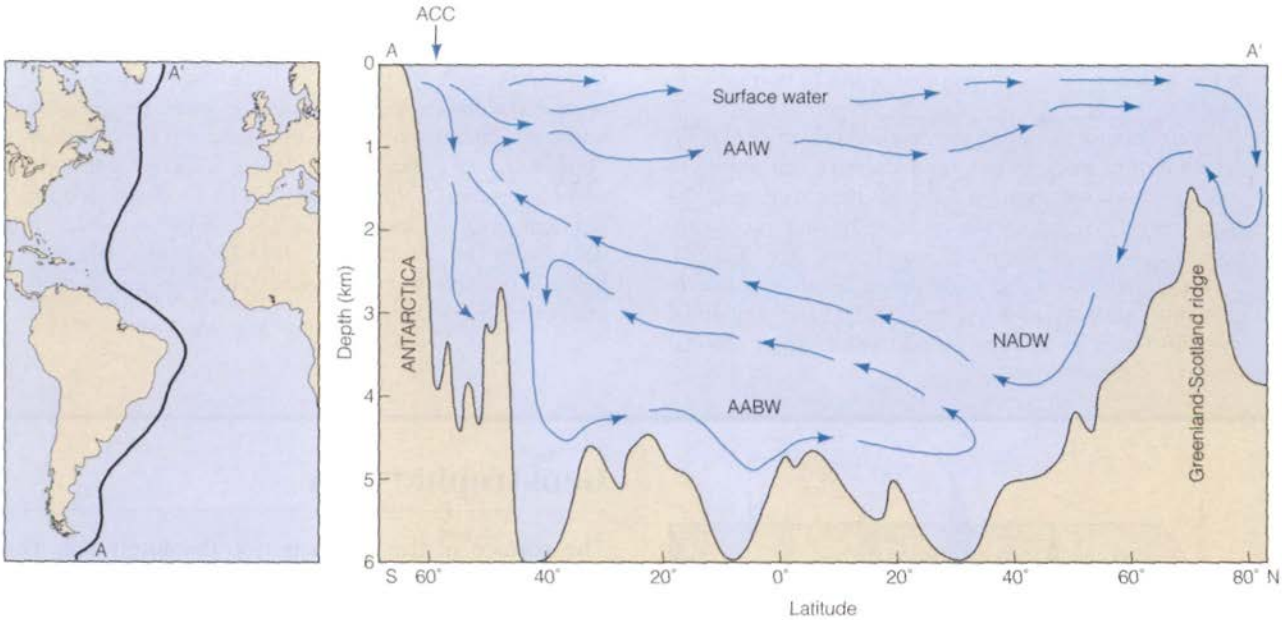


Figure 8.13A Transect along the western Atlantic Ocean showing water masses and general circulation pattern. North Atlantic Deep Water (NADW) originates near the surface in the North Atlantic as northward-flowing surface water cools, becomes increasingly saline, and plunges to depths of several km. As NADW moves into the South Atlantic, it rises over denser Antarctic Bottom Water (AABW), which forms adjacent to the Antarctic continent and flows into the North Atlantic as Antarctic Intermediate Water (AAIW) at a mean depth of about 1 km.

determined by studying the salinity and temperature structure of the water column at many places. The Atlantic Ocean provides a good example (Fig. 8.13A).

In the Atlantic, water in the surface zone forms a *central water mass* north and south of the equator to about 35° latitude. The temperature of this water typically ranges from 6 to 19°C (43 to 66°F), and the salinity ranges from 34 to 36.5‰. Cooler subarctic and subantarctic surface water masses are found at high latitudes where cool temperatures and high rainfall give rise to colder, less saline waters. The largest polar surface water mass, flowing as the Antarctic Circumpolar Current (ACC), moves clockwise around Antarctica. Its temperature is 0-2°C (32-36°F), and its salinity 34.6-34.7‰.

The central water mass of the Atlantic overlies an *intermediate water mass* that extends to a depth of about 1500 m (1640 yd). The Antarctic Intermediate Water mass (AAIW), the most extensive such body of water, originates as cold subantarctic surface water that sinks and spreads northward across the equator to about 20° N latitude. Its temperature ranges from 3 to 7°C (37-45°F), and its salinity lies within the range 33.8-34.7‰. Water entering the Atlantic from the Mediterranean Sea is so saline (37-38‰) that it flows over a shallow sill at Gibraltar (-400 m) and downward beneath intermediate water to spread laterally over much of the ocean basin.

In the North Atlantic, the deep ocean consists of a *deep-water mass* that extends from the intermediate water to the ocean floor. This dense, cold (2-4°C or 36-39°F), saline (34.8-35.1‰) **North Atlantic Deep Water (NADW)** originates at several sites near the surface of the North Atlantic, flows downward, and spreads southward into the South Atlantic.

The deepest, densest, and coldest water in the Atlantic is the *bottom water mass* that forms off Antarctica and spreads far northward; in the Pacific it reaches as far as 30° N latitude. Because of its greater density, Antarctic Bottom Water (AABW) flows beneath North Atlantic Deep Water. It forms when dense brine, produced during the formation of winter sea ice in the Weddell Sea adjacent to Antarctica, mixes with cold circumpolar surface water and sinks into the deep ocean. This dense water has an average temperature of -0.4°C (31°F) and a salinity of 34.7‰.

The Global Ocean Conveyor System

The sinking of dense cold and (or) saline surface waters provides a link between the atmosphere and the deep ocean. It also propels a global **thermohaline**

circulation system, so called because it involves both the temperature and salinity characteristics of the ocean waters. We can trace this circulation from the North Atlantic southward toward Antarctica and into the other ocean basins (Fig 8.13B). The largest mass of NADW forms in the Greenland and Norwegian seas where relatively warm and salty surface water entering from the western North Atlantic cools, becomes denser, and sinks into a confined basin north of a submarine ridge connecting Scotland and Greenland (Fig. 8.14). The dense water then spills over low places along the ridge and plunges down into the deep ocean as NADW. Warm, salty surface and intermediate water is drawn toward the North Atlantic to compensate for the south-flowing deep water. It is the heat lost to the atmosphere by this warm surface water, together with heat from the warm Gulf Stream, that maintains a relatively mild climate in northwest Europe.

In the South Atlantic, south-flowing NADW enters the Antarctic Circumpolar Current, which travels clockwise around Antarctica. Surface and intermediate water flowing into the South Atlantic from the Pacific, and from the Indian Ocean via the southern tip of Africa, replenishes the NADW moving out of the basin. Meanwhile, Antarctic Bottom Water plunging down to the ocean floor in the southernmost Atlantic moves northward, slowly upwells, mixes with overly-

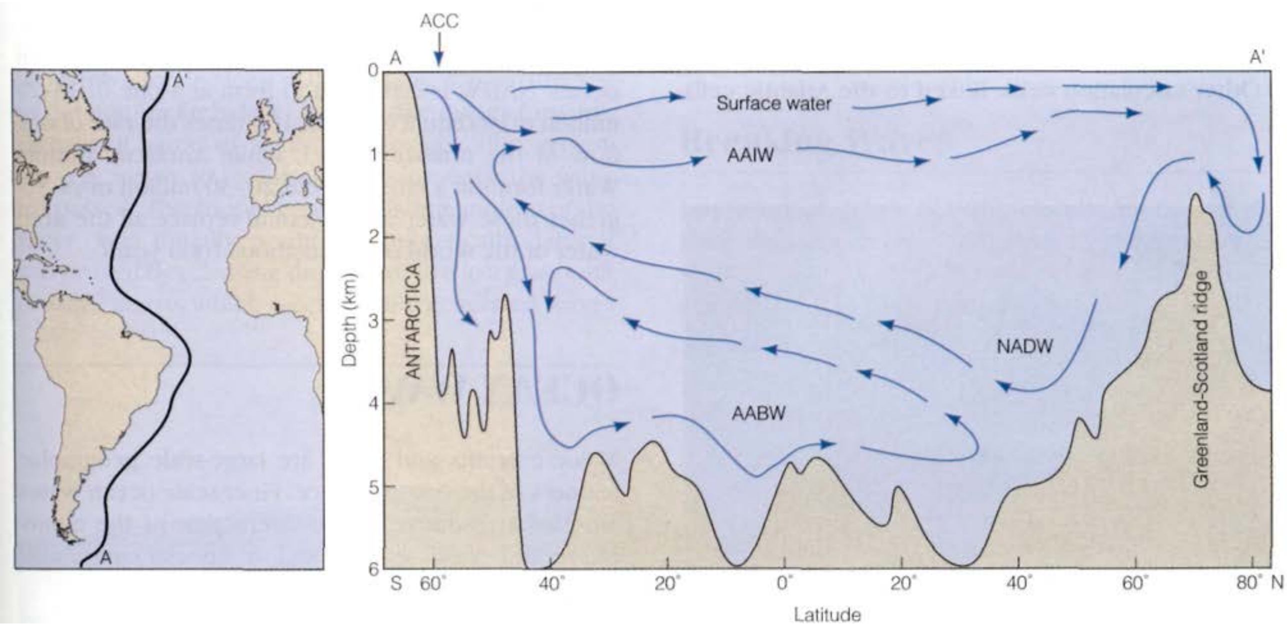


Figure 8.13B The major thermohaline circulation cells that make up the global ocean conveyor system are driven by exchange of heat and moisture between the atmosphere and ocean. Dense water forming at a number of sites in the North Atlantic spreads slowly along the ocean floor, eventually to enter both the Indian and Pacific oceans before slowly upwelling and entering shallower parts of the thermohaline circulation cells. Antarctic Bottom Water (AABW) forms adjacent to Antarctica and flows northward in fresher, colder circulation cells beneath warmer, more saline waters in the South Atlantic and South Pacific. It also flows along the Southern Ocean beneath the Antarctic Circumpolar Current to enter the southern Indian Ocean. Warm surface waters flowing into the western Atlantic and Pacific basins close the great global thermohaline cells.

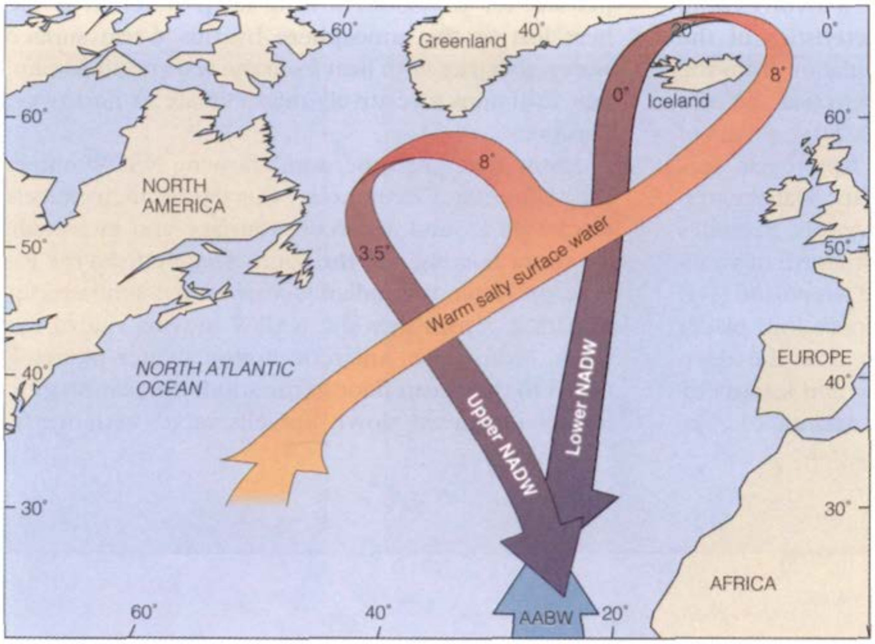


Figure 8.14 North Atlantic Deep Water (NADW) forms when the warm, salty water of the Gulf Stream/North Atlantic Current cools, becomes increasingly saline due to evaporation, and plunges downward to the ocean floor. The densest water then spills over the Greenland-Scotland ridge and flows southward as lower NADW. Less dense water forming between Greenland and North America moves south and east as upper NADW, overriding the lower, denser water. Because both water masses are less dense than northward-flowing Antarctic Bottom Water (AABW), they pass over it on their southward journey (see also Fig. 10.14B).

ing NADW, and flows back toward Antarctica as Circumpolar Deepwater (Fig. 8.13A). Thereby completing an important segment of the global system of ocean circulation, the Atlantic thermohaline circulation cell acts like a great conveyor belt.

Other circulation cells, linked to the Atlantic cells via the Antarctic Circumpolar Current and also driven by density contrasts related to temperature and salinity, exist in the Pacific and Indian oceans. In concert, they move water along the global ocean conveyor system, slowly replenishing the waters of the deep ocean. NADW is estimated to form at a rate of 15-20 million m^3/s (equal to about 100 times the rate of outflow of the Amazon River), while Antarctic Bottom Water forms at a rate of about 20-30 million m^3/s. Together these water masses could replace all the deep water of the world ocean in about 1000 years.

Figure 8.15 A ship in the open ocean struggles to maintain course through huge storm waves that tower over its deck.

OCEAN WAVES

Major currents and gyres are large-scale geographic features of the ocean surface. Finer-scale ocean waves are also a response to the interaction of the atmosphere and ocean surface and, in special cases, a response to movements of the solid earth.

Surface waves on the oceans receive their energy from winds that blow across the water surface. The size of a wave depends on how fast, how far, and how long the wind blows. A gentle breeze blowing across a bay may ripple the water or form low waves less than a meter high. At the opposite extreme, storm waves produced by hurricane-force winds (>115 km/h, or 72 mi/h) blowing for days across hundreds or thousands of kilometers of open water may become so high that they tower over ships unfortunate enough to be caught in them (Fig. 8.15).

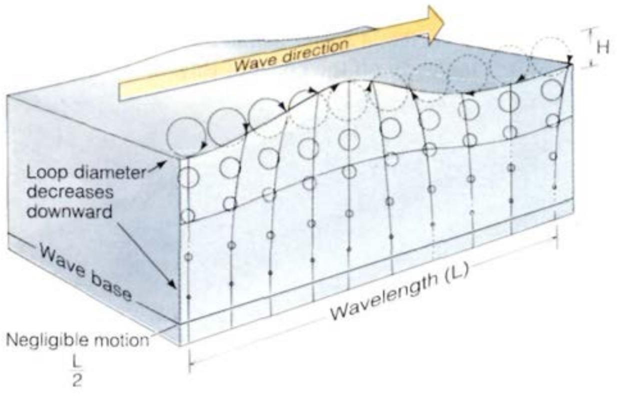


Figure 8.16 Looplike motion of water in a wave in deep water. To trace the motion of a small parcel of water at the surface, follow the arrows in the largest loops from right to left. The resultant motion is the same as watching the wave crest travel from left to right. Parcels of water in smaller loops beneath the surface have corresponding positions, marked by nearly vertical lines. Dashed lines represent wave form and parcel positions one-eighth of a period later.

Wave Motion

Figure 8.16 shows the significant dimensions of a wave traveling in deep water, where it is unaffected by the bottom far below. As the wave moves forward, each small parcel of water revolves in a loop, returning very nearly to its former position once the wave has passed. This looplike, or oscillating, motion of the water, was initially predicted theoretically; later it was proved by releasing droplets of dye in a glass tank of water across which mechanically generated waves are passing, and then photographing the paths of the dye droplets with a movie camera.

Because wave form is created by a looplike motion of water parcels, the diameters of the loops at the water surface exactly equal wave height (*H* in Fig. 8.16). Below the water surface, a progressive loss of energy occurs with increasing depth, expressed as a decrease in loop diameter. At a depth equal to half the **wavelength** (the distance between successive wave crests or troughs; *L* in Fig. 8.16), the diameters of the loops have become so small that water motion is negligible.

Wave Base

The depth *L/2* is the effective lower limit of wave motion and is generally referred to as the **wave base** (Figs. 8.16 and 8.17). In the Pacific Ocean, wavelengths as long as 600 m (1970 ft) have been measured. For them, *L/2* equals 300 m, a depth half again as great as the average depth of the outer edge of the continental shelves (about 200 m). Although the wavelengths of most ocean waves are far shorter than 600 m, it nevertheless is possible for very large waves approaching these dimensions to affect even the outer parts of continental shelves.

Breaking Waves

Landward of depth *L/2*, the circular motion of the lowest water parcels is influenced by the increasingly shallow seafloor, which restricts movement in the vertical direction. As the water depth decreases, the

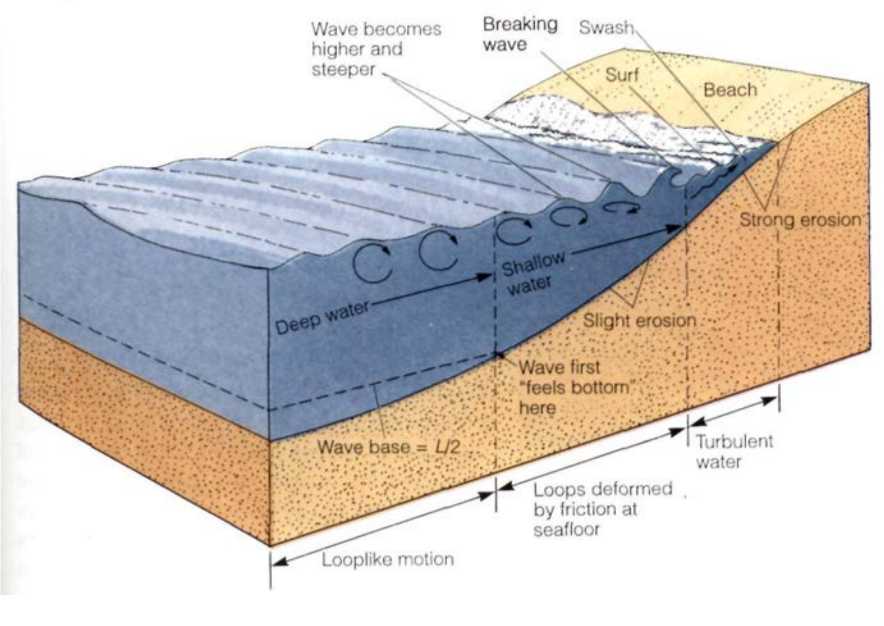


Figure 8.17 Waves change form as they travel from deep water through shallow water to shore. In the process, the circular motion of water parcels found in deep water changes to elliptical motion as the water shallows and the wave encounters frictional resistance to forward movement. Vertical scale is exaggerated, as is the size of loops relative to the scale of the waves.

Figure 8.18 Orienting the board for the best ride, a surfer skims the inside of a breaking wave off the coast of Hawaii.

loops of the water parcels become progressively flatter until, in the shallow water zone, the movement of water at the seafloor is limited to a rapid back and forth motion (Fig. 8.17). As depth decreases, the wave's shape is distorted; its height increases and the wavelength shortens. At the same time, the wave's front grows steeper. Riding the front of steep arching waves as they approach shore, a technique developed in Hawaii centuries ago, is an exhilarating but potentially dangerous sport (Fig. 8.18). Eventually the steep front is unable to support the advancing wave, and as the rear part continues to move forward, the wave collapses or *breaks* (Fig. 8.17), thereby becoming a *breaker*.

When a wave breaks, the motion of its water instantly becomes turbulent, like that of a swift river. Such "broken water" is called **surf**, defined as wave activity between the line of breakers and the shore. In surf, each wave finally dashes against rock or rushes up a sloping beach until its energy is expended; then it flows back. Water piled against the shore returns seaward in an irregular and complex way, partly as a broad sheet along the bottom and partly in localized narrow channels as rip currents, which are responsi-

Figurc 8.19 Waves arriving obliquely along a coast near Oceanside, California, change orientation as they encounter the bottom and begin to slow down. As a result, each wave front is refracted so that it more closely parallels the bottom contours. The arriving waves develop a longshore current that moves from left to right in this view.

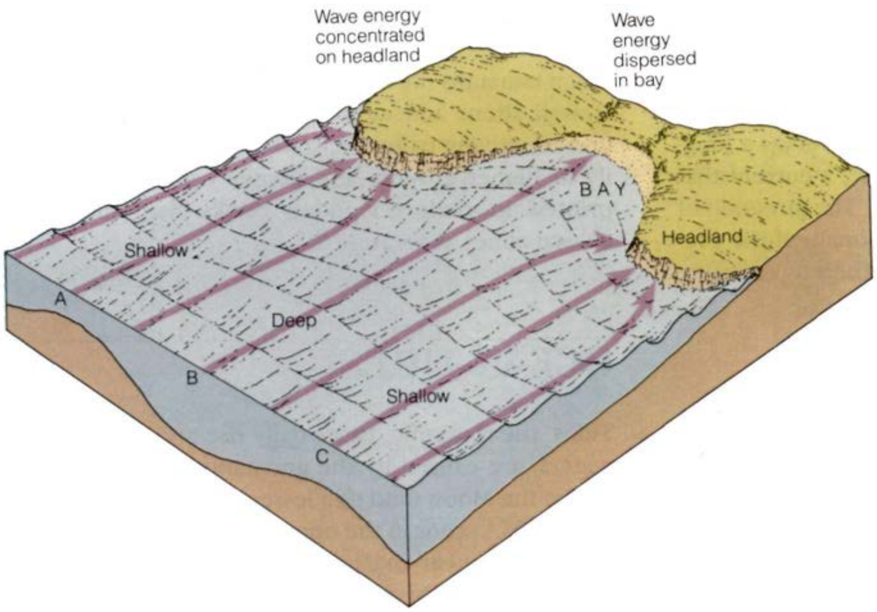


Figure 8.20 Refraction of waves concentrates wave energy on headlands and disperses it along bays. This oblique view shows how waves become progressively distorted as they approach the shore over a bottom that is deepest opposite the bay. The result is vigorous erosion on the exposed headland and sedimentation along the margin of the bay.

ble for dangerous "undertows" that can sweep unwary swimmers out to sea.

Surf possesses most of the original energy of each wave that created it. This energy is quickly consumed in turbulence, in friction at the bottom, and in moving the sediment that is thrown violently into suspension from the bottom. Although fine sediment is transported seaward from the surf zone, most of the geologic work of waves is accomplished by surf shoreward of the line of breakers.

Wave Refraction and Longshore Currents

A wave approaching a coast generally does not encounter the bottom simultaneously all along its length. As any segment of the wave touches the seafloor, that part slows down, the wavelength begins to decrease, and the wave height increases. Gradually the trend of the wave becomes realigned to parallel the bottom contours (Fig. 8.19). Known as wave **refraction**, this process changes the direction of a series of waves moving in shallow water at an angle to the shoreline. Thus, waves approaching the margin of a deep-water bay at an angle of 40° or 50° may, after refraction, reach the shore at an angle of 5° or less.

Waves passing over a submerged ridge off a headland will be refracted and will converge on the headland (Fig. 8.20). This convergence, as well as the increased wave height that accompanies it, concentrates wave energy on the headland, which is eroded vigorously. Conversely, refraction of waves approaching a bay will make them diverge, diffusing their energy at the shore. In the course of time, the net tendency of these contrasting effects is to make irregular coasts smoother and less indented.

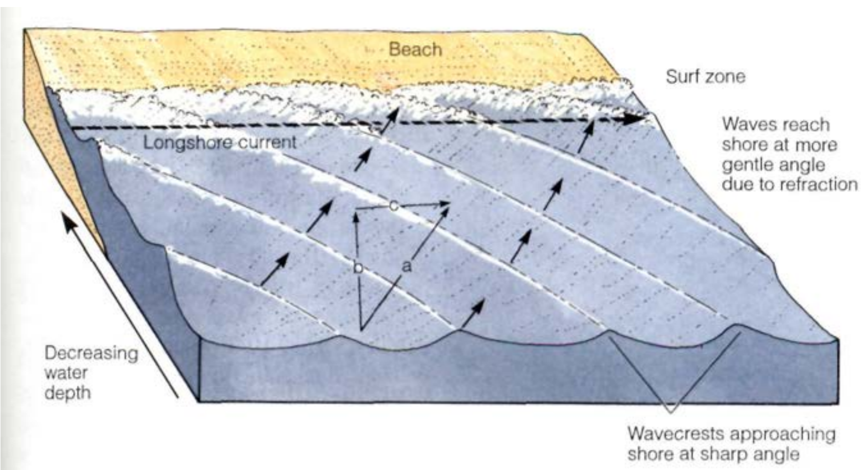


Figure 8.21 A longshore current develops parallel to the shore as waves approach a beach at a right angle and are refracted. A line drawn perpendicular to the front of each approaching wave (a) can be resolved into two components: the component oriented perpendicular to the shore (h) produces surf, whereas that oriented parallel to the shore (c) is responsible for the longshore current. Such a current can transport considerable amounts of sediment along a coast.

The path of an incoming wave can be resolved into two directional components, one oriented perpendicular to the shore and the other parallel to the shore. Whereas the perpendicular component produces the crashing surf, the parallel component sets up a longshore current within the surface zone, a current that flows parallel to the shore (Fig. 8.21). The direction of longshore currents may change seasonally if the prevailing wind directions change, thereby causing changes in the direction of arriving waves.

Seismic Sea Waves

A large earthquake or other brief, large-scale disturbance of the ocean floor, such as a landslide or volcanic eruption, can generate a potentially dangerous tsunami (Ch. 3). A tsunami is often erroneously referred to as a tidal wave, but it has nothing to do with the tides. Its threat lies in the great speed at which it travels (as much as 950 km/h, or 590 mi/h), its long wavelength (up to 200 km, or 124 mi), its low observable height in the open ocean, and its ability to pile up rapidly to heights of 30 m (100 ft) or more as it moves into shallow water along an exposed coast.

Hawaii is especially susceptible to dangerous tsunamis approaching from the numerous earthquake regions surrounding the Pacific basin (Fig. 8.22). The suddenness of arrival and consequent lack of warning time used to make tsunamis extremely hazardous. Numerous fatalities and considerable damage resulted when tsunamis moved into populated areas. Today in Hawaii sirens and radio newscasts alert the population to arriving tsunamis, so they can move temporarily to higher ground, and people can refer to maps printed in their telephone books that show the coastal zones at greatest risk.

OCEAN TIDES

Tides, the rhythmic, twice-daily rise and fall of ocean waters, are caused by the gravitational attraction between the Moon (and to a lesser degree, the Sun) and the Earth. A sailor in the open sea may not detect tidal motion, but near coasts the effect of the tides is amplified and they become geologically important.

Tide-Raising Force

The gravitational pull that the Moon exerts on the solid Earth is balanced by an equal but opposite inertial force (which tends to maintain a body in uniform linear motion) created by the Earth's rotation about the center of mass of the Earth-Moon system (Fig-8.23). At the center of the Earth, gravitational and inertial forces are balanced, but they are not balanced

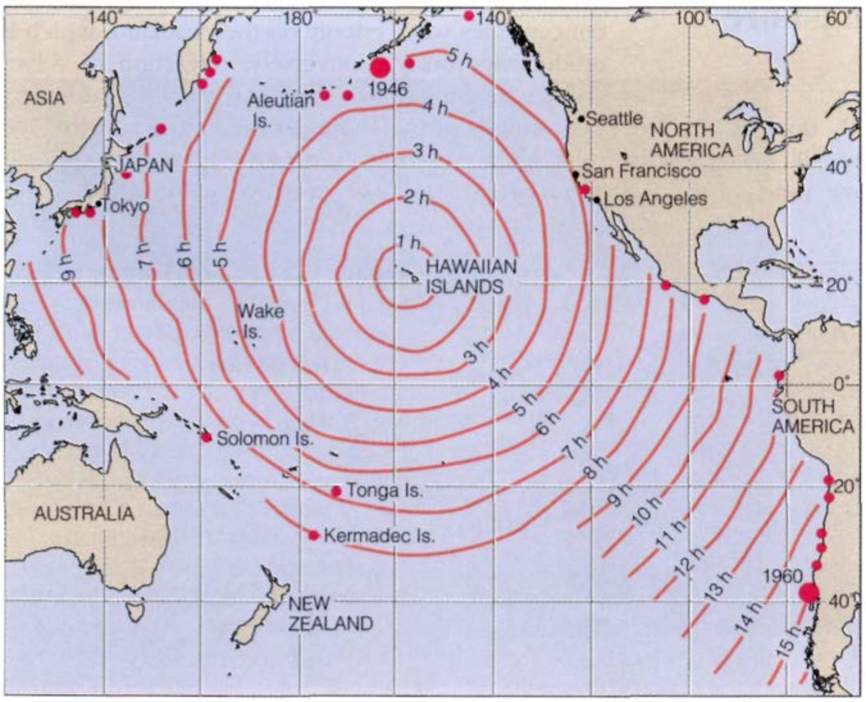


Figure 8.22 Map of the Pacific Ocean showing the time required for a tsunami to reach the island of Oahu, Hawaii. Small red dots mark the origins of tsunamis that have struck Hawaii. Large dots mark places where the disastrous tsunamis of 1946 and 1960 originated. The 1946 tsunami traveled from Alaska to Hawaii in 4.5 hours, whereas the 1960 tsunami took 15 hours to arrive from its place of origin at the coast of Chile.

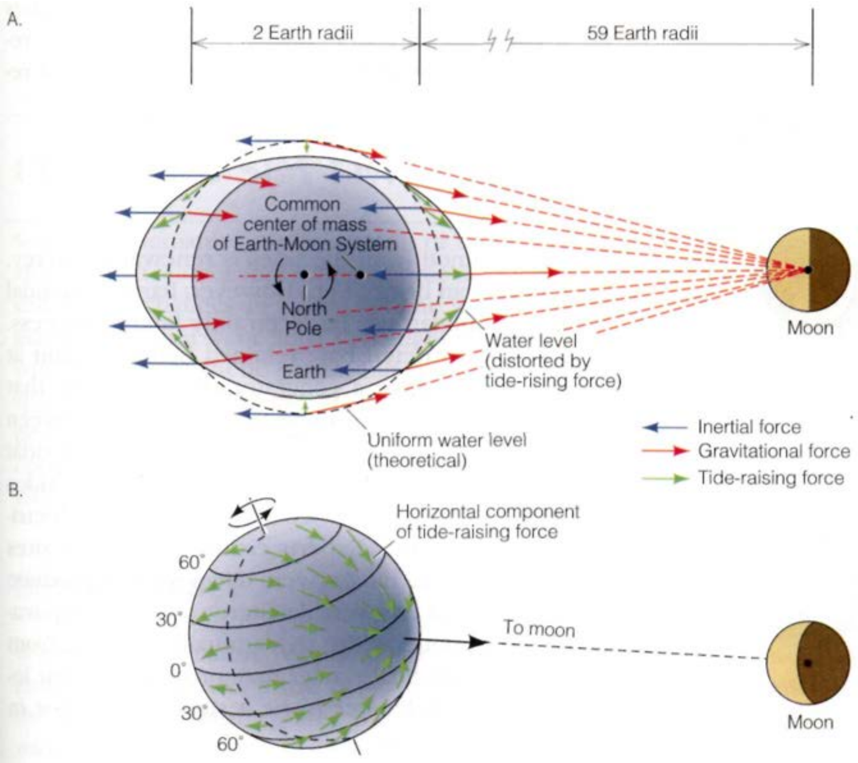


Figure 8.23 Tidal forces. A. Tide-raising forces are produced by the Moon's gravitational attraction and by inertial force. On the side toward the Moon, both forces combine to distort the water level from that of a sphere, raising a tidal bulge. On the opposite side of the Earth, where inertial force is greater than the gravitational force of the Moon, the excess inertial force (called the tide-raising force) also creates a tidal bulge. B. The horizontal component of the tide-raising force is shown by arrows on an oblique view of the Earth. The arrows are directed toward the point where a line connecting the Earth and Moon intersects the Earth's surface. This point shifts latitude with time as the relative position of the Earth and Moon change.

from place to place on the Earth's surface. A water particle in the ocean on the side facing the Moon is attracted more strongly by the Moon's graviation than it would be if it were at the Earth's center, which lies at a greater distance. Although the attractive force is small, liquid water is easily deformed, and so each water particle on this side of the Earth is pulled toward a point directly beneath the Moon. This creates a bulge on the ocean surface.

Although the magnitude of the Moon's gravitational attraction on a particle of water at the surface of the ocean varies over the Earth's surface, the inertial force at any point on the surface is the same. On the side nearest the Moon, gravitational attraction and inertial force combine, and the excess inertial force (or *tide-raising force)* is directed toward the Moon. On the opposite side of the Earth, the inertial force exceeds the Moon's gravitational attraction, and the tide-raising force is directed away from the Earth (Fig. 8.23). These unbalanced forces generate the daily ocean tides.

Tidal Bulges

The tidal bulges created by the tide-raising force on opposite sides of the Earth appear to move continually around the Earth as it rotates. In fact, the bulges remain essentially stationary beneath the tide-producing body (the Moon) while the Earth rotates. At most places on the ocean margins, two high tides and two low tides are observed each day as a coast encounters both tidal bulges. In effect, at every high tide, a mass of water runs into the coastline, where it piles up. This water then flows back to the ocean basin as the coastline passes beyond each tidal bulge.

Earth-Sun gravitational forces also affect the tides, sometimes opposing the Moon by pulling at a right angle and sometimes aiding by pulling in the same direction. Twice during each lunar month, the Earth is directly aligned with the Sun and the Moon, whose gravitational effects are thereby reinforced, producing higher high tides and lower low tides (Fig. 8.24). At positions halfway between these extremes, the gravitational pull of the Sun partially cancels that of the Moon, thus reducing the tidal range. However, the Sun is only 46 percent as effective as the Moon in producing tides, so the two tidal effects never entirely cancel each other.

In the open sea, the effect of the tides is small (less than 1 m, or 3 ft) and along most coasts the tidal range commonly is no more than 2 m (6.5 ft). However, in bays, straits, estuaries, and other narrow places along coasts, tidal fluctuations are amplified and may reach 16 m (52 ft) or more (Fig. 8.25). Associated currents are often rapid and may approach 25 km/h (16 mi/h). The incoming tide locally can create a wall of water a meter or more high (called a *tidal* *bore)* that moves up estuaries and the lower reaches of streams. Fast-moving *tidal currents,* though restricted in extent, constitute a potential source of renewable energy that is still largely untapped.

Tidal Power

Energy obtained from the tides is renewable energy, for it never can be used up. However, harnessing tidal power for human use has seen only limited success. Water in a restricted bay, retained behind a dam at high tide, can drive a generator the same way that river water can. One important difference between hydroelectric power from rivers and that from tidal power is that rivers flow continuously, whereas tides can be exploited only twice a day; therefore, electrical supply is erratic. Experts estimate that if all sites with suitable tidal ranges were developed to produce power, the total recoverable energy would be equivalent to only about a tenth that annually obtained from oil. Thus, while tidal power may prove important locally, it is unlikely ever to be a significant factor in global energy supply.

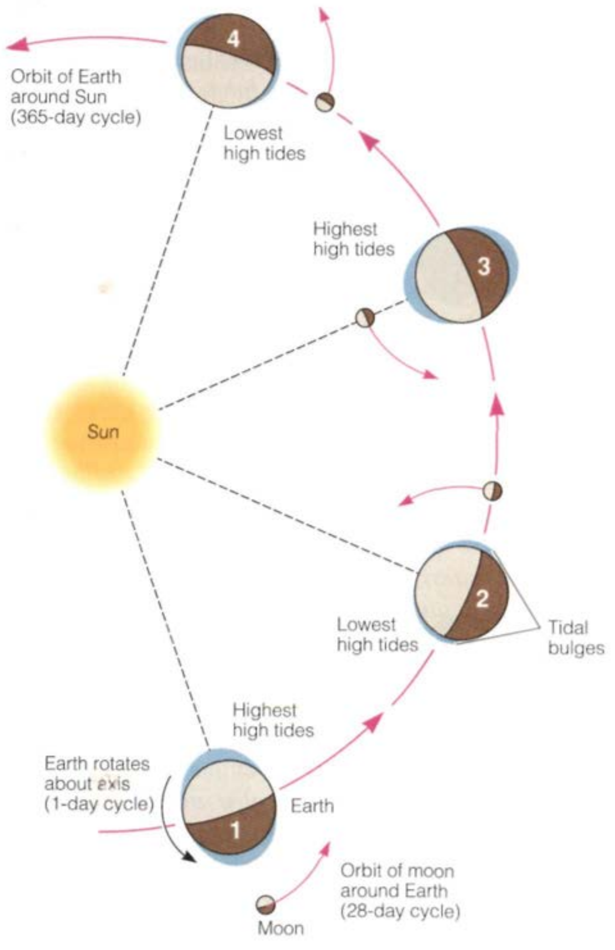


Figure 8.24 When the Earth, Moon, and Sun are aligned (positions 1 and 3), tides of highest amplitude are observed. When the Moon and Sun are pulling at right angles to each other (positions 2 and 4), tides of lowest amplitude are experienced.


A.


B.

Figure 8.25 The tidal range in the Bay of Fundy, eastern Canada, is one of the largest in the world. A. Coastal harbor of Alma, New Brunswick at high tide. B. Same view at low tide.

CHANGING SEA LEVEL

Sea level fluctuates daily as a result of tidal forces. It also fluctuates over much longer time scales as a result of (1) changes in water volume as continental glaciers wax and wane and (2) changes in ocean-basin volume as lithospheric plates shift position. Over the span of a human lifetime, these slower changes appear insignificant, but on geologic time scales they contribute in an important way to the evolution of the world's coasts.

Submergence

Whatever their nature, nearly all coasts have experienced submergence, a rise of water level relative to the land. This **submergence** is the result of a worldwide rise of sea level that occurred when glaciers melted away at the end of the last ice age (Fig. 8.26). Because of this submergence, evidence of lower sea levels during ice ages is almost universally found seaward of the present coastlines and to depths of 100 m (330 ft) or more. Former beaches, sand dunes, and other coastal landscape features now submerged on the inner continental shelves mark shorelines built by the rising sea at the end of the glacial age and later drowned.

Emergence

Evidence of past higher sea levels is related mainly to former interglacial ages similar to the present. Shoreline features well inland from the present Atlantic coast of the United States, from Virginia to Florida, reach altitudes of 30 m (100 ft) or higher. These typical coastal landforms owe their present altitude to a combination of broad upward arching of the crust, and submergence during warmer times when glaciers were smaller and sea level therefore higher. The presence of such features above sea level points to **emergence**, a lowering of water level relative to the land.

Geologists recognize evidence of repeated cycles of emergence and submergence along the world's

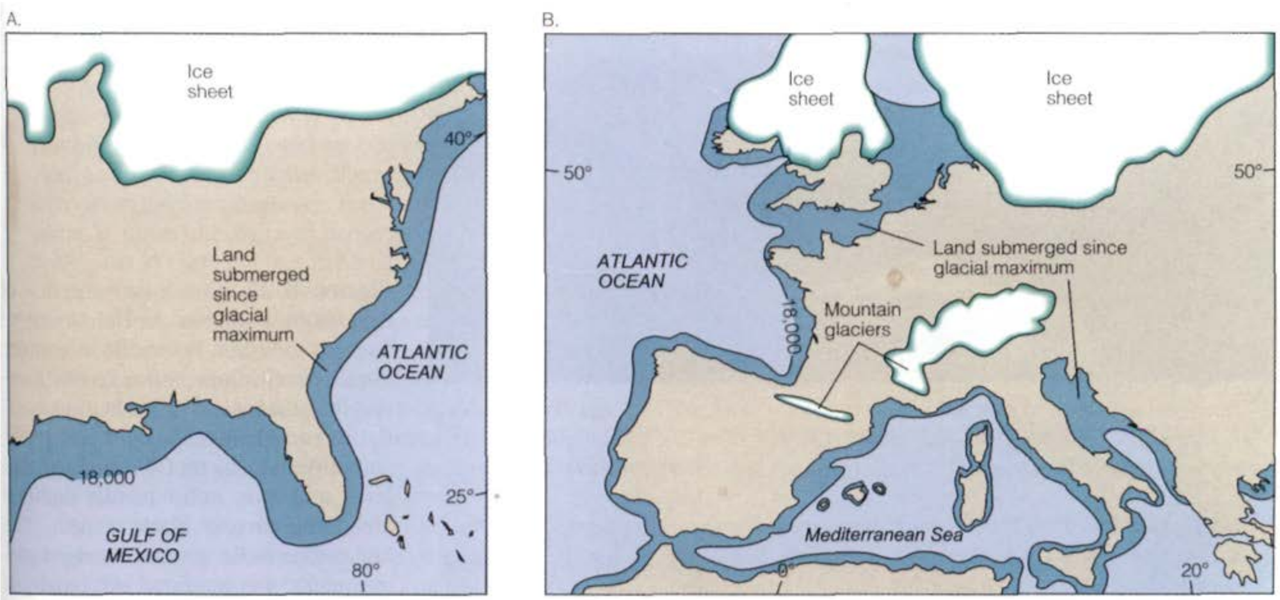


Figure 8.26 Coastal submergence of eastern North America and western Europe resulted when meltwater from wasting ice sheets returned to the ocean basins at the close of the last glaciation. A. Area of northeastern North America covered by glacier ice during the last glacial maximum, the approximate position of the shoreline at the glacial maximum (18,000 years ago; dark green line), and coastal areas submerged by the postglacial rise of sea level. B. Areas covered by ice sheets in Western Europe at the last glacial maximum and land areas that have been submerged during the postglacial rise of sea level.

coasts, each cycle related to the buildup and decay of vast ice age glacier systems. This evidence clearly demonstrates how major changes of the global climate system determine the volume of glaciers on land, which in turn affects the level of the world ocean (Chapter 14).

Relative Movements of Land and Sea

The rise and fall of sea level that cause coastal submergence and emergence are global movements, affecting all parts of the world ocean at the same time. The other causes of submergence and emergence, uplift and subsidence of the land, generally involve limited portions of landmasses. Nevertheless, uplift and subsidence can cause rapid relative changes in sea level. Vertical tectonic movements where converging plates meet have uplifted beaches and tropical reefs to positions far above sea level (Figs. 8.27). Because tectonic movements and sea-level changes may occur simultaneously, either in the same or opposite directions and at different rates, unraveling the history of sea-level fluctuations along a coast can be a challenging exercise.

A.

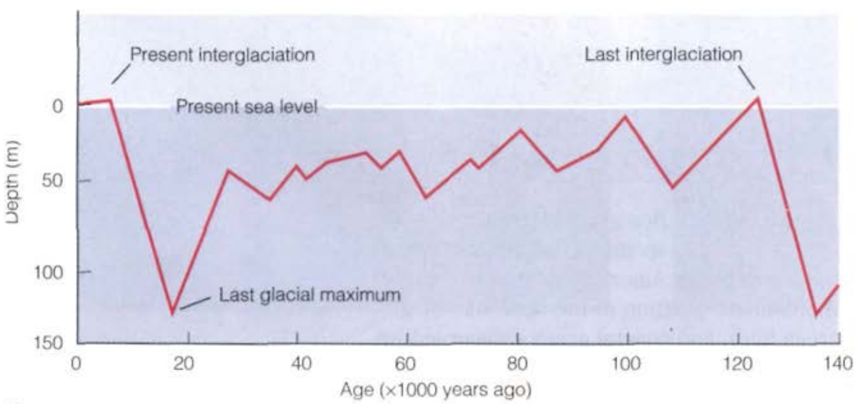


Figure 8.27 Coastal emergence of eastern New Guinea. A. The emergent coast of the Huon Peninsula in eastern Papua New Guinea is flanked by a series of ancient coral reefs that form flat terracelike benches parallel to the shoreline. Each reef formed at sea level and was subsequently uplifted along this active plate margin. The highest reefs lie several hundred meters above sea level and are hundreds of thousands of years old. B. Curve of sea-level fluctuations constructed by uranium-isotope dating the uplifted reefs of the Huon Peninsula. Prior to a recent high stand, sea level was lower than at present ever since the last interglaciation, about 120,000 years ago.

Guest Essay

Where the Sun Never Rises. Seeing is Believing

Victoria A Kaharl, author of *Water Baby: The Story of Alvin* (Oxford University Press, 1990), is science Writer in Residence at the Woods Hole Oceanographic Institution, Woods Hole, Massachusetts.

For most of the time humans have inhabited the Earth, the depths of the ocean have remained unknown and unknowable, a world cut off from the land, the sky, and the sun-lit top layers of the sea. For a long time, nobody believed there could be life in the frigid black depths. The nineteenth-century naturalist Edward Forbes probed the ocean by dangling hemp line attached to nets and dredges. Because he never brought up any life from deeper than 550 meters, Forbes concluded there was none. Beneath 550 meters, he wrote in 1841, lay the lifeless *azoic zone.*

At the time, however, there was some evidence to the contrary. The crews who repaired submarine telegraphic cables occasionally found animals, such as starfish, latched onto a deep-sea cable.

More evidence for life in the deep sea came during the research expedition of the H.M.S. *Challenger.* The British naval corvette left England three days before Christmas in 1872 for a round-the-world research cruise that would last three years and five months. Although research cruises had sailed before this, the *Challenger* cruise is often used to mark the beginning of oceanography as a truly organized and comprehensive science. The naturalists on the *Challenger* pulled up life, some marine worms, from 5500 meters.

The hemp rope used by the nineteenth-century naturalists eventually gave way to stronger piano wire for plumbing the depths. Modern oceanographers still use nets and dredges but lower them on thick steel cable. Whatever the line, however, this kind of selective sampling is done blindly, like hanging out of an airplane with a net at midnight to catch butterflies.

Because the ocean is extremely inhospitable to the explorer species *Homo sapiens,* anyone who probes this inner space is utterly dependent on technology. Oceanographic engineers are fond of saying the ocean is more hostile than the Moon. There is no saltwater on the Moon to corrode electronics, no pressure to implode instruments, no slimy animals looking for a nook to make their home

After World War II, some 70 years after the *Challenger* cruise, cameras and lights that could withstand the crushing pressures of the ocean and the corrosive effects of saltwater revealed more of the faraway deep ocean. Thousands of black and white photographs captured snatches of the deep sea. Almost all of the animals in the photographs were small; most were no bigger than flies. The photographic record showed that, although there was life in the deep ocean, there wasn't much of it. Oceanographers likened the deep ocean to a graveyard or a desert.

It would be decades before our perception and understanding of this world began to change. We had to await a new technology, one that could not only take humans to the deep sea but also allow us to move around, to actually explore and conduct experiments in that realm. This new tool was *Alvin,* a three-passenger research submarine commissioned in 1964 as a national oceanographic facility.

About the size of a UPS delivery truck, *Alvin* usually carries a pilot and two scientists on an average eight-hour dive, which is about the lifetime of its batteries. Passengers crouch beside one of the three small plastic portholes inside a 2-meter-wide titanium sphere. They breathe normally; oxygen is automatically bled into the sphere, and carbon-dioxide is removed to keep the atmosphere the same as room air. Because there is no heat inside the submarine, passengers dress warmly because it is always cold in the deep sea. With a generous lunch bag hanging from the the center of the passenger sphere, there is no room to stand up. Most people don't mind; they're too busy looking. The world outside *Alvin's* small windows, each wide enough for only one pair of eyes, is still new and will continue to be for decades to come. We have explored only a fraction of the largest environment on Earth, but the little we have seen has changed once and for all a host of misconceptions about the deep sea.

Today, with more than 2000 dives to its credit, *Alvin* takes scientists to deep-sea sites with names like the Garden of Eden, Clam Acres, and Anemone Heaven. *Graveyard* and *desert* do not apply to any of these communities; each is a hydrothermal vent teeming with life.

Summary

1. Seawater covers nearly 71 percent of the Earth's surface and is concentrated in the Pacific, Atlantic, and Indian oceans. Each is connected to the Southern Ocean, which encircles Antarctica.
2. Although the greatest ocean depth is more than 11 km, the average depth is 3.8 km. More than half the ocean water resides in the Pacific basin.
3. Evidence from sedimentary strata imply that the world has had an ocean since at least 3.95 billion years ago. Most likely, the ocean condensed from steam produced during primordial volcanic eruptions.
3. More than 99.9 percent of the saltiness of seawater is due to eight ions that are derived through chemical weathering of rocks on land and then transported by streams to the sea. Other sources of ions include airborne dust and human pollutants.
4. Sea-surface temperatures are strongly related to latitude, with the warmest temperatures measured in equatorial latitudes. Surface salinity is also strongly latitude-dependent and is related to both evaporation and precipitation.
5. Ocean water is stratified as a result of density differences that are related to salinity and temperature. A thin surface zone of relatively low density is separated from a deep zone of dense water by the pyenoctine, a zone where density changes rapidly with increasing depth. The pycnocline coincides approximately with the halocline, a zone of rapid salinity change, and the thermocline, a zone of rapidly changing temperature.
6. Huge wind-driven surface ocean currents that circulate clockwise in the northern hemisphere and counterclockwise in the southern hemisphere carry warm equatorial water toward the polar regions.
7. The Coriolis effect deflects ocean currents to the right in the northern hemisphere and to the left in the southern hemisphere, the magnitude of the effect increasing from the equator toward the poles. As a result, surface currents move toward huge midocean gyres where gravitational force balances the Coriolis effect, producing geostrophic flow.
8. El Niño/Southern Oscillation, an important element of year-to-year climatic variation, occurs when the tradewinds slacken and surface waters of the central and eastern Pacific become anomalously warm.
9. Ekman transport, which is the average flow direction through about the upper 100 m of the ocean, moves at 90° to the wind direction. Downwelling occurs when Ekman transport is toward a coast, whereas upwelling occurs if it is away from a coast.
10. Sinking of cold and (or) saline high-latitude surface waters leads to oceanwide thermohaline circulation. Operating like a great conveyor belt and driven by density contrasts, this global circulation system replenishes the deep water of the world ocean, replacing it in about 1000 years.
11. The motion of wind-driven surface waves terminates downward at the wave base, a distance equal to half the wavelength. A wave breaks in shallowing water as interference with the bottom causes the wave to grow higher and steeper. Waves approaching shore are refracted as the wave base reaches the bottom, realigning the wave so that it reaches the shore at a gentler angle.
12. Twice-daily ocean tides, resulting from the gravitational attraction of the Moon and Sun, are produced as the surface of the rotating Earth passes through tidal bulges on opposite sides of the planet.
13. Recently, nearly all coasts have experienced submergence as a result of the postglacial rise of sea level. Some coasts have experienced more complicated histories of relative sea-level change where tectonic movements have been superimposed over the worldwide sea-level rise.

Important Terms to Remember

Coriolis effect (p. 203)
downwelling (p. 206)
Ekman spiral (p. 206)
Ekman transport (p. 206)
El Niño/Southern Oscillation (ENSO) (p. 207)
emergence (p. 219)
evaporite deposits (p. 201)
geostrophic current (p. 209)
gyre (p. 204)
halocline (p. 203)
heat capacity (p. 202)
North Atlantic Deep Water (NADW) (p. 211)
pycnocline (p. 203)

salinity (p. 199)
submergence (p. 219)
surf (p. 214)
thermocline (p. 203)
thermohaline circulation (p. 211)
tsunami (p. 216)
upwelling (p. 206)
wave base (p. 213)
wavelength (p. 213)
wave refraction (p. 215)
western boundary current (p. 205)

Questions for Review

1. In terms of plate tectonics, explain why the Pacific Ocean basin is the largest of the three major ocean basins of the Earth.
2. What geologic evidence indicates that liquid water has been present at the Earth's surface for at least 3.95 billion years?
3. If the quantity of dissolved ions carried to the oceans by streams throughout geologic history far exceeds the known quantity of these substances in modern seawater, why is seawater not far more salty than it is?
4. How and why are the temperature and salinity of surface ocean water related to latitude?
5. Explain why the oceans are vertically stratified with respect to water density.
6. Explain why the seasonal temperature range is less at the western coast of North America than it is several hundred kilometers inland.
7. What geologic features would you look for to determine whether a coastal region had experienced emergence or submergence in the recent geologic past?
8. Explain the existence of large midocean gyres both north and south of the equator.
9. What contrasting ocean conditions give rise to upwelling and downwelling along continental margins?
10. Why might a ship in midocean not detect the passage of a tsunami, while one anchored in a constricted ocean-facing harbor would not miss its arrival?
11. Why does a ship tied to a dock experience two high tides and two low tides each day?
12. Describe how North Atlantic Deep Water is produced and explain its role in the global ocean conveyor system.

Questions for A Closer Look

1. Describe the effects of an El Niño event at three sites in a transect across the equatorial Pacific Ocean: in Indonesia, at an island in the middle of the ocean, and at the coast of Peru.
2. How can a past record of El Niño events be reconstructed from geologic evidence? Why would such a record be useful to us?

Questions for Discussion

1. Imagine yourself a navigational assistant on the *H.M.S. Beagle* as it sets sail, with Charles Darwin aboard as naturalist, from England toward the southeastern coast of South America. Suggest a course that Captain Fitzroy could follow in order to take advantage of winds and currents as you travel from Portsmouth (50.8°N) to Montevideo (35°S)?
2. What can we infer about the history of past sea level from the fact that 30 percent of North America is underlain by marine evaporite deposits?
3. Consider the following hypothesis: the thermohaline circulation system of the oceans has not operated continuously but from time to time has shut down. What might cause the system to shut down, and what evidence might you look for to see if this had happened in the past?

CHAPTER 9

Water on the Land: Surface Streams and Groundwater

View of Aswan High Dam, which impounds the Nile River (right) to form Lake Nasser impounds (to south, behind dam). Sediment formerly carried northward to the Mediterranean Sea and now settling out in the lake will eventually fill the reservoir and make it unusable.

Tampering with the Nile

For more than seven millennia people have lived along the banks of the lower Nile River, a ribbon of green that rises in the highlands of eastern Africa and crosses the vast desert of Sudan and Egypt on its way to the Mediterranean Sea. As far back as the age of the pyramids, people learned to live with the Nile's annual cycle. During part of each year the river flowed peacefully within its banks, but by late summer it began to rise during the annual period of flooding. Although floods could devastate those living along the stream's margins, at the same time the floods were beneficial, for they carried an abundant load of silt that provided annual nourishment to agricultural fields.

Recently, however, the Nile's ability to supply Egypt's basic water needs has been seriously threatened by the region's exploding population. In the early nineteenth century, when a French mission sent by Napoleon conducted a census, the estimated population of Egypt was less than 2.5 million. By the 1920s the number had risen to 14 million, and since then the population has tripled. The estimate of current population growth is more than 1 million people every nine months.

To provide an adequate water supply and stabilize flow through the lower, densely populated Nile basin, a high dam was constructed at Aswan in the 1960s. The Aswan Dam was designed to reduce the immense difference in flow rates between the flood and low-water seasons and to permit full utilization of the Nile water. It was intended to provide water not only for human consumption but for irrigation, hydroelectric power, and inland navigation as well. These primary goals have been realized, for the dam led to a marked change in the river's annual flow pattern. No longer is there a large seasonal flood below the dam; instead, the controlled annual outflow of water from the vast reservoir (Lake Nasser) held behind the dam is more uniform. Despite the success of the project, human tampering with the Nile brought some unanticipated consequences.

The Nile, like all other large streams, is a complex natural system, and its behavior reflects a delicate balance between water flow and sediment load. Ninety-eight percent of the Nile's sediment load is carried in suspension. Prior to construction of the Aswan Dam, an average of 125 million metric tons (275 billion lbs) of sediment passed downstream each year, but the dam reduced this value to only 2.5 million metric tons (5.5 billion lbs). Nearly 98 percent of the suspended sediment is now deposited in Lake Nasser. Under natural conditions, floodwater carried this sediment downstream, where much of it was deposited, thus adding to the rich agricultural soils at a rate of 6 to 15 cm/century (2 to 6 in/century). With this natural source of nourishment eliminated, farmers must now resort to artificial fertilizers and soil additives to keep the land productive. Some of the fertilizer seeps back into the river, causing pollution problems downstream.

Before the Aswan Dam was constructed, the annual Nile flood transported at least 90 million metric tons (200 billion lbs) of sediment to the Mediterranean Sea. The shoreline of the sea at that time reflected a balance between sediment supply and the attack of waves and currents that redistributed the sediment along the Mediterranean coast. Because the annual discharge of sediment has now been cut off, the coast has become increasingly vulnerable to erosion. Over the long term, we can expect to see continuing changes as the shoreline adjusts to the reduction in Nile sediment and a new balance is reached.

STREAMS AND DRAINAGE SYSTEMS

Almost anywhere we travel over the land surface, we can see evidence of the work of running water. Even in places where no rivers flow today, we are likely to find sediments and landforms that tell us water has been instrumental in shaping the landscape. Most of these features can be related to the activity of streams that are part of complex drainage systems.

Drainage systems evolve because a significant portion of the water falling on the land as precipitation collects and moves downslope in the general direction of the nearest ocean. The Earth's drainage systems thus form a fundamental part of the hydrologic cycle: water is evaporated from the oceans into the atmosphere, a portion is precipitated on the land surface, and part of this travels across the land on its way back to the sea.

The average annual rainfall on the area of the United States is equivalent to a layer of water 76 cm (30 in) thick equally covering this same land surface. Of this layer, 45 cm (17.7 in) returns to the atmosphere by evaporation and transpiration, and 1 cm (0.4 in) infiltrates the ground; the remaining 30 cm (11.9 in) forms **runoff,** the portion of precipitation that flows over the land surface. By standing outside during a heavy rain, you can see that water initially tends to move downslope in broad, thin sheets called *overland flow.* You will also notice, however, that after traveling a short distance overland flow begins to concentrate into well-defined channels, thereby becoming *streamflow.* Runoff is a combination of overland flow and streamflow.

A **stream** is a body of water that flows downslope along a clearly defined natural passageway and in the process transports particles of sediment and dissolved substances (Fig. 9.1). The passageway is the stream's **channel,** and the sediment constitutes the bulk of its

Figure 9.1 Maroon Creek, in Colorado's White River National Forest, produees a succession of small rapids where it Bows over and between boulders scattered along its gravelly channel.

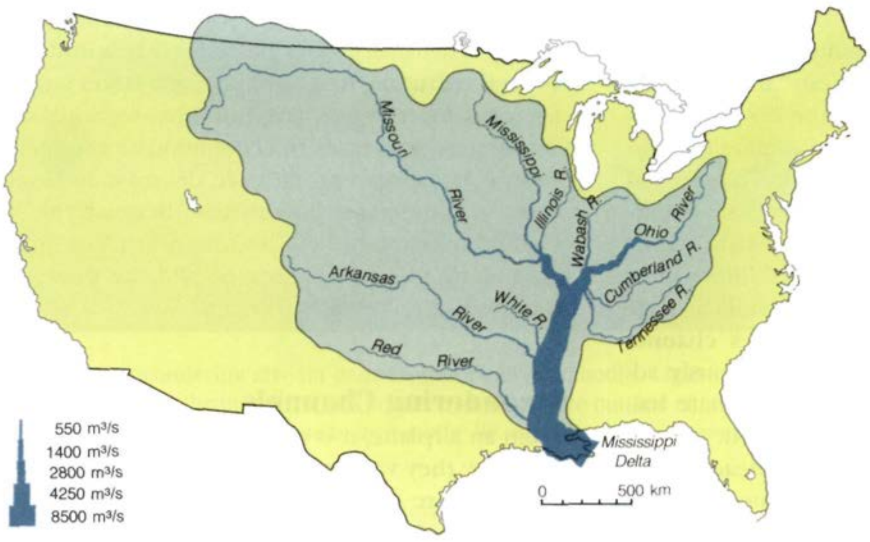


Figure 9.2 The drainage basin of the Mississippi River encompasses a major portion of the central United States. In this diagram, the width of the river and its major tributaries reflects discharge rates.

load, which is the total of all the sediment and dissolved matter that the stream transports. Geologists refer to the sediment load as **alluvium.**

Every stream or segment of a stream is surrounded by its **drainage basin,** the total area that contributes water to the stream. The line that separates adjacent drainage basins is a **divide.** Drainage basins range in size from less than a square kilometer to vast areas of subcontinental dimension. In North America, the huge drainage basin of the Mississippi River encompasses an area that exceeds 40 percent of the area of the contiguous United States (Fig. 9.2). It should come as no surprise that the area of any drainage basin is proportional to both the length of the stream that drains the basin and to the average annual volume of water that moves through the drainage system.

Stream Channels

A stream channel is an efficient conduit for carrying water. The size and shape of any particular channel cross section reflects the typical stream conditions at that place. Some very small streams are about as deep as they are wide, whereas very large streams usually have widths many times greater than their depths.

If we measure the vertical distance that a stream channel descends between two points along its course, we will have obtained a measure of the stream's **gradient.** The average gradient of a steep mountain stream may reach 60 m/km (330 ft/mi) or even more, whereas near the mouth of a large river the gradient may be less than 0.1 m/km (0.5 ft/mi).

Overall, the gradient of a river decreases downstream but not always smoothly, as any white-water rafter or kayaker can attest. A local change in gradient may occur, for example, where a channel passes from resistant rock into more erodible rock, or where a landslide or lava flow forms a natural dam across the channel. At such places, water may tumble rapidly through a stretch of rapids or form a waterfall where it plunges over a steep drop.

Dynamics of Streamflow

Five basic factors control a stream's behavior:

1. The *average width* and *depth* of the channel
2. The channel *gradient*
3. The *average velocity* of the water
4. The **discharge**, which is the quantity of water passing a point on a stream bank during a given interval of time
5. The *sediment load* (The dissolved component of the load generally has little effect on stream behavior.)

A stream experiences a continuous interplay among these factors. Measurements of natural streams show that, as discharge changes, velocity and channel shape also change. The relationship can be expressed by the formula

Discharge	=	Cross-sectional area of channel	X	Average velocity
(m^3/s)		(width X average depth) (m2)		(m/s)

The variable factors in this equation are interdependent, which means that when one changes, one or more of the others will also change. For example, with increased discharge, the velocity also typically increases. This can cause the stream to erode and enlarge its channel, rapidly if it flows on alluvium and much more slowly if it flows on bedrock. This erosion continues until the increased discharge can be accommodated in a larger channel and by faster flow. When discharge decreases, the channel dimensions decrease as some of the load is dropped, and as a result the velocity decreases. In these ways channel width, channel depth, and velocity continuously adjust to changing discharge, and an approximate balance among the various factors is maintained.

Traveling down a stream from its head to its mouth, we can see that orderly adjustments occur along the channel. For example, (1) width and depth of the channel increase, (2) gradient decreases, (3) velocity increases slightly, and (4) discharge increases (Fig. 9-3). The fact that velocity increases downstream seems to contradict the common observation that water rushes down steep mountain slopes and flows smoothly over nearly flat lowlands. However, the physical appearance of a stream is not a true indication of its velocity. Discharge is low in the headward reaches of a stream, and average velocity is also low because of the frictional resistance, or drag, caused by the water passing over a very rough stream bed. Here, where the flow is *turbulent* (agitated or disorderly), the water moves in many directions, rather than uniformly downstream. Discharge increases downstream as each **tributary** (a stream joining a larger stream) introduces more water, and the cross-sectional area of the channel increases to accommodate the increased volume. Despite a progressive decrease in slope, velocity also increases downstream because the roughness of the stream bed, and therefore the frictional drag along the bed, decreases, and the flow is more uniformly directed downstream.

Meandering Channels

From an airplane, it is easy to see that no two streams are alike; they vary in size and shape. Straight channel segments are rare and generally occur for only brief stretches before the channel assumes a sinuous shape. In many streams, the channel forms a series of smooth, looplike bends with similar dimensions (Fig. 9.4). Such a looplike bend of a stream channel is called a **meander**, after the Menderes River (in Latin, *Meander)* in southwestern Turkey which is noted for its winding course. Meanders occur most commonly in channels cut in fine-grained alluvium that have gentle gradients. The meandering pattern reflects the way in which a river minimizes resistance to flow and dissipates energy as uniformly as possible along its course.

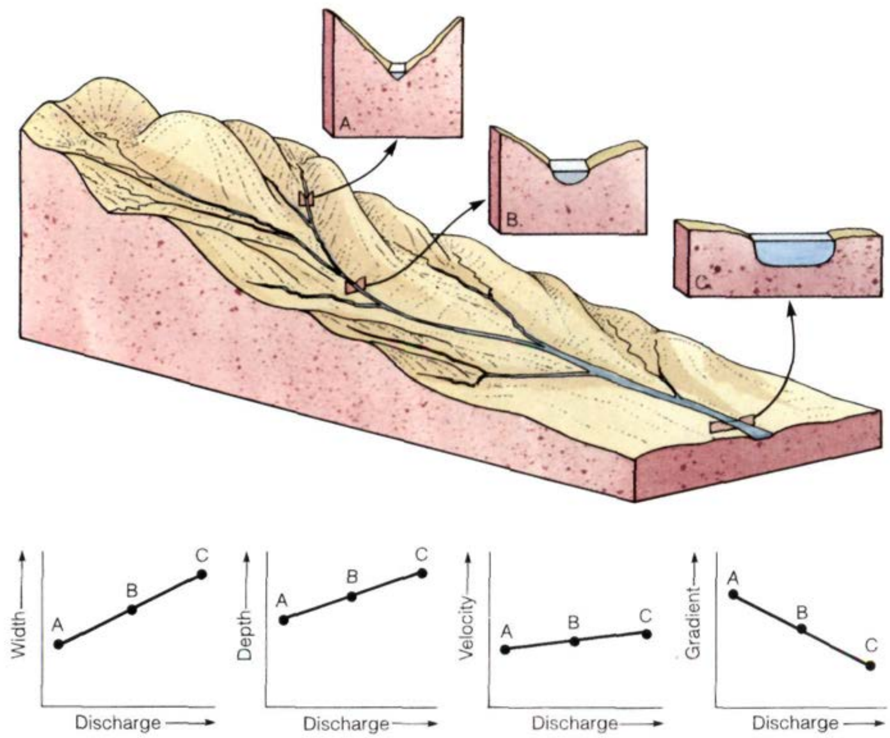


Figure 9.3 Changes in stream properties along a river system. Discharge increases as new tributaries join the main stream. Channel width and depth are shown by cross sections A, B, and C. Graphs show the relationship of discharge to channel width and depth, to velocity, and to channel gradient at the same three cross sections.

Figure 9.4 A meandering stream near Phnom Penh, Cambodia. Light-colored point bars, composed of gravelly alluvium, lie opposite steep banks on the outside of meander bends. Two oxbow lakes, the product of past meander cutoffs, lie adjacent to the present channel.

Try to wade or swim across a meandering stream, and it quickly becomes apparent that the velocity of the flowing water is not uniform. Velocity is lowest along the bed and walls of the channel because here the water encounters increasing frictional resistance to flow. The maximum velocity along a straight channel segment is found near the surface in midchannel. However, wherever the water rounds a bend, the zone of highest velocity swings toward the outside of the channel (Fig. 9.5).

The nearly continuous shift, or migration, of a meander is accomplished by erosion on the outer banks of the meander loops. Along the inner side of each meander loop, where water is shallowest and velocity is lowest, coarse sediment accumulates to form a distinctive *point bar* (Fig. 9.4). Collapse of the stream banks occurs most frequently along the outer side of a meander bend where the highest current velocity impinges on the channel bank, causing erosion and undercutting. In this way, meanders tend to migrate slowly down a valley, progressively removing and adding pieces of real estate along the banks.

Wherever a segment of a meander that is cutting into sandy alluvium encounters less erodible sediment, such as clay, downvalley migration of the meander will be slowed. Meanwhile, the upstream segment of the meander, migrating more rapidly through erodible sandy sediment, may intersect and cut into the slower moving downstream segment (Fig. 9.6A). When this happens, the stream bypasses the channel loop between the upstream and downstream, cutting it off and converting it into an arcuate *oxbow lake.* Because the new course is shorter than the older course, the channel gradient is steeper there and the overall stream length is shortened (Fig. 9-6B).

Nearly 600 km (373 mi) of the Mississippi River channel has been abandoned through cutoffs since 1776. However, the river has not been shortened ap-

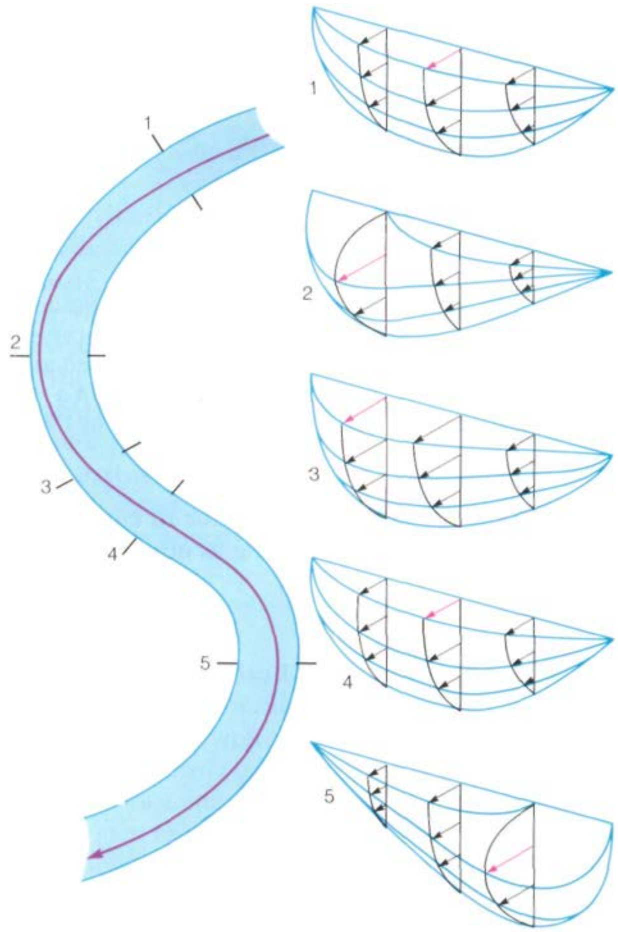


Figure 9.5 Velocity distribution in cross sections through a sinuous channel, (Lengths of arrows indicate relative flow velocities.) Where the channel is relatively straight (sections 1 and 4), the zone of highest velocity (red arrow) lies near the surface and toward the middle of the stream. At bends (sections 2 and 5), the maximum velocity swings toward the outer bank and lies below the surface.

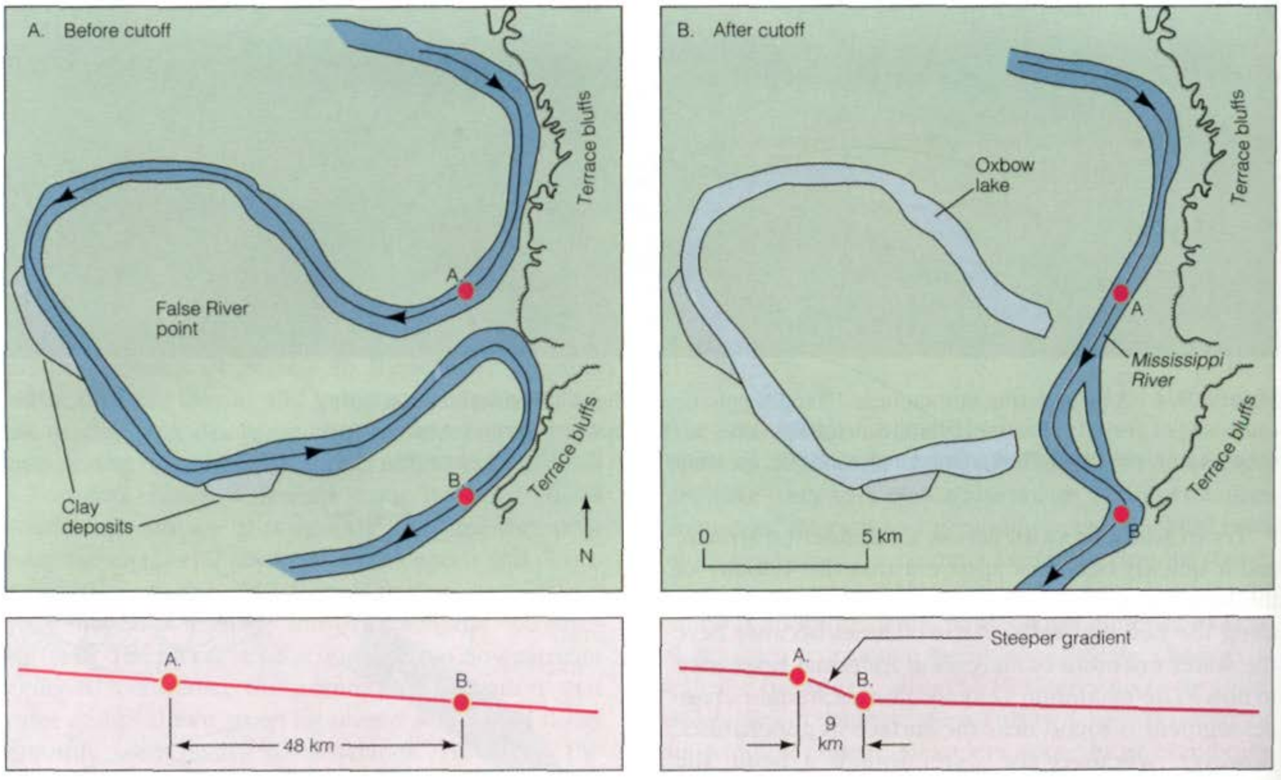


Figure 9.6 Cutoff of a meander loop of the Mississippi River in Louisiana. A. The downvalley migration of the loop encircling False River Point was halted when the channel segment on the south side of the loop encountered a body of clay in the alluvium. The next meander loop upstream continued to advance and finally cut off the entire loop surrounding False River Point. B. Over its new, shorter path, the stream had a steeper gradient than the abandoned course, and it developed a new pattern in which the single channel was replaced by two channels with an island between them.

preciably because the shortening due to cutoffs was balanced by channel lengthening as other meanders were enlarged.

Braided Channels

The intricate geometry of a **braided stream** resembles the pattern of braided hair, for the water repeatedly divides and reunites as it flows through two or more adjacent but interconnected channels separated by bars or islands (Fig. 9.7). A stream unable to transport all the available load tends to deposit the coarsest sediment as a bar, which locally divides the flow and concentrates it in the deeper segments of channel to either side of the bar. As the bar builds up, it may emerge above the stream surface as an island and become stabilized by vegetation that anchors the sediment and inhibits erosion.

Large braided rivers typically have numerous shallow channels that change size and shift position as the stream erodes and deposits sediment (Fig. 9.8). Although at any moment the active channels may cover no more than 10 percent of the width of the entire channel system, within a single season all or most of

Figure 9.7 The shifting channels of the Rakaia River, flowing from glaciers in New Zealand's Southern Alps, form a braided pattern that is constantly changing form.

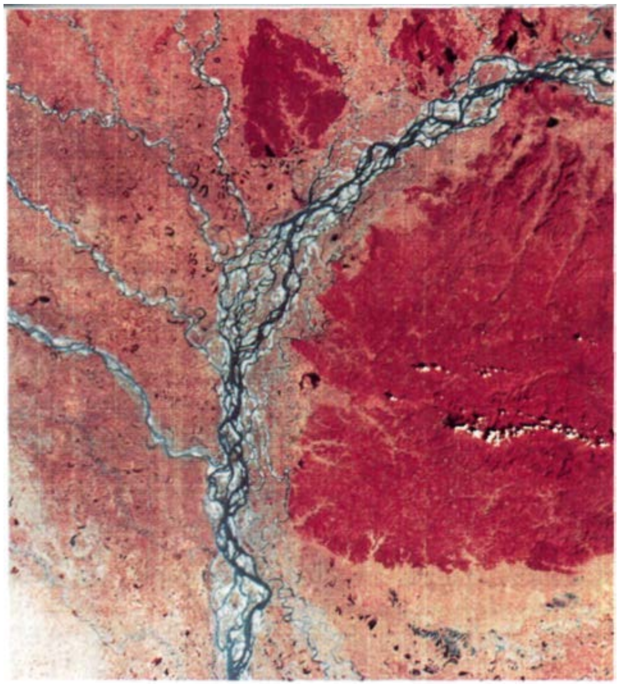


Figure 9.8 Intricate braided pattern of the Brahmaputra River where it flows out of the Himalaya en route to the Indian Ocean. Noted for its huge sediment load, the river is as wide as 8 km during the rainy monsoon season.

the surface sediment may be reworked by the laterally shifting channels.

A braided pattern tends to form in streams having highly variable discharge and easily erodible banks that can supply abundant sediment load to the channel system. Streams of meltwater issuing from glaciers typically have a braided pattern because the discharge varies both daily and seasonally, and the glacier supplies the stream with large quantities of sediment. The braided pattern, therefore, seems to represent an adjustment by which a stream becomes more efficient in transporting an overabundance of sediment.

The Stream's Load

A stream's sediment load consists of two parts. The first part is the coarse particles that move along the stream bed (the **bed load),** while the second is the fine particles that are suspended in the water (the **suspended load).** In addition, streams carry dissolved substances (the **dissolved load)** that are chiefly a product of chemical decomposition of exposed rock, as well as suspended or floating organic debris.

Bed Load

The bed load generally amounts to between 5 and 50 percent of the total sediment load of most streams. The average rate at which bed-load particles move is less than that of the water, for the particles are not in constant motion. Instead, they move discontinuously by rolling or sliding. Where forces are sufficient to lift a particle off the stream bed, the particle may move short distances by saltation, a motion that is intermediate between suspension and rolling or sliding. **Saltation** involves the progressive forward movement of a particle that travels in short, intermittent jumps along arcuate paths (Fig. 99). Saltation continues as long as currents are sufficiently turbulent to lift particles and carry them downstream.

The distribution of bed-load sediment in a stream is related to the distribution of water velocity within the channel (Fig. 9.10). Coarse-grained sediment is concentrated where the velocity is high, whereas finer grained sediment is relegated to zones of progressively lower velocity.

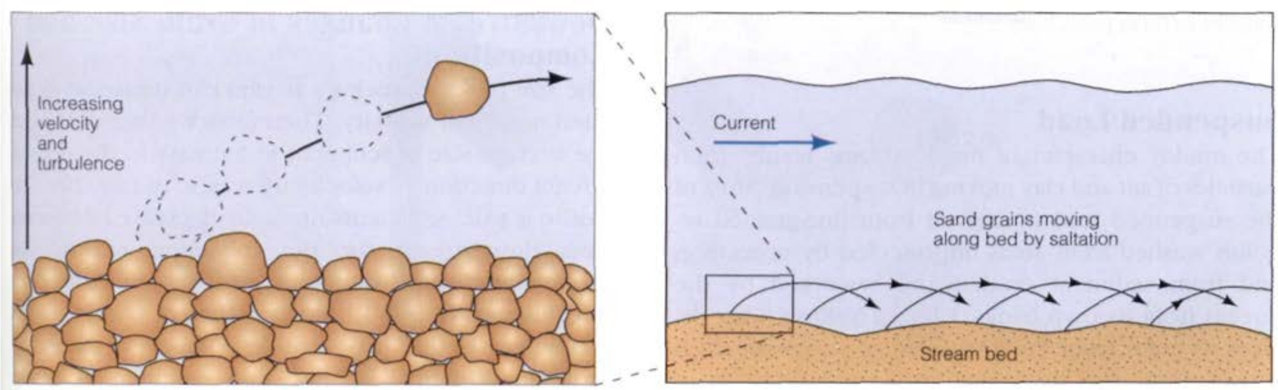


Figure 9.9 A sandy bed load moves by saltation when sand grains are carried up into a stream at places where turbulence locally reaches the bottom or where suspended grains impact other grains on the bed. Once raised into the flowing water, the grains are transported along arc-shaped trajectories as gravity pulls them toward the stream bed where they impact other particles which, in turn, are set in motion.

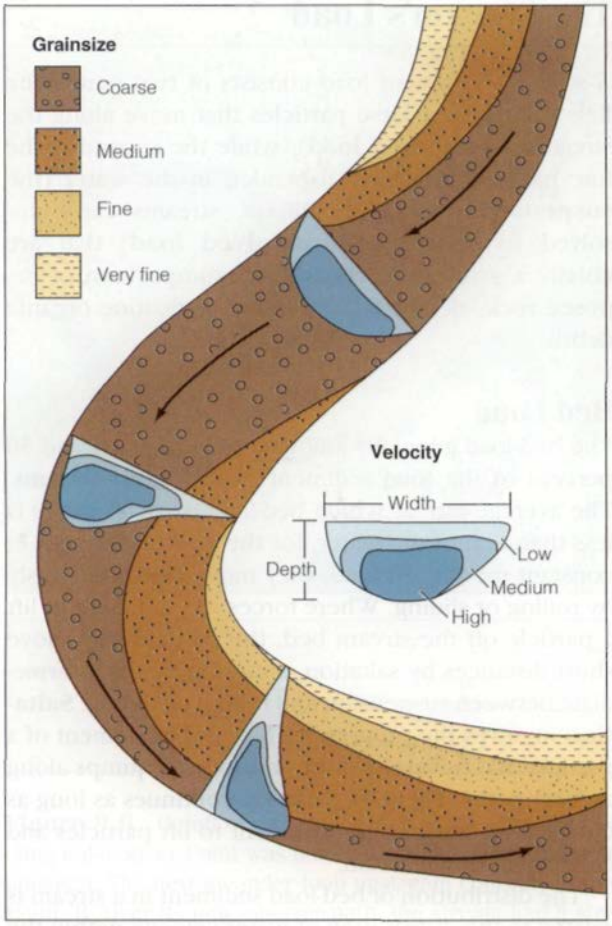


Figure 9.10 Relationship between bed-load grain size and velocity in a section of meandering channel. The coarsest sediment is associated with the zone of highest velocity; on the outside of a bend, both coarse grains and fast-moving water lie adjacent to the stream bank, but between bends both lie in the center of the channel. The finest sediment is associated with the zone of lowest velocity which lies on the inside of a bend. At such places, sediment accumulates to form point bars.

Suspended Load

The muddy character of many streams results from particles of silt and clay moving in suspension. Most of the suspended load is derived from fine-grained regolith washed from areas unprotected by vegetation and from sediment eroded and reworked by the stream from its own banks. China's Yellow River derives its color from the great load of yellowish silt it erodes from the thick unconsolidated deposits of wind-blown dust that cover much of its basin (Fig. 9.11).

Because upward-moving currents within a turbulent stream exceed the velocity at which particles of silt and clay can settle toward the bed under the pull of gravity, such particles tend to remain in suspension

Figure 9.11 A large suspended load, eroded from extensive deposits of wind-blown silt, gives the Yellow River a very muddy appearance and its English name.

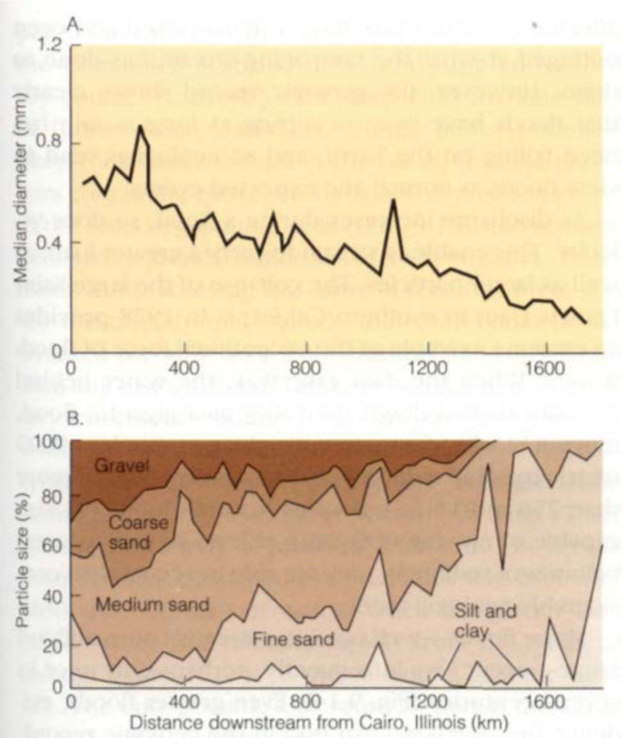
longer than they would in nonturbulent waters. They settle and are deposited only where velocity decreases and turbulence ceases, as in a lake or in the sea.

Dissolved Load

Even the clearest stream water contains dissolved substances that constitute part of its load. Seven ions comprise the bulk of the dissolved content of most rivers: bicarbonate ($(HCO_3)^-$), calcium (Ca^{2+}), sulfate ($(SO_4)^{2-}$), chloride (Cl^-), sodium (Na^+), magnesium (Mg^{2+}), and potassium (K^+). Although in some streams the dissolved load may represent only a small percentage of the total load, in others it amounts to more than half. Streams that receive large contributions of underground water generally have higher dissolved loads than those whose water comes mainly from surface runoff.

Downstream Changes in Grain Size and Composition

The size of the particles a stream can transport is related mainly to velocity. Therefore, we might expect the average size of sediment to increase in the downstream direction as velocity increases. In fact, the opposite is true: sediments normally decrease in coarseness downstream for the following reason. In mountainous headwaters of large rivers, tributary streams mostly flow through channels floored with coarse gravel that may include boulders a meter or more in diameter. Because fine sediment is easily moved, even by streams having low discharge, it is readily carried away by small mountain streams, leaving the coarser sediment behind. Through time, the coarse bed load is gradually reduced in size by abrasion and impact as it moves slowly along. When the stream eventually reaches the sea, its bed load may



consist mainly of sediment no coarser than sand. We can see such a progressive downstream change in sediment size along the channel of the Mississippi River below Cairo, Illinois (Fig. 9.12).

Because large streams generally cross a variety of exposed rocks, the composition of the load they carry changes along the drainage system in response to changes in the lithology of newly introduced sediment. The Nile River provides a good example. Flowing through lower Egypt toward its delta, the Main Nile includes water contributed by three major tributaries: the White Nile, the Blue Nile, and the Atbara (Fig. 9.13). The White Nile, which contributes nearly a third of the total discharge, is responsible for only 3

Figure 9.12 Change in sediment size along the Mississippi River downstream from Cairo, Illinois. A. Over the lower 1600 km of the channel, median diameter decreases from 0.8 to 0.2 mm. B. At Cairo, about 40 percent of the stream's load is gravel, 50 percent is sand, and 10 percent is silt and clay. At a point 1600 miles downstream, the sediment is almost entirely fine sand, silt, and clay.

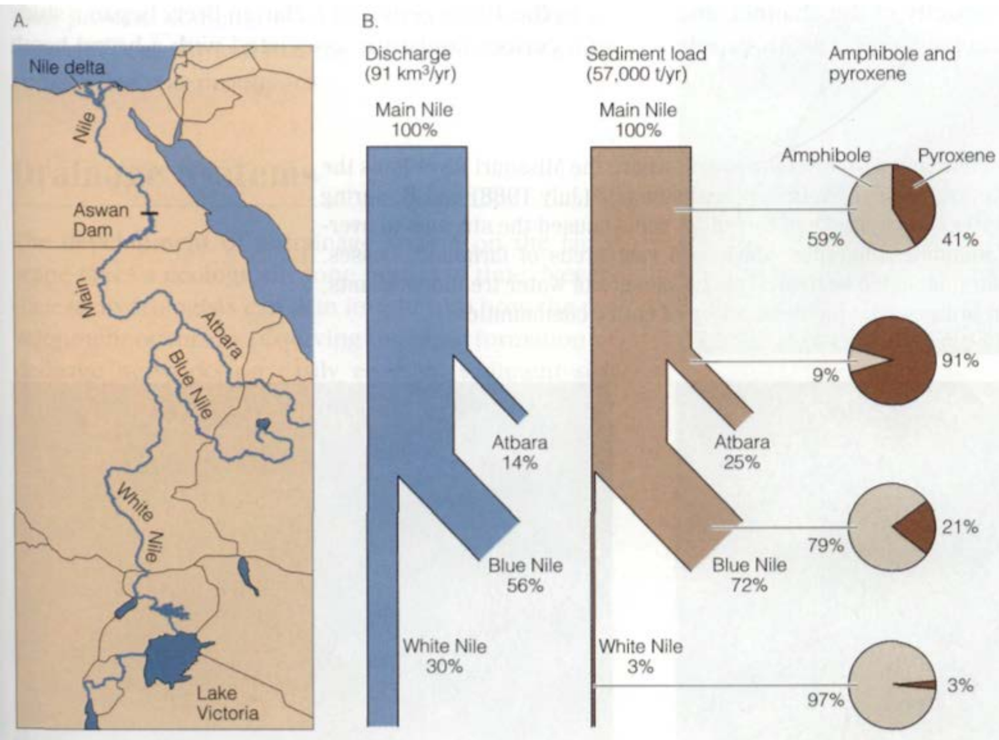


Figure 9.13 Change in sediment composition along the Nile River. A. Map of the Nile River and its principal tributaries. B. Discharge, load, and amphibole/pyroxene ratio of the Main Nile and its principal tributaries. The White Nile, which flows from Lake Victoria, contributes less than a third of the Nile's discharge and only 3 percent of its load. The greatest percentage of discharge and load is supplied by the Blue Nile, which originates in the highlands of Ethiopia. Different percentages of minerals are contributed by each tributary, causing a change in sediment composition of the Main Nile as each tributary joins it.

percent of the sediment load in the Main Nile. In the mineral component of this load, the ratio of amphibole (eroded from metamorphic bedrock of the central African Plateau) to pyroxene is 97:3. The Blue Nile, which drains the highlands of Ethiopia, contributes more than half the discharge and nearly three quarters of the bed load. Its amphibole to pyroxene ratio is 79:21, reflecting the volcanic character of the source region. The more northerly Atbara contributes 14 percent of the discharge and a quarter of the load. In this stream, pyroxene is abundant, and the amphibole to pyroxene ratio is 9:91. These differing mineral components mix together as they enter the Main Nile, resulting in an amphibole to pyroxene ratio of 59:41. This ratio, as we might expect, largely reflects the major contribution of amphibole-rich sediment from the two southerly tributaries, which together account for 82 percent of the load of the Main Nile.

Floods

The uneven distribution of rainfall through the year causes many streams to rise seasonally in flood A *flood* occurs when a stream's discharge becomes so great that it exceeds the capacity of the channel, and water overflows the stream banks (Fig. 9-14). People affected by floods are frequently surprised and even outraged at what the rampaging stream has done to them. However, the geologic record shows clearly that floods have been occurring as long as rain has been falling on the Earth, and so geologists tend to view floods as normal and expected events.

As discharge increases during a flood, so does velocity. This enables a stream to carry a greater load, as well as larger particles. The collapse of the large Saint Francis Dam in southern California in 1928 provides an extreme example of the exceptional force of floodwaters. When the dam gave way, the water behind the dam rushed down the valley as a gigantic flood, moving blocks of concrete weighing as much as 9000 metric tons (20 million lbs) through distances of more than 750 m (0.5 mi). Because natural floods are also capable of moving very large objects as well as great volumes of sediment, they are able to accomplish considerable geologic work.

Major floods—well outside a stream's normal flood range—occur very infrequently, perhaps only once in several centuries (Fig. 9.14). Even greater floods, evidence for which we can find in the geologic record, can be viewed as catastrophic events that occur very rarely even on geologic time scales.

In the 1920s geologist J. Harlen Bretz began a study of a curious landscape associated with a broad basalt

Figure 9.14 A pair of satellite images shows the region where the Missouri River joins the Mississippi River near St. Louis, Missouri A. in a typical summer (July 1988) and B. during the disastrous flood of July 1993 when weeks of torrential rains caused the streams to overflow protective levees and inundate numerous towns and vast areas of farmland. Losses, amounting to billions of dollars, included destroyed crops, closure of water treatment plants, severely damaged roads and bridges, and the destruction of entire communities.

A.

B.

plateau in eastern Washington State and locally called the Channeled Scabland. This landscape consists of dry coulees (canyons) with steep cliffs marking sites of former huge waterfalls, plunge pools, potholes, deep rock basins carved in the basalt, massive gravel bars containing enormous boulders, deposits of gravel in the form of gigantic ripples (Fig. 9.15), and scoured land that extends hundreds of meters above valley floors. After studying this array of features, Bretz was led inescapably to conclude that they could be accounted for only by a truly gigantic flood, far larger than any historic flood. Later, the source of the necessary enormous volume of floodwater was resolved with the discovery that the continental ice sheet covering western Canada during the last glaciation dammed the Clark Fork River, thereby creating a huge lake in the vicinity of Missoula, Montana. The ice-impounded lake contained between 2000 and 2500 km^3 (480 and 600 mi^3) of water. When the glacier dam failed, water was released rapidly from the basin, as though a plug had been pulled from a gigantic bathtub. The only possible exit route lay across the Channeled Scabland region and down the Columbia River to the sea. Subsequent studies have shown that the array of features scattered throughout the Scabland region provide dramatic evidence that such floods occurred repeatedly and that the geologic work they accomplished was prodigious.

A.

B.

Figure 9.15 Features attributable to catastrophic flooding in the Columbia Plateau region. A. Huge ripple marks formed by raging floodwaters as they swept around a bend of the Columbia River. Composed of coarse gravel, the ripples are up to several meters high, and their crests are as much as 100 m apart. B. Large boulder transported and deposited by floodwaters beyond the mouth of Grand Coulee, a major channel excavated by successive floods.

Drainage Systems

The development of a drainage system on the landscape takes a geologically long period of time. Nevertheless, hydrologists can gain insight into how the development occurs by observing the rapid formation of drainage networks on easily erodible sediment surfaces in the field and by conducting laboratory experiments.

An example of the first case occurred in August 1959, following an earthquake at Hebgen Lake, near West Yellowstone, Montana. The quake tilted the terrain in such a way that a large, relatively flat area of silt and sand forming the lake bed was suddenly exposed. With the first rain, small drainage systems began to form. Sample areas were surveyed and mapped one and two years after the earthquake. The results showed the same basic geometry that characterizes much larger and older stream systems. The small, newly formed valleys, together with the areas between them, were disposing of the available runoff in a highly systematic way.

A similar result has been obtained experimentally using sprinkler systems that subject large sediment-filled containers to artificial rainfall (Fig. 9.16). As erosion proceeds, the stream network spreads upslope, eventually encompassing the entire basin. In the course of drainage evolution, both the length of each channel segment and the number of tributaries increase.

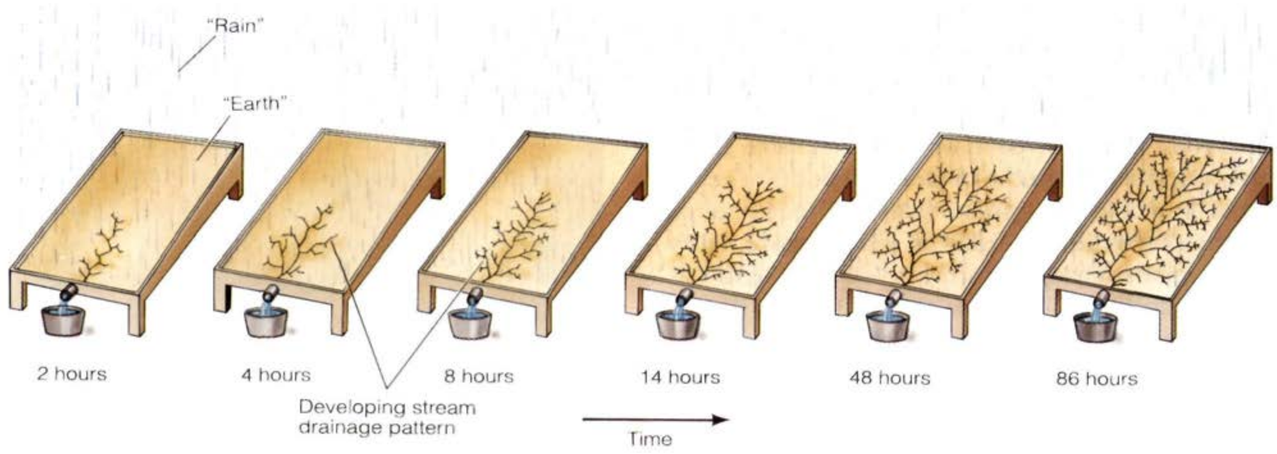


Figure 9.16 Evolution of a drainage network in an experimental rainfall-erosion container. The initial channel, which directed runoff toward the lower end of the container, grew headward and developed new tributaries as it spread to encompass the drainage basin.

As a drainage system develops, details of its pattern change. New tributaries are added, and some old ones are lost due to *stream capture*—the interception and diversion of one stream by another stream that is expanding its basin by erosion in the headward direction. In the process, some stream segments are lengthened and others are shortened. Just as the hydraulic factors within a stream are constantly adjusting to changes, so, too, is the drainage system constantly changing and adjusting as it grows. Like the stream channel, it is a dynamic system tending toward a condition of equilibrium.

GROUNDWATER

Less than 1 percent of the water in the hydrosphere is **groundwater**, defined as all the water contained in spaces within bedrock and regolith. Although the percentage of groundwater in the hydrologic system is small, it is 40 times larger than the volume of all the water in freshwater lakes or flowing in streams and nearly a third as large as the water contained in all the world's glaciers and polar ice. More than half of all groundwater, including most of the water that is usable, occurs within about 750 m (2460 ft) of the Earth's surface. The volume of water in this zone is estimated to be equivalent to a layer of water approximately 55 m (180 ft) thick spread over the world's land areas.

Groundwater operates continuously as a small but integral part of the Earth's water cycle (Fig. 1.2). Part of the water evaporated from the oceans falls on the land as rain, seeps into the ground, and enters the groundwater system. Some of this slowly moving underground water reaches stream channels and contributes to the water they carry to the ocean, where the cycle begins anew.

The Water Table

Much of what we know about the occurrence of groundwater has been learned from the accumulated experience of generations of people who have dug or drilled millions of wells. This experience tells us that a hole penetrating the ground ordinarily passes first into a zone in which open spaces in regolith or bedrock are filled mainly with air (Fig. 9-17). This is the **zone of aeration** (also called the *unsaturated zone*, for although water may be present, it does not saturate the ground). The hole then enters the saturated zone, a zone in which all openings are filled with water. We call the upper surface of the saturated zone the **water table.** Whatever its depth, the water table is a significant surface, for it represents the upper limit of all readily usable groundwater.

In humid regions the water table is a subdued imitation of the ground surface above it. It is high beneath hills and low at valleys because water tends to move toward low points in the topography, where the pressure on it is least. If all rainfall were to cease, the water table would slowly flatten and gradually approach the levels of the valleys. Seepage of water into the ground would diminish, then cease, and the streams in the valleys would dry up. In times of drought, when rain may not fall for several weeks or even months, we can sense the flattening of the water table in the drying up of wells. When that occurs, we know that the water table has fallen to a level below

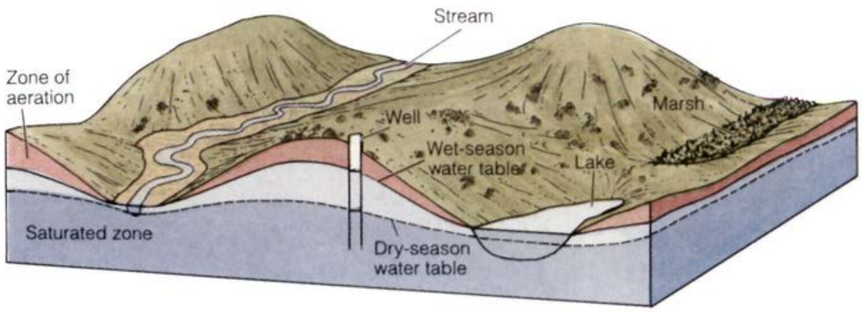


Figure 9.17 In a typical groundwater system, the water table separates the zone of aeration from the saturated zone and fluctuates in level with seasonal changes in precipitation. Corresponding fluctuations are seen in the water level in wells that penetrate the water table. Lakes, marshes, and streams occur where the water table intersects the land surface. In shape, the water table is a subdued imitation of the overlying land surface.

the bottoms of the wells. It is repeated rainfall, dousing the ground with fresh supplies of water, that maintains the water table at a normal level.

Movement of Groundwater

Most of the groundwater within a few hundred meters of the surface is in motion. Unlike the swift flow of rivers, which is measurable in kilometers per hour, groundwater moves so slowly that velocities are expressed in centimeters per day or meters per year. The reason for this contrast is simple. Whereas the water of a stream flows unimpeded through an open channel, groundwater must move through small, constricted passages, often along a tortuous route. Therefore, the flow of groundwater to a large degree depends on the nature of the rock or sediment through which it moves.

Porosity and Permeability

The amount of water that can be contained within a given volume of rock or sediment depends on the **porosity**, the percentage of the total volume of a body of bedrock or regolith that consists of open spaces, or *pores* (Fig. 9.18). In some well-sorted sands and gravels, the porosity may reach 20 percent, while some very porous clays have a porosity as high as 50 percent. Porosity is affected by the sizes, shapes, and arrangement of the rock particles, as well as by the extent to which the pores are filled with cementing substances (Fig. 9.18C). The porosity of igneous and metamorphic rocks is generally low, except when joints and fractures are common.

Permeability is a measure of how easily a solid allows fluids to pass through it. A rock or sediment of very low porosity is also likely to have low permeability. Gravel, with very large openings, is more permeable than sand and can yield large volumes of water. However, high porosity does not necessarily mean a corresponding high permeability, because the size and continuity of the openings influence permeability in an important way. Cement deposited between grains can restrict flow of water between pores, thereby reducing permeability.

Movement in the Saturated Zone

The movement of groundwater in the saturated zone is similar to the flow of water when a saturated sponge is squeezed gently. Water moves slowly through very small open spaces along parallel, thread-like paths. Movement is easiest through the central parts of the spaces but diminishes to zero immediately adjacent to the sides of each space. Normally, flow velocities range between half a meter a day and several meters a year. The highest rate yet measured in the United States, in exceptionally permeable material, was only about 250 m/yr (820 ft/yr).

Responding to gravity, water flows from areas where the water table is high toward areas where it is lowest. In other words, it flows toward surface streams or lakes (Fig. 9.19). Only part of the water travels directly down the slope of the water table by the shortest route. Much of it flows along innumerable long, curving paths that go deeper through the ground. Some of the deeper paths turn upward against the force of gravity and enter the stream or lake from beneath. This happens because the water in the saturated zone at any given altitude is under greater pressure beneath a hill than beneath a stream.

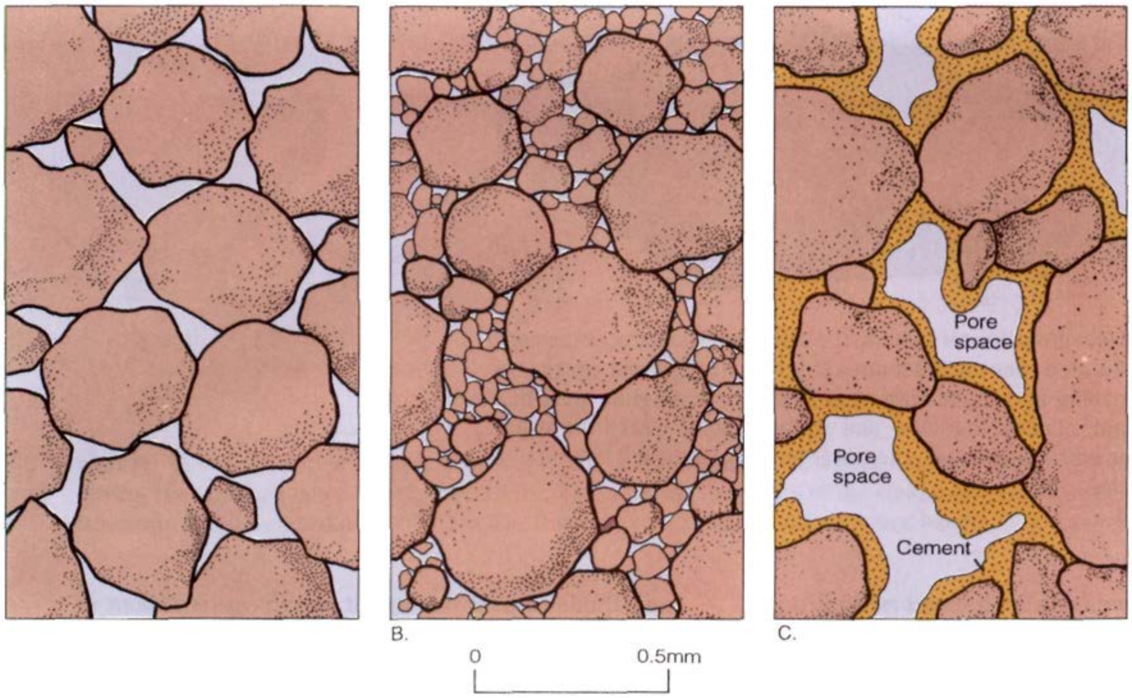


Figure 9.18 Porosity in different sediments. A. A porosity of 30 percent in a reasonably well-sorted sediment. B. A porosity of 15 percent in a poorly sorted sediment in which fine grains fill spaces between larger grains. C. Reduction in porosity in an otherwise very porous sediment due to cement that binds grains together.

The water therefore tends to move toward points where pressure is least. However, most of the groundwater entering a stream travels along shallow paths not far beneath the water table.

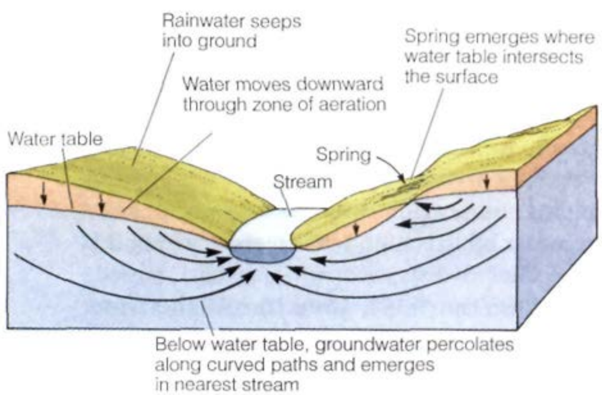


Figure 9.19 Paths of groundwater flow in a humid region in uniformly permeable rock or sediment. Long, curved arrows represent only a few of many possible paths. Springs are located where the water table locally intersects the land surface.

Recharge and Discharge Areas

Replenishment of groundwater, referred to as **recharge,** occurs as rainfall and snowmelt enter the ground in **recharge areas.** These are areas of the landscape where precipitation seeping into the ground reaches the saturated zone (Fig. 9.20). The water moves through the system to **discharge areas**, areas where subsurface water emerges as springs or is discharged to streams or to lakes, ponds, or swamps. The areal extent of recharge areas is invariably larger than that of discharge areas. In humid regions, recharge areas encompass nearly all the landscape except streams and their adjacent floodplains. In more arid regions, recharge occurs mainly in mountains and in alluvial deposits that border them. In such regions, recharge also occurs along channels of major streams that are underlain by permeable alluvium through which water leaks downward and recharges the groundwater.

The time it takes for water to move through the ground from a recharge area to the nearest discharge area depends on rates of flow and the travel distance.

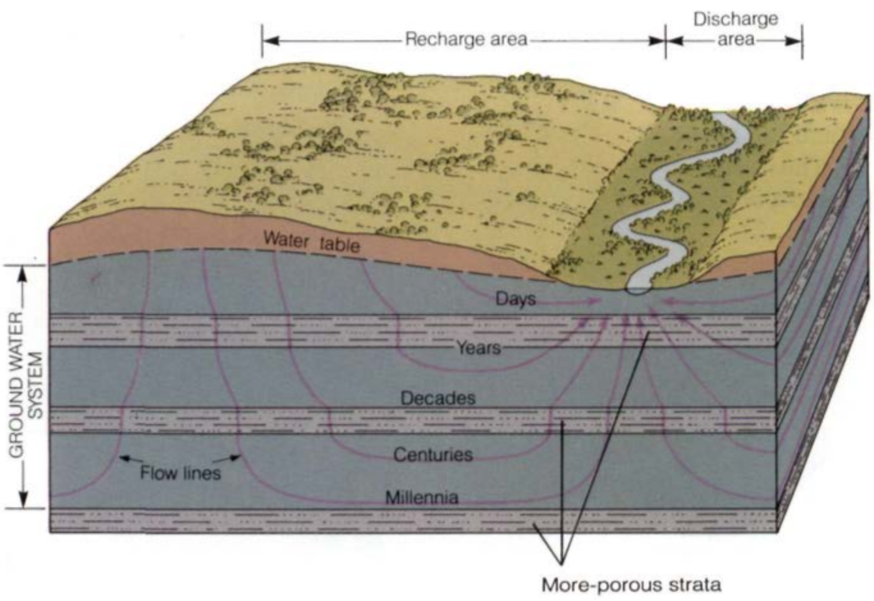


Figure 9.20 Distribution of recharge and discharge areas in a humid landscape. The purple lines are possible pathways groundwater may take from recharge area to discharge area. The times required along various pathways are labeled and depend on the permeability of the rock or regolith along the path and on the distance traveled. Downward and upward movement of groundwater is faster and more direct in the most porous strata.

It may take only a few days, or possibly thousands of years in cases where water moves through the deeper parts of the groundwater system (Fig. 9.20).

Wells

A well supplies water if it intersects the water table. Figure 9.21 shows that a shallow well can become dry when the water table is low, whereas a nearby deeper well may yield water throughout the year.

When water is pumped from a new well, the rate of withdrawal initially exceeds the rate of local groundwater flow. This imbalance in flow rates creates a **cone of depression** in the water table immediately surrounding the well (Fig. 9.21). The locally steepened slope of the water table increases the flow of water to the well. Once the rate of inflow balances the rate of withdrawal, the slope of the water table stabilizes, but it will change if either the rate of pump-

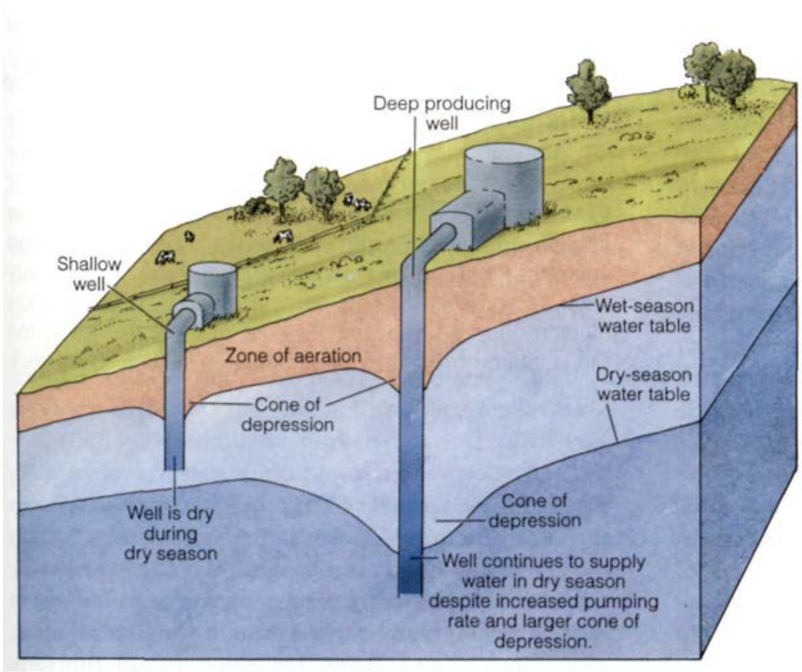


Figure 9.21 Effect of seasonal changes in precipitation on the position of the water table. During the wet season, recharge is high and the hydraulic gradient is relatively steep, so that water is present both in a shallow well and in a deeper well upslope. During the dry season, the water table falls, the hydraulic gradient decreases, and the shallow well is dry. The deeper well continues to supply water, but increased pumping during the dry season enlarges the cone of depression.

ing or the rate of recharge changes. In most small domestic wells, the cone of depression is hardly discernible. Wells pumped for irrigation and industrial uses, however, withdraw so much water that the cone can become very wide and steep and can lower the water table in all wells of a district. This is an example of human action as part of the biosphere effecting the hydrosphere.

Aquifers

If we wish to find a reliable supply of groundwater, we search for an **aquifer** (Latin for water carrier), a body of rock or regolith sufficiently permeable to conduct economically significant quantities of groundwater to springs or wells. Bodies of gravel and sand generally are good aquifers, for they tend to be highly permeable and often are very extensive. Many sandstones are also good aquifers; however, in some sandstone bodies, a cementing agent between the grains reduces the diameter of the openings, thereby reducing permeability and decreasing their potential as aquifers.

Types of Aquifers

An aquifer that has an upper surface that coincides with the water table, and therefore in places is in direct contact with the atmosphere (at lakes and streams, where the water table intersects the land surface), is called an **unconfined aquifer.** A **confined aquifer** is one that is bounded by bodies of impermeable or distinctly less permeable rock adjacent to an aquifer.

About 30 percent of the groundwater used for irrigation in the United States is obtained from the High Plains aquifer, an unconfined groundwater system that lies at shallow depths beneath the High Plains, to the east of the Rocky Mountains (Fig. 9.22). The aquifer, which is tapped by about 170,000 wells, is the principal source of water for a major agricultural region that encompasses 20 percent of the irrigated land of the country. The aquifer consists of a number of sandy and gravelly rock units between which groundwater can readily flow, and its saturated thickness averages about 65 m (213 ft). The water table slopes gently from west to east, and water flows through the aquifer at an average rate of about 30 cm/day (12 in/day). Recharge comes from precipitation and from seepage from streams.

Development of groundwater for irrigation in the High Plains was spurred by severe regional drought in the 1930s and again in the 1950s. Annual recharge of

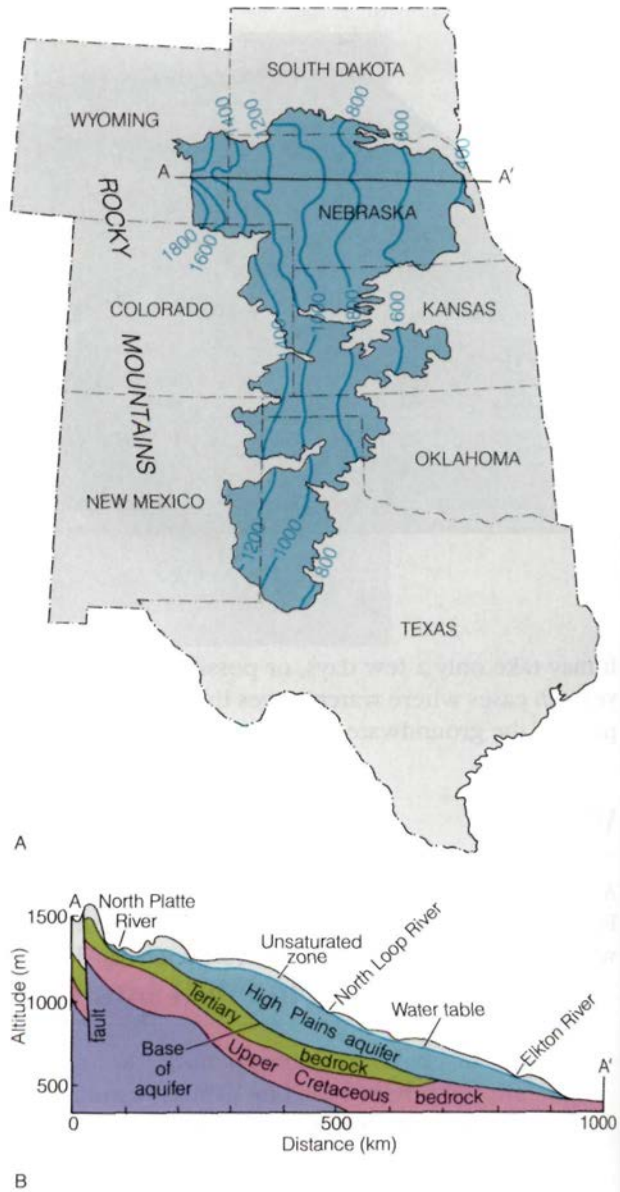


Figure 9.22 The High Plains aquifer, an example of an unconfined aquifer. A. Regional extent of the aquifer and contours (in m) on the water table. Water flow is generally cast, perpendicular to the contour lines. B. Cross section along profile *A-A'* showing the slope of the water table and the relation of the High Plains aquifer to underlying rock units.

the High Plains aquifer from precipitation is much less than the amount of water being withdrawn, and so the inevitable result is a long-term fall of the water table. A dramatic increase in pumping rates has led to serious declines in water level. In parts of Kansas, New Mexico, and Texas, the thickness of the satu-

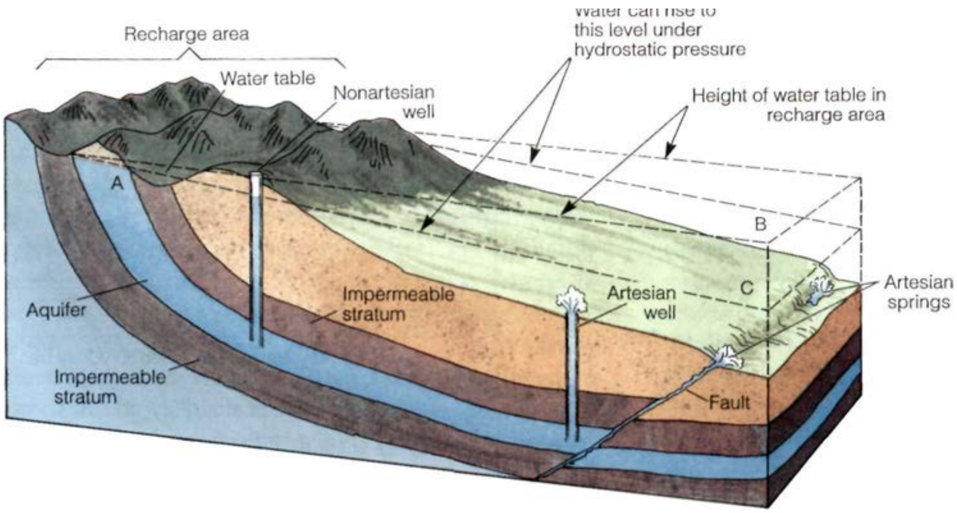


Figure 9.23 Two conditions are necessary for an artesian system: a confined aquifer and water pressure sufficient to make the water in a well rise above the aquifer. The water in a nonartesian well rises to the same height as the water table in the recharge area (line *AB*), minus an amount determined by the loss of energy in friction of percolation. Thus, the water can rise only to the line *AC*, which slopes downward and away from the recharge area. In the artesian well downslope, water flows out at the surface without pumping, for the well top lies below line *AC*.

rated zone has declined by more than 50 percent. The resulting decreased water yield and increased pumping costs have led to major concern about the future of irrigated farming on the High Plains.

Artesian Systems

Water that enters a confined aquifer in an upland recharge area flows downward under the pull of gravity. As it reaches greater depths, the water comes under increasing *hydrostatic pressure* (pressure due to the weight of water at higher levels in the zone of saturation). If a well is drilled to the aquifer, the difference in pressure between the water table in the recharge area and the level of the well intake will cause water to rise in the well. Potentially, the water could rise to the same height as the water table in the recharge area. If the top of the well is lower in altitude than the recharge area, the water will flow out at the surface, without pumping. Such an aquifer is called an **artesian aquifer**, and the well is called an *artesian well* (Fig. 9.23). Similarly, a freely flowing spring supplied by an artesian aquifer is an *artesian spring.* The term *artesian* comes from a French town, Artois (called Artesium by the Romans), where artesian flow was first studied. Under unusually favorable conditions, water pressure can be great enough to create fountains that rise as much as 60 m (200 ft) above ground level.

Chemistry of Groundwater

Analyses of many wells and springs show that the elements and compounds dissolved in groundwater consist mainly of chlorides, sulfates, and bicarbonates of calcium, magnesium, sodium, potassium, and iron. We can trace these substances to the common minerals in the rocks from which they were weathered. As might be expected, the composition of groundwater varies from place to place according to the kind of rock in which it occurs. In much of the central United States, the water is rich in calcium and magnesium bicarbonates that have been dissolved from the local carbonate bedrock. Taking a bath in such water, termed *hard water*, can be frustrating, because soap does not lather easily and a crustlike ring forms in the tub. Hard water also leads to deposition of scaly crusts in water pipes, eventually restricting water flow. By contrast, water that contains little dissolved matter and no appreciable calcium is called *soft water.* With it, we can easily get a nice soapy lather in the shower.

Groundwater flowing through rocks having a high arsenic or lead content may dissolve these toxic elements, making it dangerous to drink. Water circulating through sulfur-rich rocks may contain dissolved hydrogen sulfide (H_2S) which, though harmless to drink, has the disagreeable odor of rotten eggs. In some arid regions, the concentration of dissolved sulfates and chlorides is so great that the groundwater is unusually noxious.

WATER AND PEOPLE

As the human population has grown and become increasingly industrialized, people have altered the natural flow of rivers and communities have generated ever-larger amounts of human and industrial wastes, a good deal of which has inevitably found its way into the very water that people must rely on for their existence. In many places, water is dwindling both in quantity and quality, creating important questions for the communities involved: Will there be enough clean water to sustain future needs? Is its quality adequate for the uses to which we put it? Is the water being used with a minimum of waste? (See the "Guest Essay" at the end of this chapter for more discussion.)

Human Impact on Rivers

Few of the world's large rivers in the densely populated regions of the Earth now flow unrestricted to the sea, and most are contaminated, to various degrees, with human-generated wastes. This human impact has produced serious problems for natural ecosystems, as well as for people who live beside rivers and use their water for personal and industrial purposes.

Hydroelectric Power and Artificial Dams

Hydroelectric power is recovered from the potential energy of water in streams as they flow downslope to the sea. Water ("hydro") power is an expression of solar power because it is the Sun's heat energy that drives the water cycle. That cycle is continuous, and so energy obtained from flowing water is also continuous. Water power has been used in small ways for thousands of years, but only in the twentieth century has it been used widely for generating electricity. All the water flowing in the streams of the world has an estimated 9.2×10^{19} J/yr of recoverable energy, an amount equivalent to burning 15 billion barrels of oil per year. Unlike coal and oil, however, hydro power cannot be used up; it is a renewable resource.

Large artificial dams are being constructed in ever-increasing numbers to provide water storage, electrical power, and flood control. A dam can help modulate the flow of water along a stream system, thereby reducing the impact of peak flow events. The results can be impressive. Before construction of Egypt's Aswan Dam, high discharge during the late summer and early autumn caused extensive annual flooding in the lower Nile drainage basin. The dam, with its vast reservoir (Lake Nasser) backed up behind it, permitted the controlled release of water, thereby markedly reducing the seasonal variability in discharge (Fig. 9.24).

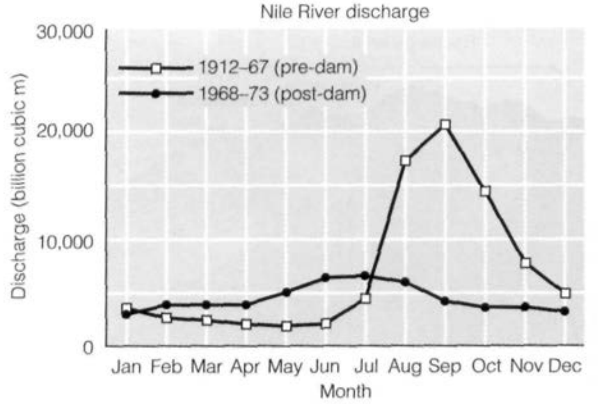


Figure 9.24 Prior to construction of the Aswan Dam, the discharge of the Nile River varied seasonally, with peak discharge coming during the late summer and early fall interval of flooding. Controlled release of water after the dam was built greatly reduced the seasonal variability in discharge.

One major problem with an artificial dam built across a river is that it creates a reservoir in which is trapped nearly all the sediment that the river formerly carried uninterruptedly to the ocean. The accumulating sediment will eventually fill the reservoir, sometimes within less than 200 years, making it useless. In the case of the Aswan Dam, the projected lifetime of the reservoir is about 500 years. Thus, although water power is continuously available, the reservoirs needed for the conversion of water power to electricity have limited lifetimes.

Impacts of Mining and Logging

People have been mining the Earth for mineral resources for millennia, but within the past two centuries mining activity has increased to the point where the impact on natural stream systems has sometimes been dramatic. When gold was discovered in stream channels draining California's Sierra Nevada in the 1840s, the rush for riches led ultimately to the development of hydraulic mining. This technique involves directing water under high pressure against unconsolidated gold-bearing sediments along the stream channels. The sediments are washed into a wooden conduit where the heavy flakes and nuggets of gold are concentrated, while the remaining coarse alluvium is flushed away (Fig. 9.25). However, the unwanted alluvium dumped into the streams was more load than the streams could handle, leading to widespread deposition downstream. The effects ultimately were felt as far away as San Francisco Bay where, in the years between 1850 and 1914, an estimated 1.28 billion cubic meters (45 billion ft^3) of hydraulically mined sediment was deposited in the northeastern arm of the bay. This is at least eight times the amount of sediment excavated during construction of the Panama Canal. The sediment destroyed fish spawning grounds, obstructed navigation, and ultimately reduced the area and volume of the bay, thereby modifying the tidal circulation.

Logging can also have a deleterious effect on natural streams. In the rugged coastal mountains of northwestern United States, logging operations have often involved clearcutting—the removal of all trees within an area, followed by replanting of seedlings to restore the forest. However, one effect of this practice has been a dramatic increase in sediment supplied to streams. Once the trees are felled, the natural protection from erosion provided by trees rooted on steep hillslopes is lost. Both hillslopes and logging roads then become sites of increased runoff and erosion. The end result is stream channels clogged with debris and disruption of natural ecosystems. The impact on salmon populations has been especially serious. Activity in the biosphere effects the rock sphere which in turn effects the biosphere! This has led to new approaches designed to permit both forestry and fisheries to operate fairly and effectively, with minimum disruption of the environment.

Human Impact on Groundwater

Mining Groundwater

In the dry regions of western North America, where streams are few and average discharge is low, groundwater is a major source of water for human consumption. In many of these regions, withdrawal exceeds natural recharge, so that the volume of stored water is diminishing and the water table is falling, often at an increasing rate. In the same way that petroleum is being steadily withdrawn from the most accessible oil pools and minerals are being mined from accessible rocks of the upper crust, groundwater also is being mined. We regard petroleum, coal, and minerals as nonrenewable resources, for they form only over geologically long intervals of time. We don't often stop to think that groundwater can also be a nonrenewable resource. In some regions, natural recharge would take so long to replenish depleted aquifers that formerly vast underground water supplies have essentially been lost to future generations. Even where the problem has been recognized and measures taken to stem the loss, centuries or millennia of natural recharge will be required to return aquifers to their original state.

To halt the fall of the water table, methods have been developed to recharge groundwater artificially. For example, runoff from rainstorms in urban areas that normally would flow away in surface streams can be channeled and collected in basins where it will

Figure 9.25 Hydraulic mining in the foothills of California's Sierra Nevada. Gold-bearing alluvium is washed into wooden channels (sluices) where flakes and nuggets of gold are concentrated. Although this was an efficient way of mining in the years following the Gold Rush, it led to widespread environmental degradation of streams, and its impact was felt as far away as San Francisco Bay.

Figure 9.26 The Leaning Tower of Pisa, Italy, the tilting of which accelerated as groundwater was withdrawn from aquifers to supply the growing city.

seep into permeable strata below, thereby raising the water table. Groundwater withdrawn for nonpolluting industrial use can be pumped back into the ground through injection wells, recharging the saturated zone.

The water pressure in the pores of an aquifer helps support the weight of the overlying rocks or sediments. When groundwater is withdrawn, the pressure is reduced, and the particles of the aquifer shift and settle slightly. As a result, the land surface subsides. The famous Leaning Tower of Pisa, Italy, built on unstable fine-grained alluvial sediments, began to tilt when construction began in 1174 (Fig. 9.26). The tilting increased during the present century as groundwater was withdrawn from deep aquifers. Recent strengthening of the foundation is designed to keep the tower stable in the future, providing groundwater withdrawal is strictly controlled.

Groundwater Contamination

Citizens of modern industrialized nations take it for granted that when they turn on a faucet, safe, drinkable water will flow from the tap. Throughout much of the world, however, the available water is often unfit for human consumption. Not only do natural dissolved substances make some water undrinkable, but also many water supplies have become severely contaminated by human and industrial wastes.

The most common source of water pollution in wells and springs is sewage. Drainage from septic tanks, broken sewers, privies, and barnyards contaminates groundwater. If water contaminated with sewage bacteria passes through sediment or rock with large openings, such as very coarse gravel or cav-

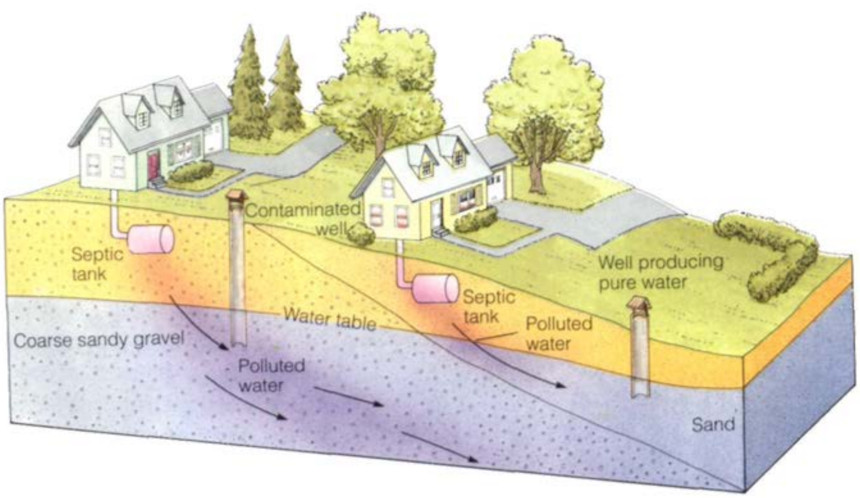


Figure 9.27 Purification of groundwater contaminated by sewage. Pollutants seeping through a highly permeable sandy gravel contaminate the groundwater and enter a well downslope from the source of contamination. Similar pollutants moving through permeable fine sand higher in the stratigraphic section are removed after traveling a relatively short distance and do not reach a well downslope.

ernous limestone, it can travel long distances and remain polluted (Fig. 9.27). On the other hand, if the contaminated water seeps through sand or permeable sandstone, it can become purified within short distances, in some cases in less than about 30 m (30 yd). Sand is ideal for cleaning up polluted water, for it promotes purification by (1) mechanically filtering out bacteria (water gets through, but most of the bacteria do not); (2) oxidizing bacteria so they are rendered harmless; and (3) containing other organisms that consume the bacteria. For this reason, purification plants that treat municipal water supplies and sewage pass these fluids through sand. Although natural purification is as effective, it occurs much more slowly than in a water treatment plant.

Toxic Wastes and Agricultural Poisons

Vast quantities of human garbage and industrial wastes are deposited each year in open basins or excavations at the land surface. When a landfill site reaches its capacity, it generally is covered with earth and revegetated. Many of the waste products, now underground, are mobilized by precipitation that percolates though the site, carrying away soluble substances. In this way, harmful chemicals slowly leach into groundwater reservoirs and contaminate them, potentially making them unfit for human use. The pollutants travel from landfill sites as plumes of contaminated water in directions that depend on the regional groundwater flow pattern, and they are dispersed at the same rates as the moving water (Fig. 9.28). The pollutants often are toxic not only to humans but also to plants and animals in the natural environment.

In the United States, the pollution problems associated with landfill wastes have become so severe that the government has begun a major long-term effort (the Superfund Program) to clean up such sites and render them environmentally safe. However, the identified sites number in the tens of thousands, and it is difficult to judge how much time and money ultimately will be required to accomplish the task.

Yet another hazard is that posed by chemicals. Each year pesticides and herbicides are sprayed over agricultural fields to help improve quality and productivity. However, some of the chemicals have been linked with cancers and birth defects in humans, and some have led to disastrous population declines of wild animals. Because of the manner in which they are spread, the toxic chemicals invade the groundwater system over wide areas as precipitation flushes them into the soil.

Underground Storage of Hazardous Wastes

One of the leading environmental concerns of industrialized countries is the necessity of dealing with highly toxic or radioactive waste products. Experi-

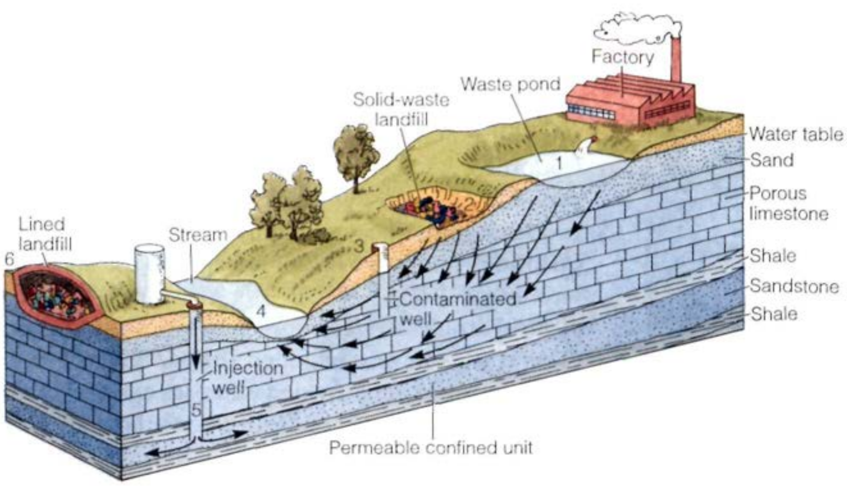


Figure 9.28 A groundwater system contaminated by toxic wastes. Toxic chemicals in an open waste pond (1) and an unlined landfill (2) percolate downward and contaminate an underlying aquifer. Also contaminated are a well downslope (3) and a stream (4) at the base of the hill. Safer, alternative approaches to waste management include injection of waste into a deep, confined rock unit *(5)* that lies well below aquifers used for water supplies, and a carefully engineered surface landfill (6) that is fully lined to prevent downward seepage of wastes. Because neither of the latter approaches is completely foolproof, constant monitoring at both sites would be required.

Guest Essay

River Aesthetics—A Janus Perspective

W. R. Osterkamp is a geomorphologist with the National Research Program, U.S. Geological Survey. He has undergraduate degrees in geology and chemistry from the University of Colorado, and an M.S. and Ph.D. in geology from the University of Arizona. His principal research interests include stream-channel dynamics, geomorphic-vegetative relations, and methods for tracing the movement and storage of sediment.

Janus, the Roman god of doors and gates, had two faces, allowing him to look in opposite directions. His namesake month, January, affords views of both the past and the future. Nondeities among us, and society in general, tend not to concentrate their attention on Janus's multidirectional manner. Historians look to the past, whereas developers and economists perhaps look to the future. However, land regulators, aquatic ecologists, and many others involved in the conflict between use and aesthetics of drainage basins and their river systems need the eyes of Janus to see both back and ahead—to remember and revere the unspoiled, and to recognize and minimize the inevitable encroachments into riparian habitats by an expanding human population.

Aesthetics is a concept of personal preference, a judgment, sense, or love of beauty. The qualities that render a landscape or a river aesthetically rewarding to someone vary with experiences. My sense of aesthetics, as applied to natural systems, tempers the idealism of childhood perceptions—perhaps an innate recognition of beauty—with the experiences and awarenesses of an earth scientist. As it did for Thoreau, it includes a spiritual emotion akin to holism, an awareness of complete systems of interdependent parts as opposed to a concentrated attention on an individual or isolated part of the system. I look back and cherish the natural wonders I perceived as a child; I look forward and question whether those perceptions, if faithful, can be experienced again. Despite apprehensions, I remain optimistic that the sights of a youthful Janus can be compatible with his seasoned view, encompassing the realities of modern society.

Why the optimism? Because a stream is a highly dynamic core of a much larger, more complex organism: the drainage basin. In *Le Phenomene Humain,* the French geologist/cleric, Pierre Teilhard de Chardin, likened human society to an evolving organism, one whose history necessarily helps direct its future. The

ence has demonstrated that surface dumping quickly leads to contamination of surface and subsurface water supplies and thereby to the possibility of serious and potentially fatal health problems (Fig. 9.29)- Countries with nuclear capacity have the special problem of disposing of high-level radioactive waste products, substances so highly toxic that even minute quantities can prove fatal if released to the surface environment.

Most studies concerning disposal of toxic and nuclear wastes have concluded that underground storage is appropriate, provided safe sites can be found. In the case of high-level nuclear wastes that can remain dangerous for tens or hundreds of thousands of years, a primary requirement is a site that will be stable over very long time intervals. The only completely safe sites would be those where waste products and their containers would not be affected chemically by groundwater, physically by natural deformation such as earthquakes, or accidentally by people.

Figure 9.29 Toxic substances in this New Jersey waste pond could constitute a serious health hazard for nearby residents if the fluid wastes leak into the groundwater system.

interplay of physical and biological processes in a drainage basin in turn permits analogy to a discrete, active organism, one that is most dynamic where water and energy are concentrated. When one part of the organism is injured, regardless of cause, natural or induced, the system adjusts and heals, and stream channels, the arteries of the organism, generally show the greatest and fastest recuperative powers. To me, the beauty of a stream, its aesthetics, is displayed by its vibrance, its energy, a life inseparable from and nourished by all other parts of the drainage basin or watershed.

This organism is exceedingly complex but is conceptualized by earth scientists as elements or interacting systems of soil and rock, water and air, plants and animals. If one element suffers disturbance, all elements suffer disturbance. Recovery by one system is shared by recovery of all systems. But as with other organisms, healing requires circulation. And it is the arteries of circulation that most enrich and become enriched, conveying nutrients and toxins.

Terms typically employed to describe the aesthetic quality of a stream and its adjacent riches seem inappropriate when applied to a scientific view of a watershed-organism. An aesthetic view of a stream may express ideal qualities based on beauty, but an aesthetically pristine, timeless, or unrestrained stream constantly faces real threats to its health. Objective measures of the character of a stream channel are easily collected, but the beauty, not utility, of the channel must remain ephemeral if other parts of the organism become diseased. A river distributes nourishment, and it discharges wastes, both of which are derived from other parts of the being. Its health rests in the balance. Floods, fires, infections, and disturbances are components of the history of any drainage basin, and none is timeless. No organism is free of wastes or decay, and thus none is pristine.

A drainage basin and its streams, like all organisms, is dynamic and is constantly changing according to the energy available and stresses imposed on it. If the basin that feeds the rivers is healthy, its component systems unstressed and adjusted to each other, the stream, too, is healthy—it has beauty and aesthetic value. In mimicking Janus, we must not lose sight of our highly individualized perceptions and the qualities that have always been associated with the beauty of natural flowing water. Similarly, as scientifically literate members of society, we must look ahead and realize that change induced anywhere in a drainage basin induces change everywhere. Conservation practices in all parts of a drainage basin are necessary if we want to minimize harmful human impacts and aesthetic impairment to the select circulatory parts of the drainage organisms. Perhaps the effects of conservation are not apparent immediately, or within a human life span, but ultimately change and balance will be assured. Janus, or society, must continue to see clearly in all directions if both the health of stream channels and the aesthetic value such health provides are truly to be sustained.

Summary

1. Streams are the chief means by which water returns from the land to the sea. They help shape the Earth's surface and transport sediment to the oceans.
2. A drainage basin encompasses the area supplying water to the stream system that drains it. Its area is closely related to the stream's length and annual discharge.
3. Discharge, velocity, and channel cross-sectional area are interrelated, so that when discharge changes, the product of the other factors also changes to restore equilibrium. As discharge increases downstream, channel width and depth increase, and velocity increases slightly.
4. Straight channels are rare. Meandering channels commonly form on gentle slopes and where sediment load is small to moderate. Braided patterns develop in streams that have variable discharge and a large bedload.
5. Stream load is the sum of bed load, suspended load, and dissolved load. Bed load comprises as much as 50 percent of the total load and moves by rolling, sliding, and saltation. Most suspended load is derived from erosion of fine-grained regolith or from stream banks. Streams that receive large contributions of underground water commonly have higher dissolved loads than those deriving their discharge principally from surface runoff.
6. The average grain size of a stream's load tends to decrease downstream as particles are reduced in size by abrasion and as the sediment is sorted by weight. Changes in composition of a stream's load commonly reflect changes in the lithology of rocks across which the stream flows.
7. Floods result when discharge exceeds the capacity of a stream's channel. Streams experiencing large floods are capable of transporting great loads of sediment as well as very large boulders. Exceptional floods, well outside a stream's normal range, occur very infrequently.
8. Although water may be present in the zone of

aeration, it does not saturate the ground. In the underlying saturated zone all openings are filled with water. The water table marks the boundary between the zone of aeration and the saturated zone and in humid regions is a subdued imitation of the ground surface above it.

9. Groundwater flows far more slowly than the water in surface streams, normally at rates between half a meter a day and several meters a year. With constant permeability, the rate of flow increases as the slope of the water table increases.
10. In moist regions, groundwater seeps downward under the pull of gravity. It moves away from hills toward valleys, where it may emerge to supply streams.
11. Groundwater is replenished in recharge areas and moves downward and laterally to emerge in discharge areas where it emerges as springs, streams, lakes, ponds, or swamps.
12. Groundwater flows into most wells directly by gravity, but into artesian wells under hydrostatic pressure. The pumping of water from wells creates cones of depression in the water table.
13. Major supplies of groundwater are found in aquifers, among the most productive of which are porous sand, gravel, and sandstone. The water table defines the upper surface of an unconfined aquifer, whereas a confined aquifer is bounded by bodies of impermeable or distinctly less permeable rock adjacent to an aquifer.
14. Because of high hydrostatic pressure, water of an artesian aquifer may flow freely out the top of a well that is lower than the recharge area.
15. Artificial dams are built across stream channels to increase water supplies for irrigation, harness hydroelectric power, and help control floods, but accumulation of sediment behind a dam reduces the useful lifetime of a reservoir.
16. Hydraulic mining and clearcut logging can lead to increased sedimentation along stream channels that adversely affects natural ecosystems.
17. Excess withdrawal of groundwater can lead to lowering of the water table and to land subsidence. In some cases, natural replenishment to original levels would take centuries or even millennia.
18. Water quality is influenced by natural dissolved substances and human, agricultural, and industrial pollutants that seep into groundwater reservoirs.
19. Extremely toxic and radioactive wastes pose special problems. They should be stored underground only if geologic conditions imply little or no change in groundwater systems over geologically long intervals of time.

Important Terms to Remember

alluvium (p. 227)
aquifer (p. 240)
artesian aquifer (p. 241)
bed load (p. 231)
braided stream (p. 230)
channel (p. 226)
cone of depression (p. 239)
confined aquifer (p. 240)
discharge (p. 227)
discharge area (p. 238)
dissolved load (p. 231)
divide (p. 227)
drainage basin (p. 227)
gradient (p. 227)
groundwater (p. 236)
load (p. 227)
meander (p. 228)
permeability (p. 237)
porosity (p. 237)
recharge (p. 238)
recharge area (p. 238)
runoff (p. 226)
saltation (p. 231)
saturated zone (p. 236)
stream (p. 226)
suspended load (p. 231)
tributary (p. 228)
unconfined aquifer (p. 240)
water table (p. 236)
zone of aeration (p. 236)

Questions for Review

1. How does streamflow differ from overland flow, and how are these two kinds of surface flow related?
2. How do a stream's channel dimensions (depth, width) and velocity adjust to changes in discharge?
3. Why does stream velocity generally increase downstream, despite a decrease in stream gradient?
4. Studies of meandering streams show that a relationship exists between the width of a channel and the diameter of meander loops: the larger the channel width, the larger the loop diameter. Why should this be true?
5. Why, if velocity increases downstream, does sediment in transport typically decrease in size in that direction?
6. If you were to collect samples of sandy, gravelly alluvium along the length of a large stream and determine the composition of the gravel and mineralogy of the sand, you would likely see obvious changes in these properties in the downstream direction. What factor or factors might explain such changes?
7. What ultimately limits the period of time that an artificial dam can provide useful levels of hydroelectric power?'
8. Why do the flow paths of groundwater moving beneath a hill tend to turn upward toward a stream in an adjacent valley?
9. What determines how long it takes water to pass from a recharge area to a discharge area?
10. What causes a cone of depression to form around a producing well?
11. Why are sandstones generally better aquifers than siltstones or shales?
12. For what reason does water rise to or above the ground surface in an artesian well?
13. Why is sand especially effective in purifying water flowing through it?
14. What is the origin of "hard" water in regions of carbonate bedrock?

Questions for Discussion

1. Let's assume you are considering the purchase of a 30-year-old house adjacent to a large river that experiences seasonal flooding. What information might you seek out to evaluate the possibility that an unusually large flood could inundate the house sometime during the next half century?
2. Find out where your community or city derives its supply of fresh water. Will the existing source of supply prove adequate if the population doubles during the next several decades? If not, what other potential sources exist?
3. A large area on a hillside has been suggested as a landfill site for garbage generated by a small nearby city. You are asked for a geologic appraisal of the site to determine if local subsurface water supplies might be affected. What geologic factors would you investigate and why?
4. What geologic and biologic factors are likely to disqualify a site from being selected for underground storage of high-level radioactive waste? Do potential sites exist in your state that could prove suitable on geologic grounds?

CHAPTER 10

The World of Snow and Ice

Two icebergs neat Antarctica have calved off the margin of a nearby ice shelf. Scientists visualize towing much larger bergs to supply fresh water to arid lands in Australia, South America, southwestern United States, and the Middle East.

Glacial Water for Arid Lands

As world population increases and our insatiable demand for fresh water continues to rise, a critical problem we face is how to meet this demand. Some imaginative people think the answer may lie in a vast, unutilized resource: Antarctic icebergs.

Large icebergs are plentiful around Antarctica, where an estimated 1000 km^3 (240 mi^3) of glacier ice breaks off as icebergs each year. The English explorer James Cook (1728-1779) was among the first to recognize the potential for using the fresh water locked up in these icebergs. In 1773, while his ship lay off the coast of Antarctica, his sailors hoisted 15 tons of berg ice on board, and Cook noted in his log that "this is the most expeditious way of watering I ever met with."

Icebergs also occur in Arctic waters, but they are far less numerous and on average much smaller than Antarctic ones. Today, scientists visualize towing large Antarctic icebergs north to supply water to the arid coasts of the Middle East, South America, Africa, Australia, and the southwestern United States. A French engineering firm, commissioned by Saudi Arabia, has already studied the feasibility of towing huge icebergs north to the Red Sea, where they could supply drinking and agricultural water.

As impressive as such a proposal sounds, the technological problems are formidable. How can blocks of ice hundreds of meters thick and several square kilometers in area be moved thousands of kilometers through warm and often stormy seas and still have enough mass to make the venture economically feasible? At an average speed of 0.5 m/s (1 knot), a huge berg 1 to 2 km (0.6 to 1.2 mi) long would require 70 days to be towed 3000 km (I860 mi) through relatively cool waters to Australia, but the outermost 60 m (200 ft) of ice would likely melt away during the trip. An unprotected iceberg of this size is unlikely to survive the long trip (up to a year) through much warmer waters to southern California or the Middle East. To avoid excessive loss, the ice would have to be insulated against melting and evaporation and protected against collapse and disintegration.

In addition, the towing vessels or propulsion systems large enough to move huge icebergs over great distances have yet to be developed. Although these technological challenges appear immense, the day may come when ice that has been stored in Antarctic glaciers for tens of thousands of years wiL bring life-sustaining waters to the Earth's low-latitude, arid lands.

THE EARTH'S COVER OF SNOW AND ICE

Among the planets of our solar system, only the Earth has a bluish color, imparted by the vast expanse of world ocean. Nevertheless, a polar view of either hemisphere appears largely white because of an extensive, perennial cover of snow and ice. In the northern hemisphere, much of the ice floats as a thin sheet on the Arctic Ocean, whereas in the southern hemisphere it consists of a vast glacier system overlying the continent of Antarctica and adjacent islands and seas as well as floating sea ice that extends far beyond the coastline. The vast mantle of polar ice is linked in important ways to both the oceans and the atmosphere. It is involved in the generation of cold, dense water that drives deep ocean circulation (Chapter 8), and it is an important factor in world climate (Chapter 14). It is also an extremely dynamic part of our planet, for its expanse fluctuates seasonally as well as on geologic time scales.

The part of the Earth's surface that remains perennially frozen constitutes the **cryosphere** *(cold* or *frozen sphere).* It includes not only glaciers and sea ice, but also vast areas of frozen ground that lie beyond the limits of glaciers. At present, glaciers cover about 10 percent of the Earth's land surface, while perennially frozen ground covers an additional 20 percent. Thus, nearly a third of the Earth's land area belongs to the cryosphere.

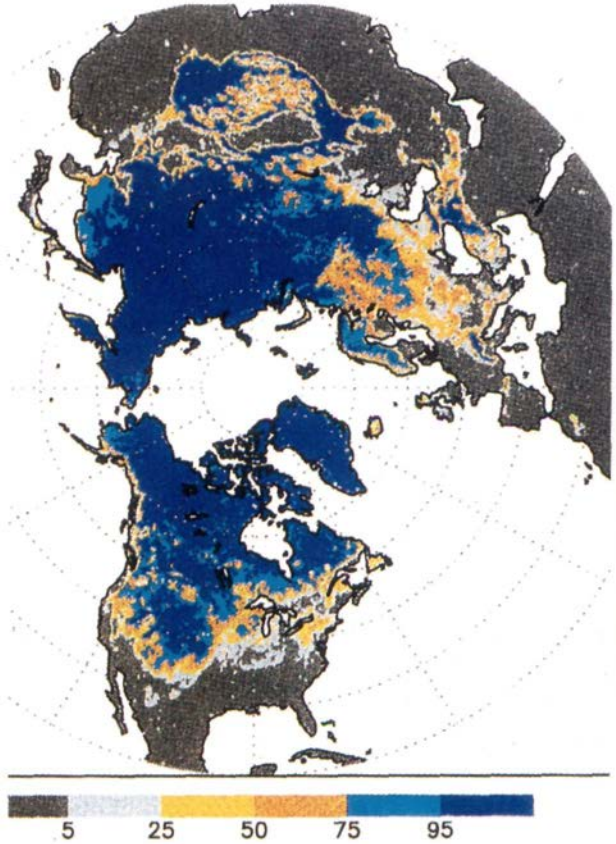


Figure 10.1 A map of average snow cover in the northern hemisphere (expressed as percentage of land area covered by snow) during December, 1992 is based on data received from a microwave sensor aboard an orbiting satellite. Greatest snow cover lies in regions of continental climate in middle to high latitudes (northern North American and northeastern Asia) and high-altitude regions such as the Tibetan Plateau of central Asia.

SNOW AND THE SNOWLINE

The winter snow that must be shoveled from our sidewalks and delays our morning commute to work also plays an important role in the Earth's climate system. Its highly reflective surface bounces sunlight back into space, thereby reducing surface air temperature. When the snow melts, it becomes a major source of water for rivers and moisture for agricultural soils. The timing or amount of snowfall can affect people adversely; for example, a heavy late-winter snowfall can generate widespread flooding, while a snowfall deficit can lead to water rationing during the following summer.

Annual Snow Cycle

Prior to the mid-1960s, estimates of variations in continental snow cover were obtained from limited ground-based measurements. Today, variations in snow depth and area are monitored weekly by satellites.

During a typical year, snow cover in the northern hemisphere first appears in northern Alaska and northeastern Siberia in mid-September to mid-October (Fig. 10.1) Through November, the snow cover expands southward and begins to thicken. By December, the expanding snow cover reaches southern Russia, central Europe, and the northern United States, and snow covers nearly all of the high Tibetan Plateau. From December through March, the snowpack thickens in continental interiors, but its southern limit begins to retreat as air temperature rises. The snowpack then recedes rapidly northward during late spring, and by mid-June the remaining snow is confined mainly to high mountains and lands bordering the cold Arctic Ocean.

The Snowline

If we view a high mountain at the end of the summer, just before the earliest autumn snowfall, we commonly will see a snowy zone on its upper slopes. The lower boundary of this zone is the **snowline,** which is defined as the lower limit of perennial snow (Fig. 10.2). Above the snowline, part of the past winter's

Figure 10.2 The lower limit of snow in late spring forms an irregular line across the flank of Mount Cook, the highest peak in New Zealand's Southern Alps. As the weather warms and the snow melts, the snow limit rises to its highest level at the end of the summer. This late-summer limit marks the annual snowline. Above the snowline, most of the ground remains snow-covered all year.

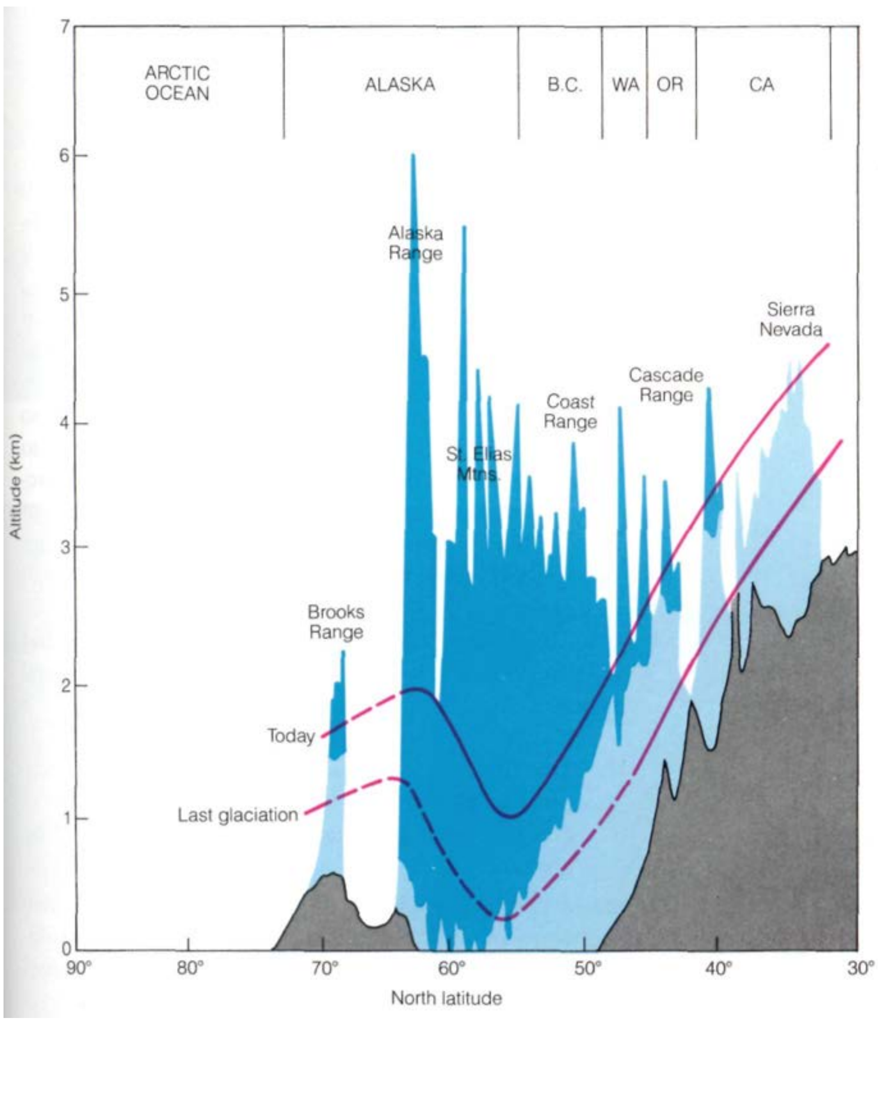


Figure 10.3 A transect along the coastal mountains of western North America from Alaska to Mexico showing the relationship between (a) the present snowline and existing glaciers (blue) and (b) the ice-age snowline and glaciers of that time (light blue).

snow survives the warm temperatures of summer, along with snow from earlier winters that also persisted through past summers. In detail, the snowline is an irregular surface, its shape controlled both by variations in the thickness of the winter snowpack and by local topography. When viewed from a distance, however, the snowline appears as a line delimiting snow-covered land from snow-free land.

The altitude of the snowline and its horizontal position on the landscape commonly change from year to year depending on the weather. While a number of climatic factors are involved, the two principal ones are winter snowfall, which affects total snow accumulation, and summer temperature, which influences melting.

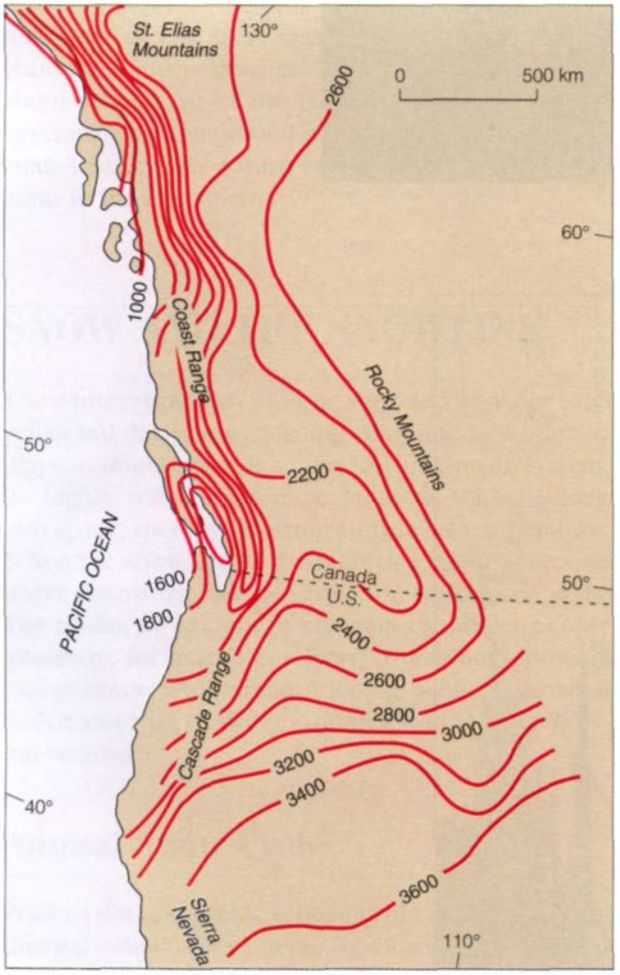


Figure 10.4 Contours (in meters) show the regional altitude of the snowline throughout northwestern United States, British Columbia, and southern Alaska for a representative balance year. The surface defined by the contours rises steeply inland from the Pacific coast in response to increasingly drier climate, and also from north to south in response to progressively higher mean annual temperatures.

In the polar regions, annual snowfall is generally very low because the air is too cold to hold much moisture, but because summer temperatures are also low, little melting occurs. Consequently, the snowline generally lies at low altitudes. Mean summer temperatures increase toward the equator, and so snowline altitude also rises toward the equator, but not uniformly (Fig. 10.3). Because its level is also controlled by precipitation, the snowline rises inland as precipitation decreases away from ocean moisture sources (Fig. 10.4). This effect is especially noticeable as one traces the snowline inland from a midlatitude continental coast. For example, winter snowfall in the Coastal Ranges of southern Alaska and British Columbia, adjacent to the Pacific Ocean source of moisture, is veiy high, and the snowline lies as low as 600 m (1170 ft). However, it rises steeply inland to 2600 m (8530 ft) in the Rocky Mountains as precipitation decreases by more than half. The snowline is highest [in places more than 6000 m (19,700 ft)] on the lofty summits of central Asia, which lie farther from an oceanic source of moisture than any other high mountains.

GLACIERS

Wherever the amount of snow falling each winter is greater than the amount that melts during the following summer, the snow gradually grows thicker. As it accumulates, its increasing weight causes the basal snow to recrystallize to ice. This process is analogous to the way sedimentary strata, buried deep in the Earth's crust, recrystallize to metamorphic rocks. When the accumulating snow and ice become so thick that the pull of gravity causes the frozen mass to move, a glacier is born. Accordingly, we define a **glacier** as a permanent body of ice, consisting largely of recrystallized snow, that shows evidence of either downslope or outward movement owing to the pull of gravity.

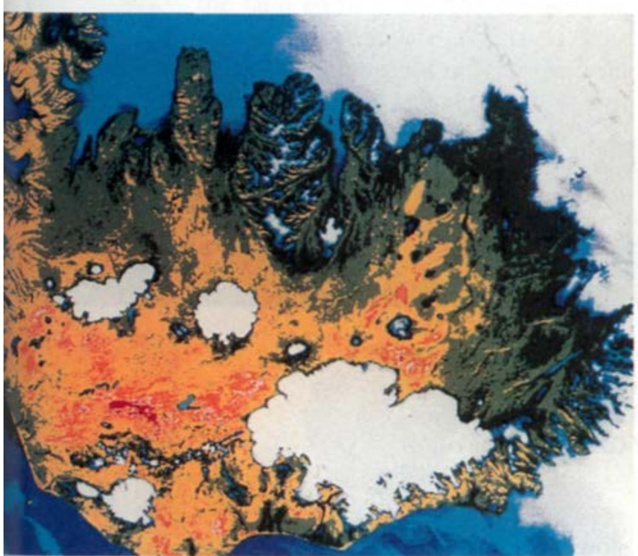
Glaciers vary considerably in shape and size, and on these bases we recognize several fundamental types:

1. The smallest, a *cirque glacier* (Fig. 10.5A), occupies a protected, bowl-shaped depression (a **cirque**) on a mountainside that is produced by glacial erosion.
2. A cirque glacier that expands outward and downward into a valley becomes a *valley glacier* (Fig. 10.5B). Many of the Earth's high mountain ranges

A.

B.

Figure 10.5 A. A small cirque glacier below a mountain summit in Alaska's Denali National Park. B. Dark bands of rock debris delineate the boundaries between adjacent tributary ice streams that merged to form Kaskawulsh Glacier, a large valley glacier in Yukon Territory, Canada. C. Several ice caps cover areas of high land on Iceland. Vatnajokull, in the southeastern part of the island, is the largest ice cap (8300 km^2) and overlies an active volcano.



C.

contain glacier systems that include valley glaciers tens of kilometers long (Fig. 10.6).

3. Along some high-latitude sea coasts, nearly every large valley glacier occupies a deep **fjord,** the seaward end of a glacier-carved bedrock trough, the floor of which lies far below sea level. Such glaciers are called *fjord glaciers.*
4. *Ice caps* mantle mountain highlands or low-lying land at high latitude and generally flow radially outward from their center (Fig. 10.5C).
5. Huge continent-sized *ice sheets* overwhelm nearly all the land surface within their margins (Fig.

Figure 10.6 A vertical satellite view of the valley-glacier complex that covers much of Denali National Park in south-central Alaska. Mount McKinley, the highest peak in North America, lies near the center of the glacier-covered region.

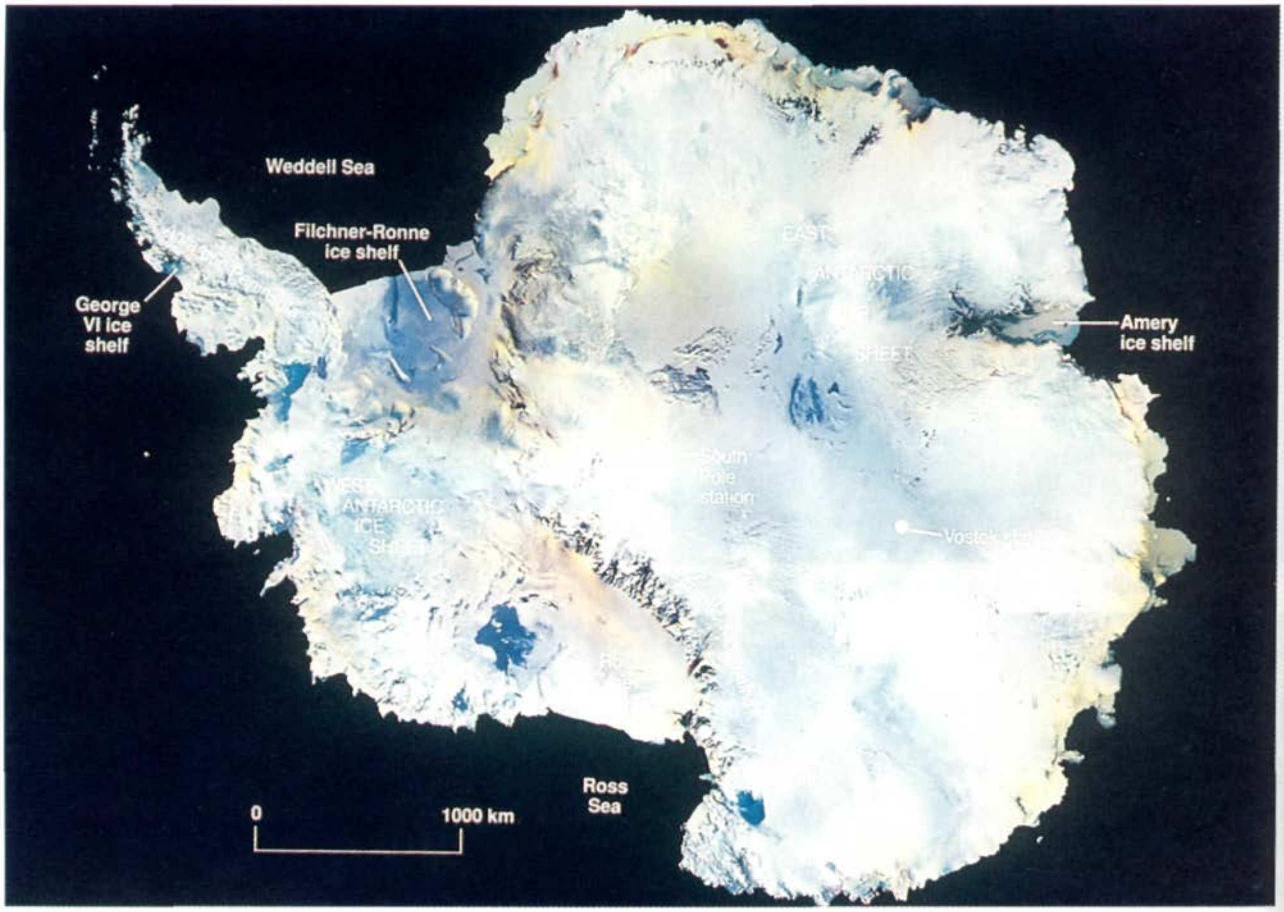


Figure 10.7 Satellite view of Antarctica. The East Antarctic Ice Sheet overlies the continent, while the much smaller West Antarctic Ice Sheet covers a volcanic island arc and surrounding seafloor. Major ice shelves occupy large coastal embayments. The ice-covered regions of Antarctica nearly equal the combined areas of Canada and the coterminous United States.

10.7). Modern ice sheets, which are confined to Greenland and Antarctica, include about 95 percent of the ice in existing glaciers and reach thicknesses of 3000 m (> 9800 ft) or more.

6. Floating *ice shelves* hundreds of meters thick occupy large embayments along the coasts of Antarctica (Fig. 10.7), and smaller ones are found among the Canadian Arctic islands.

How Glaciers Form

Newly fallen snow is porous and has a density less than a tenth that of water. Air easily penetrates the pore spaces, and the delicate points of each snowflake gradually evaporate. The resulting water vapor condenses, mainly in constricted places near a snowflake's center. In this way, the fragile ice crystals slowly become smaller, rounder, and denser, and the pore spaces between them disappear (Fig. 10.8).

Snow that survives a year or more gradually becomes denser and denser until it is no longer permeable to air, at which point it becomes **glacier ice.** Although now a rock, glacier ice has a far lower melting temperature than any other naturally occurring rock, and its density—about 0.9 g/cm^3—means it floats in water.

Further changes take place as the ice becomes buried deeper and deeper within a glacier. Figure 10.9 shows a core obtained by Russian glaciologists drilling deep in the East Antarctic Ice Sheet at Vostok Station (Fig. 10.7). As snowfall adds to the glacier's thickness, the increasing pressure causes initially small grains of glacier ice to grow until, near the base of the ice sheet, they reach a diameter of 1 cm (0.4 in) or more. This increase in grain size is similar to what

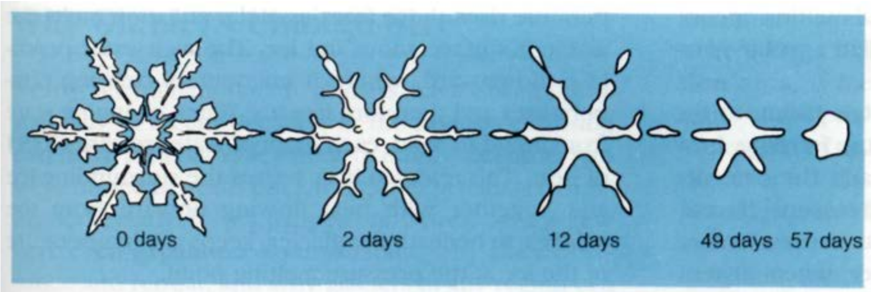


Figure 10.8 As a snowflake is slowly converted to a granule of old snow, melting and evaporation cause its delicate points to disappear. The resulting meltwater refreezes, and vapor condenses near the center of the crystal, making it denser.

happens in a fine-grained rock that is carried deep within the Earth's crust: as the rock is subjected to high pressure over a long time, large mineral grains slowly develop (Chapter 7).

Distribution of Glaciers

Potentially, a glacier can develop anywhere the snowline intersects the land surface and the topography permits snow and ice to accumulate (Fig. 10.3). This explains why glaciers are found not only at sea level in the polar regions, where the snowline is low, but also near the equator, where some lofty peaks in New Guinea, East Africa, and the Andes rise above the snowline. As we might expect, most glaciers are found in high latitudes, the coldest regions of our planet. However, because low temperatures also occur at high altitudes, many glaciers also can exist in lower latitudes on high mountains.

Low temperature is not the only limiting factor determining where glaciers can form, for a glacier must receive a continuing input of snow. Proximity to a moisture source is therefore another requirement. The abundance of glaciers in the coastal mountains of northwestern North America, for example, is related mainly to the abundant precipitation received from air masses moving landward from the Gulf of Alaska. Farther inland, the Rocky Mountains contain fewer and smaller glaciers because the climate there is much drier. Thus, in northwestern North America the existence of glaciers is linked to the interaction of several Earth systems: to tectonic forces that have produced high mountains; to the adjacent ocean, which provides an abundant source of moisture; and to the atmosphere, which delivers the moisture to the land as snow.

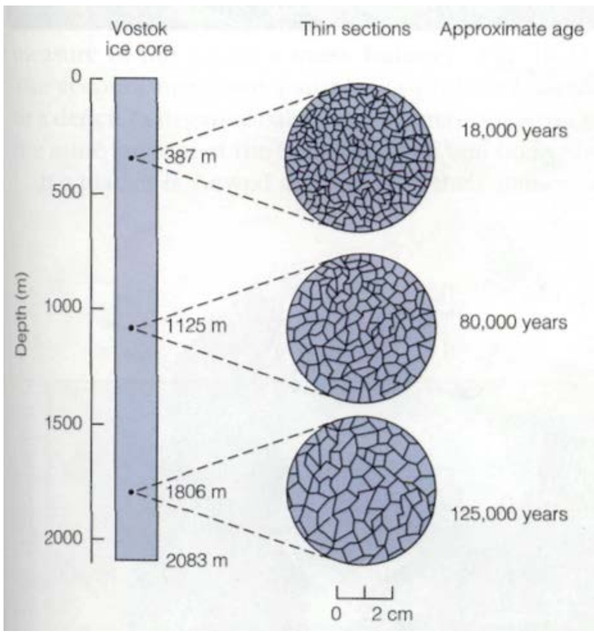


Figure 10.9 A deep ice core drilled at Russia's Vostok Station penetrates through the East Antarctic Ice Sheet to a depth of 2083 m. Thin-section samples taken from different depths in the core show a progressive increase in the size of ice crystals, the result of slow recryslallization as the thickness and weight of overlying ice slowly increase with time.

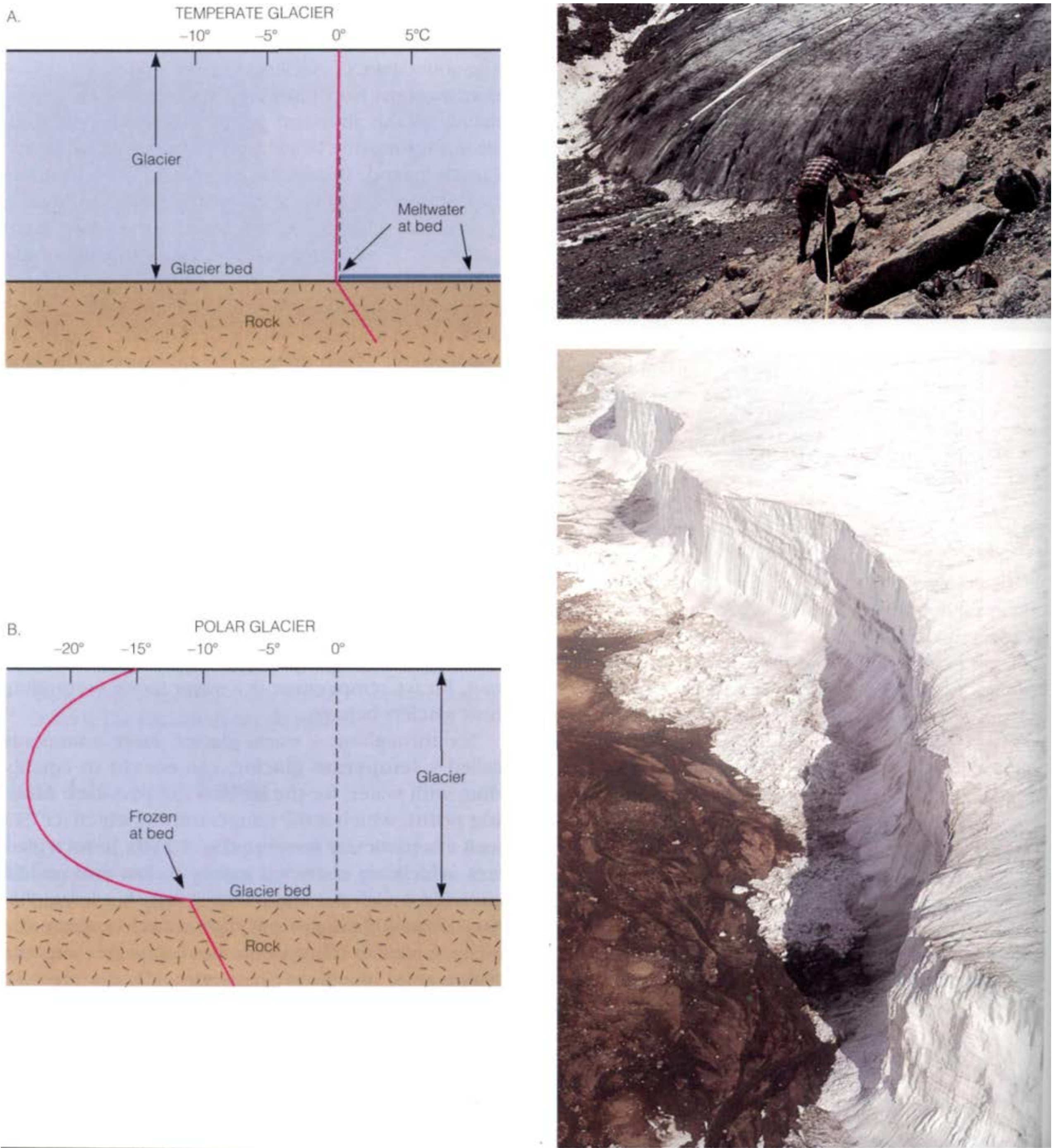
Warm and Cold Glaciers

Glaciers obviously are cold because they consist of ice and snow. However, when we drill holes through glaciers in a variety of geographic environments, we find a large range in ice temperatures. This temperature range allows us to divide glaciers into warm and cold types. The difference between them is important, for ice temperature is a major factor controlling how glaciers behave.

Ice throughout a warm glacier, more commonly called a **temperate glacier,** can coexist in equilibrium with water, for the ice is at the **pressure melting point,** which is the temperature at which ice can melt at a particular pressure (Fig. 10.10). In such glaciers, which are restricted mainly to low and middle latitudes, meltwater and ice exist together in equilibrium. At high latitudes and high altitudes, where the mean annual air temperature lies below freezing, the temperature in a glacier remains below the pressure melting point and little or no seasonal melting occurs. Such a cold glacier is commonly called a **polar glacier.**

If the temperature of snow crystals falling to the surface of a temperate glacier is below freezing, how does ice throughout the glacier reach the pressure melting point? The answer lies in the seasonal fluctuation of air temperature and in what happens when water freezes to form ice. In summer, when air temperature rises above freezing, solar radiation melts the glacier's surface snow and ice. The meltwater percolates downward, where it encounters freezing temperatures and therefore freezes. When changing state from liquid to solid, each gram of water releases 335 J of heat. This released heat warms the surrounding ice and, together with heat flowing upward from the solid earth beneath the glacier, keeps the temperature of the ice at the pressure melting point.

Figure 10.10 Temperate and polar glaciers. A. Ice in a temperate glacier is at the pressure melting point from surface to bed. The terminus is rounded, as illustrated by Pre de Bar Glacier in the Italian Alps, because melting occurs at the surface. B. Ice in a polar glacier remains below freezing, and the ice is frozen to its bed. Subfreezing temperatures inhibit melting at the terminus, which forms a steep cliff of ice, as illustrated by Commonwealth Glacier in Antarctica.



Why Glaciers Change Size

Nearly all high-mountain glaciers have shrunk substantially in recent decades, in the process exposing extensive areas of valley floor that only a century ago were buried beneath thick ice. Other glaciers have remained relatively unchanged, however, and a few have even expanded. To understand why glaciers advance and retreat, and why glaciers in any region can show dissimilar behavior, we need to examine how a glacier responds to a gain or loss of mass.

Annual Balance of a Glacier

The mass of a glacier is constantly changing as the weather varies from season to season and, on longer time scales, as local and global climates change. We can think of a glacier as being like a checking account. The balance in the account at the end of the year is the difference between the amount of money added during the year and the amount removed. The balance of a glacier's account is measured in terms of the amount of snow added, mainly in the winter, and the amount of snow (and ice) lost, mainly during the summer. The additions are collectively called **accumulation**, and the losses are **ablation.** The total in the account at the end of a year—in other words, the difference between accumulation and ablation—is a measure of the glacier's **mass balance** (Fig. 10.11). The account may have a surplus (a positive balance) or a deficit (a negative balance), or it may hold exactly the same amount at the beginning and end of a year.

If a glacier is viewed at the end of the summer ablation season, two zones are generally visible on its surface (Fig. 10.12). An upper zone, the **accumulation area,** is the part of the glacier covered by remnants of the previous winter's snowfall and is an area of net gain in mass. Below it lies the **ablation area,** a region of net loss where bare ice and old snow are exposed because the previous winter's snow cover has melted away.

When, over a period of years, a glacier gains more mass than it loses, its volume increases. The front, or **terminus,** of the glacier is then likely to advance as the glacier grows. Conversely, a succession of years in which negative mass balance predominates will lead to retreat of the terminus. If no net change in mass occurs, the terminus is likely to remain relatively stationary.

The Equilibrium Line

The **equilibrium line** marks the boundary between the accumulation area and the ablation area (Fig. 10.12) It lies at the level on the glacier where net mass loss equals net mass gain. The equilibrium line on temperate glaciers coincides with the local snowline. Being very sensitive to climate, the equilibrium line fluctuates in altitude from year to year and is higher in warm, dry years than in cold, wet years (Fig. 10.13). Because of this sensitivity, we can use the altitude of the equilibrium line to estimate a glacier's mass balance without having to make detailed field measurements of accumulation and ablation. Thus, fluctuations in the altitude of the equilibrium line over time can provide us with a measure of changing climate.

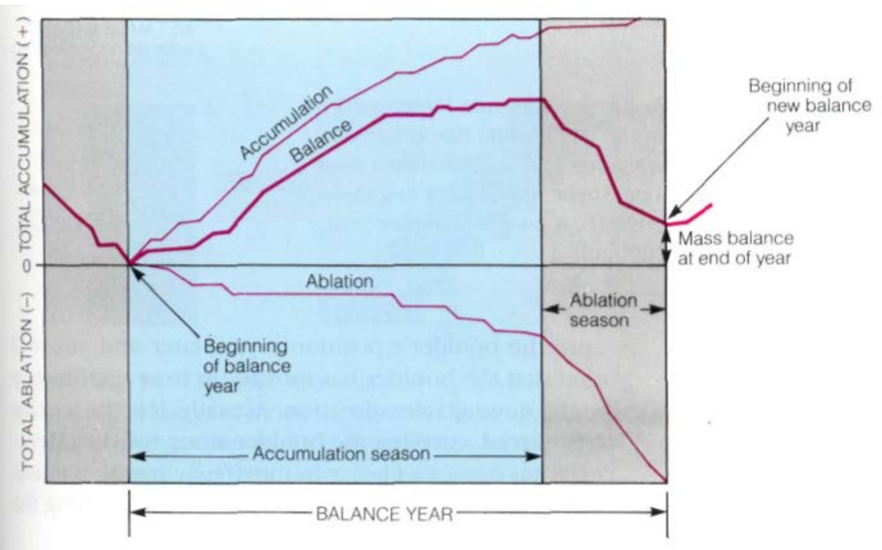


Figure 10.11 Accumulation and ablation determine glacier mass balance (heavy line) over the course of a balance year. The balance curve, obtained by summing values of accumulation (positive values) and ablation (negative values), rises during the accumulation season as mass is added to the glacier and then falls during the ablation season as mass is lost. The mass balance value at the end of the balance year reflects the difference between annual mass gain and mass loss.

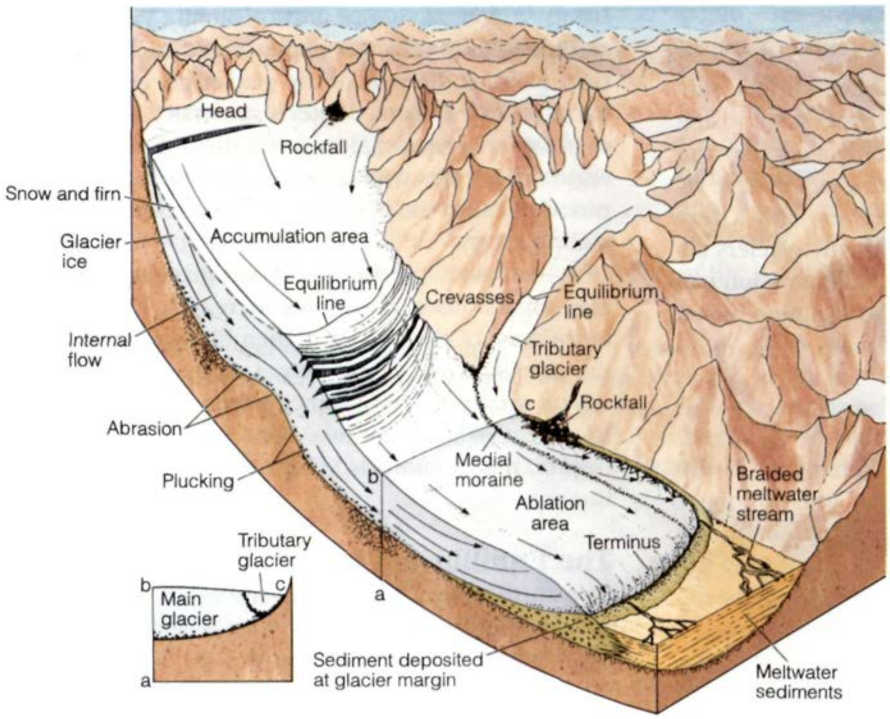


Figure 10.12 Main features of a valley glacier. The glacier has been cut away along its center line so that only half is shown. Crevasses form where the glacier flows over an abruptly steepened slope. Arrows show the local directions of ice flow. A band of rock debris forms a medial moraine that marks the boundary between the main glacier and a tributary glacier joining it from a lateral valley.

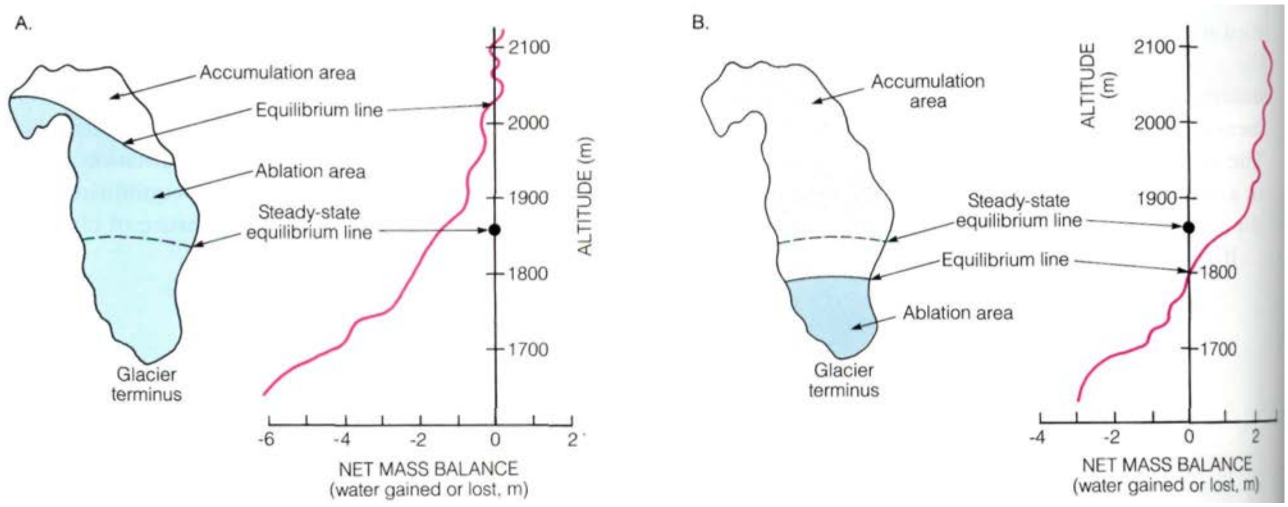


Figure 10.13 Maps of South Cascade Glacier in the Washington Cascade Range at the end of two successive balance years showing the position of the equilibrium line relative to the position it would have under a balanced condition. The curves plot mass balance as a function of altitude. During the first year, A, a negative balance year, the glacier lost mass and the equilibrium line was high (2025 m). The following year, B, a positive balance year, the glacier gained mass and the equilibrium line was low (1800 m).

How Glaciers Move

One way to prove that glaciers move is to walk onto a glacier near the end of the summer and carefully measure the position of a surface boulder with respect to some fixed point beyond the glacier margin. Remeasure the boulder's position a year later and you will find that the boulder has moved up to several meters in the downglacier direction. Actually, it is the ice that has moved, carrying the boulder along for the ride.

What causes a glacier to move may not be immediately obvious, but we can find clues by examining the ice and the terrain on which it lies. These clues tell us that ice moves in two primary ways: by internal flow and by sliding of the basal ice across rock or sediment.

Internal Flow

When an accumulating mass of snow and ice on a mountainside reaches a critical thickness, the mass begins to deform and flow downslope under the pull of gravity. The flow takes place mainly through movement within individual ice crystals, which are subjected to higher and higher stress as the weight of the overlying snow and ice increases. Under this stress, ice crystals are deformed by slow displacement (termed *creep)* along internal crystal planes in much the same way that cards in a deck of playing cards slide past one another if the deck is pushed from one end (Fig. 10.14). As the compacted, frozen mass begins to move, stresses between adjacent ice crystals cause some to grow at the expense of others, and the resulting larger crystals end up with their internal planes oriented in the same direction. This alignment of crystals leads to increased efficiency of flow, for the internal creep planes of all crystals now are parallel.

In contrast to deeper parts, where the ice flows as a result of internal creep, the surface portion of a glacier has relatively little weight on it and is brittle. Where a glacier passes over an abrupt change in slope, such as a bedrock cliff, the surface ice cracks as tension pulls it apart. When a crack opens up, it forms a **crevasse**, a deep, gaping fissure in the upper surface of a glacier, generally less than 50 m (165 ft) deep (Fig. 10.12). At depths greater than about 50 m, continuous flow of ice prevents crevasses from forming. Because it cracks at the surface yet flows at depth, a glacier is analogous to the upper layers of the Earth, which include a surface zone that cracks and fractures (the lithosphere) and a deeper zone (the asthenosphere) that can flow slowly.

Basal Sliding

Ice temperature is very important in controlling the way a glacier moves and its rate of movement. Meltwater at the base of a temperate glacier acts as a lubricant and permits the ice to slide across its *bed* (the rocks or sediments on which the glacier rests) (Fig. 10.15). In some temperate glaciers, such sliding accounts for up to 90 percent of the total observed movement. By contrast, polar glaciers are so cold they are frozen to their bed. Their motion largely involves internal deformation rather than basal sliding, and so their rate of movement is greatly reduced.

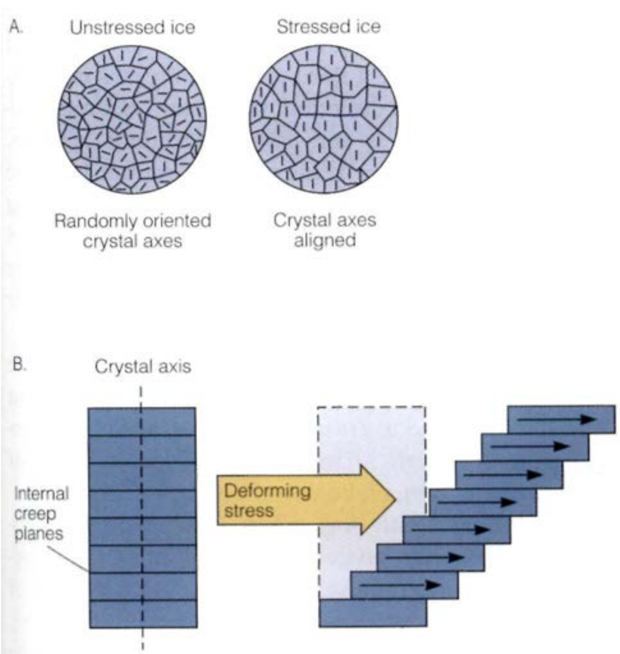


Figure 10.14 Internal creep in the ice crystals of a glacier. A. Randomly oriented ice crystals in the upper layers of a glacier are reorganized under stress so that their axes are aligned. B. When stress is applied to an ice crystal, creep along internal planes causes slow deformation.

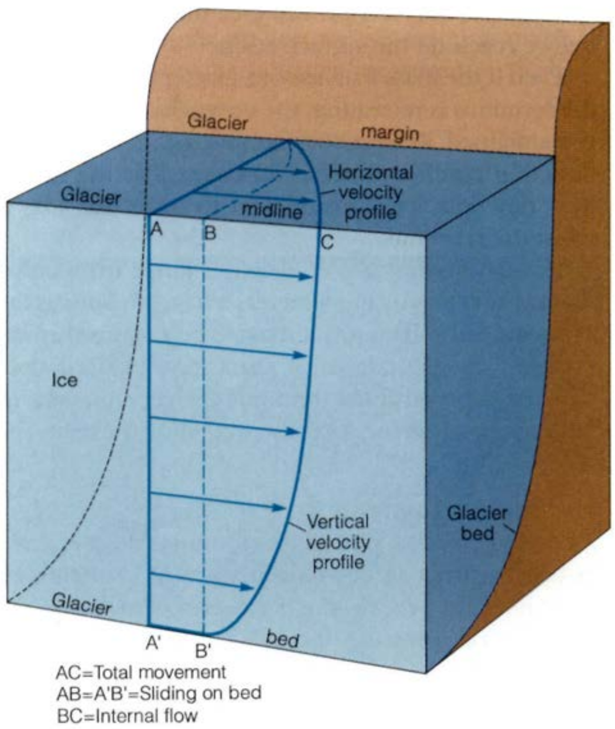


Figure 10.15 Three-dimensional view through half of a glacier showing horizontal and vertical velocity profiles. Glacier movement is due partly to internal flow and partly to sliding of the glacier across its bed.

Ice Velocity

Measurements of the surface velocity across a valley glacier show that the uppermost ice in the central part of the glacier moves faster than ice at the sides, similar to the velocity distribution in a river (Figs. 10.15 and 9.5). The reduced rates of flow toward the margins are due to frictional drag of the ice against the valley walls. A similar reduction in flow rate toward the bed is observed in a vertical profile of velocity (Fig. 10.15).

Although snow piles up in the accumulation area each year, and melting removes snow and ice from the ablation area, a glacier's surface profile does not change much because ice is transferred from the accumulation area to the ablation area. In the accumulation area, the mass of accumulating snow and ice is pulled downward by gravity, and so the dominant flow direction is toward the glacier bed. However, the ice does not build up to ever greater thickness because a downglacier component of flow is also present. Ice flowing downglacier replaces ice being lost from the glacier's surface in the ablation area, and so in this area the flow is upward toward the surface (Fig. 10.12). Ice crystals falling as snowflakes on the glacier near its head therefore have a long path to follow before they emerge near the terminus. Those falling close to the equilibrium line, on the other hand, travel only a short distance through the glacier before reaching the surface again.

Even if the mass balance of a glacier is negative and the terminus is retreating, the downglacier flow of ice is maintained. Retreat does not mean that the ice-flow direction reverses; instead, it means that the rate of flow downglacier is insufficient to offset the loss of ice at the terminus.

In most glaciers, flow velocities range from only a few centimeters to a few meters a day, or *i* bout as fast as the rate at which groundwater percolates through crustal rocks. Hundreds of years have elapsed since ice now exposed at the terminus of a very long glacier fell as snow near the top of its accumulation area.

Response Lags

Advance or retreat of a glacier terminus does not necessarily give us an accurate picture of changing climate because a lag occurs between a climatic change and the response of the glacier terminus to that change. The lag reflects the time it takes for the effects of an increase or a decrease in accumulation above the equilibrium line to be transferred through the slowly moving ice to the glacier terminus. The length of the lag depends both on the size of a glacier and on the way the ice moves; the lag will be longer for large glaciers than for small ones and longer for polar glaciers than for temperate ones. Temperate glaciers of modest size (like those in the European Alps) have response lags that probably range from several years to a decade or more. This lag time can explain why, in any area that has glaciers of different sizes, some glaciers may be advancing while others are either stationary or retreating. Another explanation for a few seemingly anomalous advances is discussed in "A Closer Look: Earthquakes, Rockfalls, and Glacier Mass Balance."

Rapid Changes in Glacier Size

Calving

During the last century and a half, many coastal Alaskan glaciers have receded at rates far in excess of typical glacier retreat rates on land. Their dramatic recession is due to frontal **calving,** defined as the progressive breaking off of icebergs from the front of a glacier that terminates in deep water. Although the base of a fjord glacier may lie far below sea level along much of its length, its terminus can remain stable as long as it is resting (or "grounded") against a shoal (a shallow submarine ridge) (Fig. 10.16). However, if the terminus retreats off the shoal, water will replace the space that had been occupied by ice. With the glacier now terminating in water, conditions are right for calving. Because a fjord glacier increases in thickness in the upfjord direction, the water becomes progressively deeper as the calving terminus retreats. The deepening water leads to faster retreat because the greater the water depth, the faster the calving rate. Once started, calving will continue rapidly and irreversibly until the glacier front recedes into water too shallow for much calving to occur, generally near the head of the fjord.

Icebergs produced by calving glaciers constitute an ever-present hazard to ships in subpolar seas. In 1912, when the S.S. *Titanic* sank after striking an iceberg in the North Atlantic ocean, the detection of approaching bergs relied on the sharpness of sailors' vision. Today, with sophisticated electronic equipment, large bergs can generally be identified well before an encounter. Nevertheless, ice has a density of 0.9, so that 90 percent of an iceberg lies under water, making it difficult to detect. In coastal Alaska, where calving glaciers are commonplace, icebergs pose a potential threat to huge oil tankers. For this reason, Columbia Glacier, which lies adjacent to the main shipping lanes from Valdez at the southern end of the

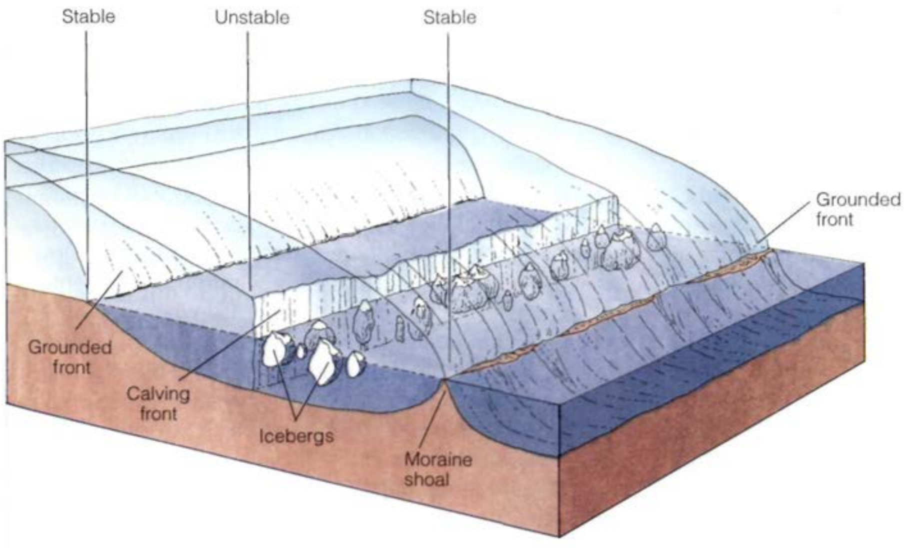


Figure 10.16 The terminus of a fjord glacier remains stable if it is grounded against a shoal (submarine ridge), but if the glacier retreats into deeper water, calving will begin. The unstable terminus then retreats at a rate that depends mainly on water depth. Once it becomes grounded farther up the fjord, the terminus is stable again.

Alaska Pipeline, is being closely monitored as its terminus pulls steadily back and multitudes of bergs are released.

Glacier Surges

Although most glaciers slowly grow or shrink as the climate fluctuates, some experience episodes of very unusual behavior marked by rapid movement and dramatic changes in size and form. Such an event, called a **surge,** is unrelated, or only secondarily related, to a change in climate. When a surge occurs, a glacier seems to go berserk. Ice in the accumulation area begins to move rapidly downglacier, producing a chaos of crevasses and broken pinnacles of ice in the ablation area. Medial **moraines**, which are bands of rocky debris marking the boundaries between adjacent tributary glaciers (Fig. 10.5B), are deformed into intricate patterns (Fig. 10.17). The termini of some glaciers have advanced up to several kilometers during surges. Rates of movement as great as 100 times those of nonsurging glaciers and averaging as much as 6 km (3.7 mi) a year have been measured.

Figure 10.17 Contorted medial moraines of Susitna Glacier in the Alaskan Range provide striking evidence of periodic surges during which tributary ice streams advance at rates far greater than those of adjacent nonsurging glaciers.

A Closer Look

Earthquakes, Rockfalls, and Glacier Mass Balance

The great Prince William Sound earthquake of March 27, 1964, which resulted in widespread death and destruction throughout southern Alaska, triggered the collapse of a massive mountain buttress above Sherman Glacier in the Chugach Mountains. The resulting landslide spread debris across 8.5 km^2 (3.3 mi^2) of the glacier surface, covering about a third of the ablation area (Fig. C10.1). Before the earthquake, the mass balance of the glacier was slightly negative, the annual loss of ice in the ablation area was about 4 m (4.4 yd), and the terminus was retreating about 25 m/year (27 yd/yr). Within a few years following the earthquake, the debris cover (which averaged 1.3 m, or 1.4 yd thick) reduced the annual melting to only a few cm. The result was a shift to a positive mass balance, which caused the glacier terminus to advance.

Many mountain glaciers, like the Sherman, flow through valleys bordered by steep cliffs that give rise to periodic rockfalls. If a rockfall spreads across the accumulation zone of a glacier, its initial effect will be minimal, for the debris is quickly buried by successive winter snowfalls. However, as the glacier continues to flow downvalley, the debris traveling within the ice eventually emerges at the surface (Fig. 10.12) where it will reduce ablation in a manner similar to that of rock debris falling directly onto the ablation zone. The principal difference is that the response of the glacier's terminus will lag the rockfall event by however long it takes the debris to emerge at the surface.

The abrupt addition of a debris cover in the ablation area will cause a glacier to respond differently than other nearby glaciers that respond only to climate. Thus, a sudden rockfall event could explain why a glacier would behave nonsynchronously compared to other glaciers in the same climatic environment.

Figure C10.1 A vast sheet of rocky debris covers the lower ablation zone of Sherman Glacier following the collapse of a large mountain buttress during the 1964 Alaska earthquake. The debris cover impeded melting, leading to a negative mass balance and a subsequent advance of the glacier terminus.

The cause of surges is still imperfectly understood, but available evidence supports a reasonable hypothesis. We know that the weight of the ice can produce high pressure in water at the base of a glacier. Over a period of years, this steadily increasing pressure may cause the glacier to separate from its bed. The resulting effect is similar to the way a car hydroplanes on a wet road surface. According to this hypothesis, as the ice is floated off its bed, its forward mobility is greatly increased and it moves rapidly forward before the water escapes and the surge stops.

Glaciers As Environmental Archives

Trapped in the snow that piles up each year in the accumulation area of a glacier is evidence of both local and global environmental conditions. The evidence includes physical, chemical, and biological components that can be extracted in a laboratory and studied as a record of the changing natural environment. While the oldest ice in most cirque and valley glaciers is several hundred to several thousand years old, large ice caps and ice sheets contain ice that dates far back into the ice ages. The record they contain is often unique, and it is of critical importance for understanding how the atmosphere, oceans, and biosphere have changed over hundreds of thousands of years.

If we dig a pit several meters deep in the accumulation area of a glacier and look closely at the snow, a cyclic layering can be seen. In each layer, the relatively clean snow that records a succession of winter snowfalls passes upward to a darker layer that contains dust and refrozen meltwater, a record of relatively dry summer weather. Depending on the local accumulation rate, from one to many such annual layers may be exposed in our pit. Because digging a pit into a glacier is an inefficient way to examine the stratigraphy, glaciologists drill cores of ice that can be extracted and returned to a laboratory for analysis. Some drilling operations have penetrated to the base of the thick Greenland and Antarctic ice sheets, while others have focused on high-latitude and high-altitude ice caps (see "Guest Essay" at the end of this chapter).

Ice cores have proved a boon for atmospheric scientists who would like to know whether the concentrations of important atmospheric trace gases like carbon dioxide and methane, which are trapped in air bubbles in the ice, have fluctuated as the climate changes (Chapter 14). Measurements of oxygen isotopes in the ice can tell us the air temperature when the snow accumulated on the glacier surface. Ice cores also provide a record of major volcanic eruptions that generate large volumes of sulfur dioxide gas which, combined with water, accumulates as a layer of acid snowfall on glaciers. High concentrations of fine dust in ice layers dating back to the last ice age show that the windy climate of glacial times was also extremely dusty. Tiny fragments of organic matter and fossil pollen grains trapped in the ice layers can tell us about local vegetation composition near a glacier and can be radiocarbon-dated to provide ages for the enclosing ice. Because these natural historical archives are trapped in annual ice layers that can be read like the pages of a book, they offer an unparalleled, detailed look at past surface conditions on our planet.

SEA ICE

Approximately two-thirds of the Earth's permanent ice cover floats as a thin veneer of **sea ice** at the ocean surface in polar latitudes (Fig. 10.18). Despite its vast extent, sea ice comprises only about 1/1000 of the Earth's total volume of ice.

How Sea Ice Forms

Once the ocean surface cools to the freezing point of seawater, slight additional cooling leads to ice formation. The first ice to form consists of small crystalline platelets and needles up to 3 or 4 mm (0.1 or 0.2 in) in diameter that collectively are termed *frazil ice*. As more ice crystals form, they produce a soupy mixture at the ocean surface. In the absence of waves or turbulence, the crystals freeze together to form a continuous cover of ice 1 to 10 cm (0.4 to 4 in) thick. If waves are present, the crystals form rounded, pancake-like masses up to 3 m (10 ft) in diameter that eventually weld together into a continuous sheet of sea ice. Once a continuous ice cover forms, the cold atmosphere is no longer in contact with the seawater, and sea-ice growth then proceeds by the addition of ice to the sea-ice base. In the Arctic, over the course of a yearly cycle about 45 cm (18 in) of ice is lost from the ice surface, but an equal amount is added to the base. As a result, an ice crystal added to the sea ice at

Figure 10.18 A glaciologists camp on a slab of sea ice in the southern Beaufort Sea (ca. 76°N Lat., 150°W Long.) As ice broke up at the end of summer, a fissure (left foreground) split the camp in two, causing one segment of camp to drift 400 m away from the one shown here.

its base will move upward through the ice column with an average velocity of about 45 cm/yr (18 in/yr) until it reaches the surface and melts away.

Sea-Ice Distribution and Zonation

The contrasting geography of the Earth's two polar regions leads to important differences in the distribution of sea ice in the two hemispheres. The South Pole lies near the middle of the continent of Antarctica, which is mantled by a vast, thick ice sheet, whereas the North Pole falls in the middle of the deep Arctic Ocean basin, which is underlain by oceanic crust. The open Southern Ocean adjacent to Antarctica contrasts with the largely land-locked Arctic Ocean, which is connected to the world ocean only by relatively narrow straits. In the Antarctic, sea ice forms a broad ring around the continent and adjacent ice-covered archipelago, a ring that varies in width with the seasons. At its greatest northward extent in winter it covers 20 million km^2 (7.7 million mi^2), but it shrinks to only 4 million km^2 (1.5 million mi^2)n summer (Figs. 10.19A and 10.20). By contrast, the Arctic Ocean is ice-covered most of the year, and several marginal seas (i.e., the Sea of Japan, the Sea of Ohkotsk, the Bering Sea, Davis Strait, Hudson Bay) are partially or wholly ice-covered during the winter. At its minimum extent in August, Arctic sea ice covers about 7 million km^2, (2.7 million mi^2) while during its winter maximum it expands to 14 million km^2 (5.4 million mi^2) (Fig. 10.19B).

Scientists commonly categorize sea-ice zones as being either perennial or seasonal. *The perennial ice zone* contains sea ice that persists for at least several years (multiyear ice). In the Arctic, this zone lies north of 75° latitude and contains about two-thirds of all perennial sea ice. Near the center of the basin, the ice has an average thickness of 3 to 4 m (10 to 13 ft) and an age of up to at least several decades. In the Antarctic, multiyear ice is restricted to semi-enclosed seas (the Ross, Weddell, and Bellinghausen), where it reaches a thickness of up to 5 m (16.5 ft) but an age of less than five years. In the *seasonal ice zone,* the ice cover varies annually. In the Arctic, ice of this zone is less than 2 m (6.5 ft) thick where undeformed, but deformation within the pack ice often increases thickness substantially. In the Southern Ocean, the limit of seasonal ice on average shifts through 10° of latitude, Here, the ice front retreats poleward in summer largely in response to heat derived from the ocean water, whereas in the Arctic, surface melting in summer is a major factor in the retreat of the ice margin.

Sea-Ice Motion

Sea ice is in constant motion, driven by winds and ocean currents. Average drift rates in the Arctic Ocean are about 7 km/day (4.3 mi/day), whereas in the Greenland and Bering seas velocities reach 15 km/day (93 mi/day). Each year, about 10 percent of the Arctic Sea ice moves south into the Greenland Sea, where it eventually breaks up and melts away. Sea ice generally moves clockwise around Antarctica, but a large gyre in the Weddell Sea, east of the Antarctic Penin-

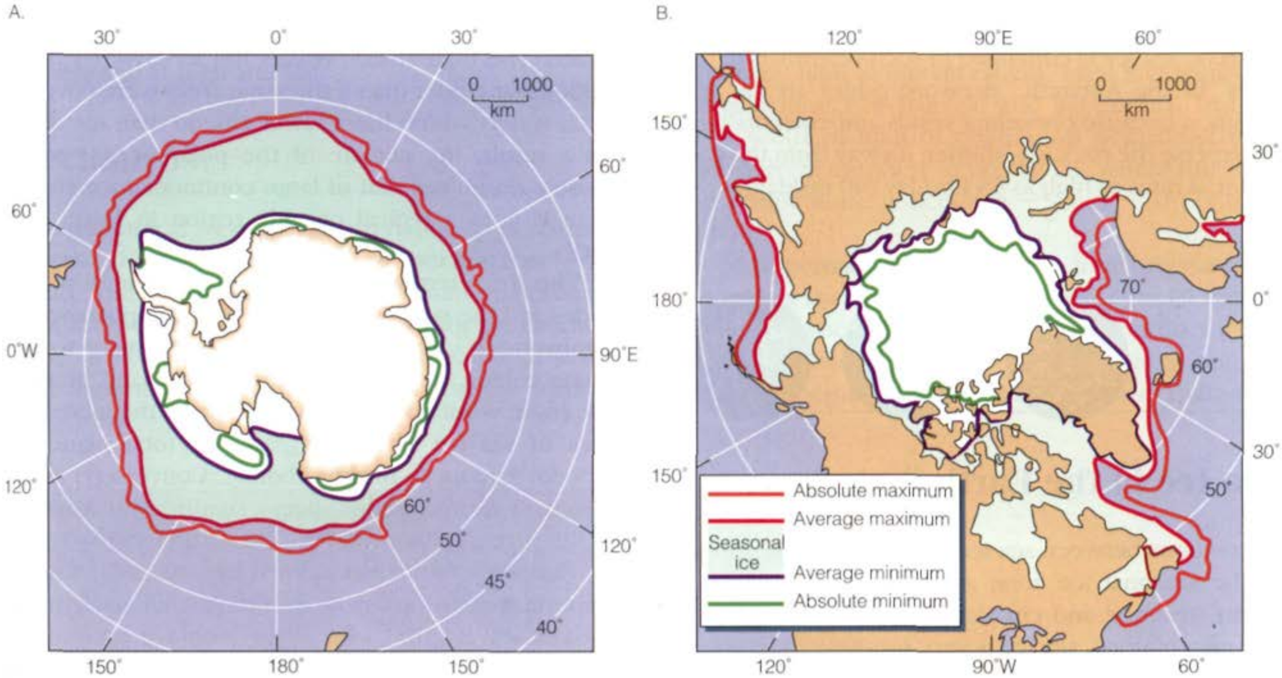


Figure 10.19 Seasonal extent of sea ice in A. southern hemisphere and B. northern hemisphere.

sula, causes the drifting ice to pile up to form a large region of multiyear ice.

Stresses resulting from diverging movement of the thin ice cover cause it to break, exposing the underlying water. Such a linear opening, called a **lead**, tends to be long and narrow and may extend for many kilometers. An exceptionally large lead may grow to become a huge area of open water called a *polynya* (Fig. 10.20). Because of the large temperature gradient between the air and sea water in a lead, the water loses heat rapidly, causing a new, thin cover of ice to form quickly. As a result, the fractured ice pack becomes a changing complex mosaic of new ice and older ice. Although the exposure of surface water to the atmosphere permits substantial amounts of solar energy to reach the upper ocean, such open water commonly comprises less than 1 percent of the winter sea-ice cover.

Early explorers who tried to reach the North Pole by crossing the Arctic ice pack quickly found it rough going. The ice is not a vast smooth surface; rather, it is broken by numerous *pressure ridges*, formed when the shifting, fractured ice converges, shears, and piles up, in much the same way that converging lithospheric plates produce mountain chains on the continents. Beneath each pressure ridge is a submerged *keel* of deformed ice, much like the keel of a sailboat, up to five times as thick as the overlying ridge. Esti-

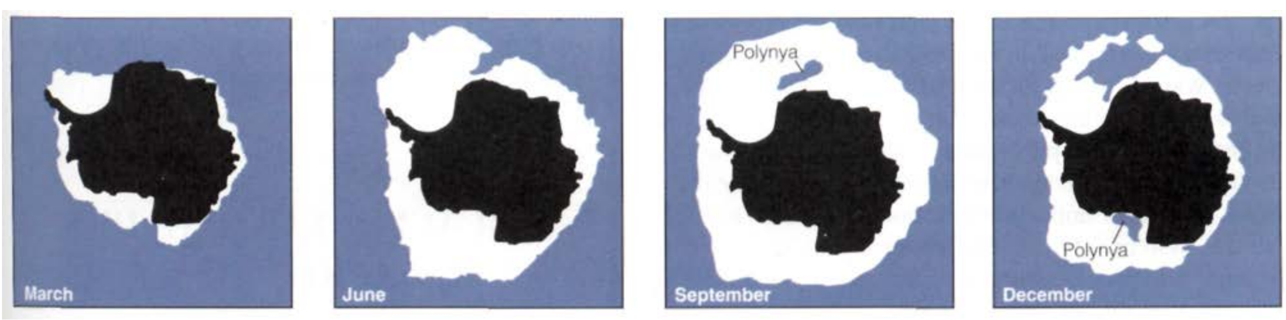


Figure 10.20 Seasonal variations in the sea-ice cover around Antarctica in a typical year. The ice is least extensive during the summer months (January-March) but steadily increases, reaching a maximum in winter (July-September). At the time of maximum sea ice (September), a large polynya has developed northeast of the Weddell Sea.

mates suggest that as much as 40 percent of the mass of Arctic sea ice is contained in such deformation features. In the Antarctic, pressure ridges are far less common because prevailing winds and currents tend to disperse the pack ice, shifting it away from the continent at rates as high as 65 km/day (40 mi/day).

Sea Ice in the Earth System

Interactions between sea ice, ocean, and atmosphere in the seasonal ice zone are believed to influence ocean structure and circulation. Because it is only a few meters thick, sea ice is very sensitive to temperature changes in the overlying atmosphere and in the ocean water below. In turn, the ice cover affects both the atmosphere and ocean in important ways. The growth of sea ice increases salinity at the top of the mixed layer (Chapter 8) when salt is excluded as seawater freezes. Conversely, when the ice melts, the salinity is decreased as fresh water is added to the ocean surface. Thus, continual variations in the extent of sea ice cause corresponding variations in the salinity near the top of the water column.

Exclusion of salt as seawater freezes leads to the production of cold, saline (and therefore dense) water on the continental shelves. This dense water spills downward off the shelves into the ocean basins to produce deep water and bottom water. The process is enhanced in the marginal Antarctic seas, where offshore winds generate polynyas: Here, rapid ice growth under extremely cold conditions produces large quantities of dense water that likely is the source of much of the Antarctic Bottom Water. Similar processes operating in the Greenland and Norwegian seas are responsible for producing North Atlantic Deep Water, which is crucial to maintaining the global thermohaline circulation system (Chapter 8).

Both its rapid response to changing conditions and its direct influence on the atmosphere and oceans make sea ice an important component of the Earth's climate system. The floating cover of ice effectively isolates the ocean surface from the atmosphere, thereby cutting off the exchange of heat between these two reservoirs; the more extensive the ice cover, the stronger the effect. At the same time, the ice surface is highly reflective (i.e., it has a high **albedo,** which is a measure of surface reflectivity) and bounces incoming solar radiation back into space. The high albedo makes the ice-covered polar regions far colder than if the same areas were covered with water, which has a lower albedo than ice does. As a result, the climate of the polar oceans more closely resembles that of large continental ice sheets than it does a typical oceanic region in lower latitudes.

The steep temperature gradient between low latitudes and the polar regions is of major importance to atmospheric circulation. If the climate were to become colder, causing the sea ice to expand in area, the result would be a positive feedback: the increased area of sea ice would increase the total planetary albedo, leading to further cooling. Conversely, if the ice cover shrinks or disappears, significant disruption in the pattern of atmospheric circulation might occur.

That sea ice influences global climate leads to some important questions: How stable is the ice pack in the land-locked Arctic basin? What would it take to remove the thin ice cover completely, and how long would it take? In a time of climatic warming (Chapter 14), might we expect the ice cover to disappear suddenly, thereby causing abrupt changes in climate in some of the Earth's most densely populated regions? Climate models suggest that in a warming climate, the warming at high latitudes will be several times that at low and middle latitudes. A possible scenario indicates that, as the Arctic warms up, we can expect a gradual shrinking of the ice pack followed by a discontinuous transition to ice-free conditions. We can easily calculate how much of a change would lead to disappearance of perennial Arctic ice: it could occur, for example, with a 3° to 5°C increase in annual temperature, a 25 to 30 percent increase in solar radiation reaching the ice surface, a 15 to 20 percent decrease in summer albedo (brought about by increased surface melting), or a significant change in cloudiness. While exactly how a change in albedo would take place is not known, modeling suggests that once the albedo reaches a sufficiently low value, the shift to an ice-free Arctic Ocean would occur rapidly, measurable in years rather than in decades.

PERIGLACIAL LANDSCAPES AND PERMAFROST

Land areas beyond the limit of glaciers where low temperature and frost action are important factors in determining landscape characteristics are called **periglacial** zones. Periglacial conditions are found over more than 25 percent of the Earth's land areas, primarily in the circumpolar zones of each hemisphere and at high altitudes.

Permafrost

A common feature of periglacial regions is perennially frozen ground, also known as **permafrost**—sediment, soil, or even bedrock that remains continuously at a temperature below 0°C (32°F) for an extended time (from two years to tens of thousands of years). The largest areas of permafrost occur in northern North America, northern Asia, and the high, cold Tibetan Plateau (Fig. 10.21). It has also been found on many high mountain ranges, even including some lofty summits in tropical and subtropical latitudes. The southern limit of continuous permafrost in the northern hemisphere generally lies where the annual air temperature is between —5 and —10°C (23 and 14°F).

Most permafrost is believed to have originated during either the last glacial age or earlier glacial ages. Remains of woolly mammoth and other extinct ice age animals found well preserved in frozen ground indicate that permafrost existed at the time of their death.

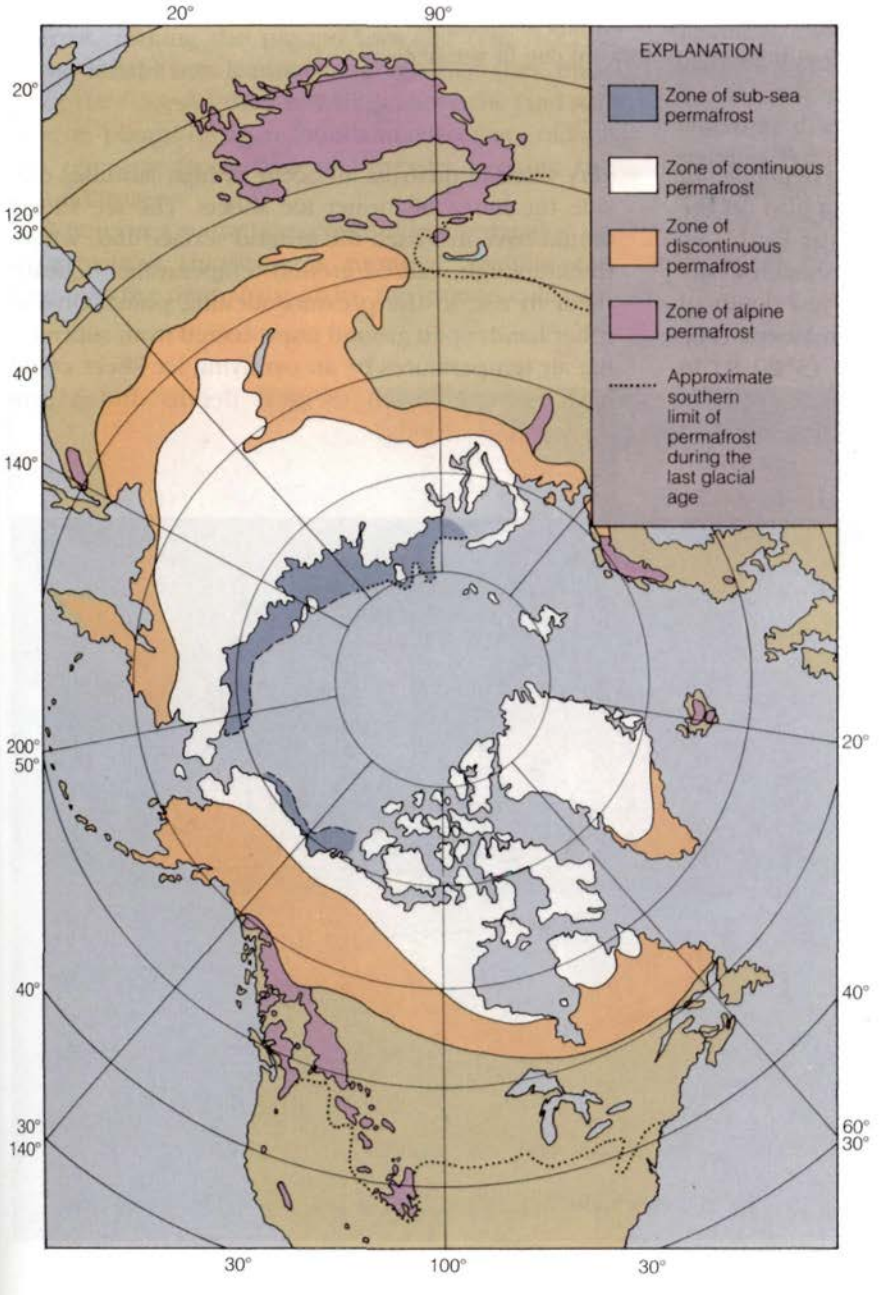


Figure 10.21 Distribution of permafrost in the northern hemisphere. Continuous permafrost lies mainly north of the 60th parallel and is most widespread in Siberia and Arctic Canada. Extensive alpine permafrost underlies the high, cold plateau region of central Asia. Smaller isolated bodies occur in the high mountains of the western United States and Canada.

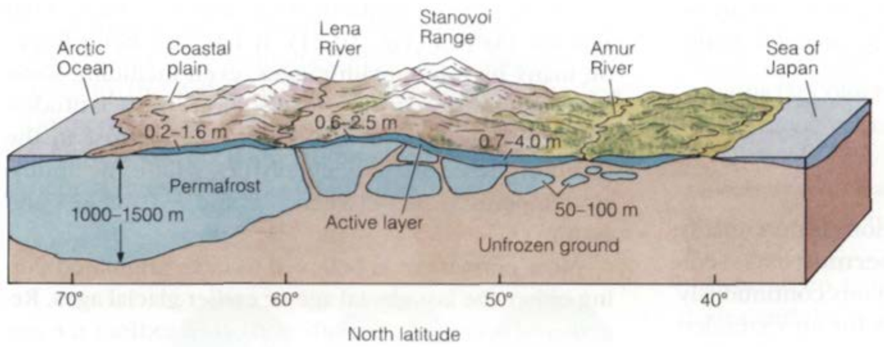


Figure 10.22 Diagrammatic transect across northeastern Siberia (vertical scale is greatly exaggerated) showing distribution and thickness of permafrost and thickness of the active layer. Thick, continuous permafrost under the Arctic coastal plain thins southward where it becomes discontinuous in response to warmer mean annual air temperature. The active layer increases in thickness southward due to warmer summer temperatures.

The depth to which permafrost reaches depends not only on the average air temperature but also on the rate at which heat flows upward from the Earth's interior and on how long the ground has remained continuously frozen. The maximum reported depth of permafrost is about 1500 m (4900 ft) in Siberia (Fig. 10.22). Thicknesses of about 1000 m (3300 ft) in the Canadian Arctic and at least 600 m (2000 ft) in northern Alaska have been reported. These areas of very thick permafrost all occur in high latitudes outside the limits of former ice sheets. The ice sheets would have insulated the ground surface and, where thick enough, caused ground temperatures beneath them to rise to the pressure melting point. On the other hand, open ground unprotected from subfreezing air temperatures by an overlying ice sheet could have become frozen to great depths during prolonged cold periods.

Figure 10.23 This cabin in central Alaska settled more than a meter in 8 years as permafrost beneath its foundation thawed.

Living with Permafrost

In permafrost terrain, a thin surface layer of ground that thaws in summer and refreezes in winter is known as the *active layer.* In summer the thawed layer tends to become very unstable. The permafrost beneath, however, is capable of supporting large loads without deforming. Many of the landscape features we associate with periglacial regions reflect movement of regolith within the active layer during annual freeze/thaw cycles.

Permafrost presents unique problems for people living on it. If a building is constructed directly on the surface, the warm temperature developed when the building is heated is likely to thaw the underlying permafrost, making the ground unstable (Fig. 10.23). Arctic inhabitants learned long ago that they must place the floors of their buildings above the land surface on pilings or open foundations so that cold air can circulate freely beneath, thereby keeping the ground frozen.

Wherever a continuous cover of low vegetation on a permafrost landscape is ruptured, melting can begin. As the permafrost melts, the ground collapses to form impermeable basins containing ponds and lakes. Thawing can also be caused by human activity, and the results can be environmentally disastrous. Large wheeled or tracked vehicles crossing Arctic tundra can quickly rupture it. The water-filled linear depressions that result from thawing can remain as features of the landscape for many decades.

The discovery of a commercial oil field on the North Slope of Alaska in the 1960s generated the need to transport the oil southward by pipeline to an ice-free port. The company formed to construct the pipeline was faced with some unique problems. In order for the sticky oil to flow through a pipeline in the frigid Arctic environment, the oil had to be heated. However, an uninsulated, heated pipe in the frozen ground could melt the surrounding permafrost. Even if the pipe were insulated before placing it underground, the surface vegetation cover would be disrupted, likely leading to melting and instability. For these reasons, along much of its course the pipeline was constructed on piers above ground, thereby greatly reducing the possibility of ground collapse (Fig. 10.24).

Figure 10.24 The Alaska Pipeline carries petroleum from the North Slope oil fields near Prudoe Bay southward across two mountain ranges enroute to the port of Valdez. To increase the ease of flow through the pipe, the oil is heated. Because much of the pipeline route lies across permafrost terrain, in many sectors the huge pipe is suspended above ground on large piers to keep it from melting the frozen ground beneath.

Guest Essay

Ice Core Archives: The Keys to Our Future Are Frozen in Our Past

Dr. Lonnie G. Thompson received his Ph.D. degree from the Ohio State University in 1976. He holds a research scientist position in the Byrd Polar Research Center and is a professor in the Department of Geological Sciences. He has led a series of major expeditions to recover ice cores for paleoenvironmental study and to investigate modern environments. These expeditions include the Quelccaya ice cap Cordillera Blanca, Peru; the Dunde and Guliya ice caps in western China; the Pamirs and Tien Shan of the former USSR; Lewis Glacier on Mount Kenya, Africa; as well as polar region programs in both Greenland and Antarctica.

Reliable meteorological observations for climate reconstruction are limited or absent prior to A.D. 1850 for much of the Earth. This is especially true for tropical South America and the Tibetan Plateau region of Central Asia. Scientists widely recognize ice sheets and ice caps as libraries of atmospheric history from which past climatic and environmental conditions can be extracted. However, much climatic activity of significance to humanity may not be strongly expressed in (or extend to) the polar ice caps. Fortunately, ice records can be recovered from both polar ice sheets and a select few high-altitude, low-, and midlatitude ice caps. These latter sites provide long-term records of El Niño/Southern oscillation and monsoon variability from regions where most of the Earth's population is concentrated.

The Tibetan Plateau, the largest on Earth, covers an area half the size of the United States. With a mean elevation of approximately 5 km, the plateau contains many glaciers and ice caps. The heating of the plateau drives the regionally intense monsoon system and causes perturbations in global circulation patterns. The climatic and environmental histories reconstructed from the frozen archives in Tibet provide a very important perspective of past variations in the monsoon system. Model results suggest that the central part of the Asian continent, because it is far from the mitigating influence of oceans, may be one of the first places to exhibit an unambiguous signal of the anticipated "enhanced greenhouse warming."

In 1992 I, along with my Chinese colleague Dr. Yao Tandong, led a team of 21 Chinese, Russian, and American scientists to the Guliya ice cap. Located in the far Western Kunlun Mountains of Tibet, the ice cap stands at an elevation of 6710 m (a site higher than Mount McKinley). Our team joined a caravan of two six-wheel-drive trucks and two four-wheel-drive vehicles to make the five-day journey along the only road south of Kashi (Kashgar). Once we reached our destination, we recovered many ice cores, including a 309-m-long sample.

Although the Lanzhou Institute of Glaciology and Geocryology assigned the six-wheelers to ensure successful transport of the cores, recovering and keeping them frozen presented a major challenge. We used two drills in the recovery process. The first was an electromechanical device that drilled to 200 m. As temperatures warmed with depth, a thermal drill was used to drill to bedrock, 309 m below the surface. Pits excavated in the surface of the ice cap provided the necessary refrigeration to keep the cores temporarily frozen. Two shallow holes were drilled adjacent to the main borehole, where special cylindrical containers were lowered on the end of ropes to a depth of 10 m. These cryopak (blue ice) containers were filled with chemicals that froze in the —16°C temperatures that are common at this depth. After freezing, the cryopak containers were packed with the ice cores into special insulated boxes that provided the necessary refrigeration for the four-day journey to Kashi. At the end of the field season, we intended to transport these ice cores, which weighed a total of nearly 2050 kg, by truck, cross-country through a muddy plain to the main road.

Another unanticipated (and critical) aspect of the logistics involved restricted access through the northernmost pass across the Western Kunlun Mountains. The pass was open only three days per month for northbound traffic. The maximum time the ice would remain intact without refreezing the cryopak was five days, and the trip to Kashi, driving 24 hours a day, required four days. Thus, the timing of our transit through the pass was critical. We arrived with our trucks at the base of the pass at midnight on the day it opened, only to find a line of 45 trucks ahead of us, their drivers sound asleep beneath and beside their vehicles! It was our unfortunate job to wake the irritable drivers and persuade them to keep the traffic moving throughout the night so that we could get our cargo through the pass before it closed again. After reaching Kashi, the ice cores were placed in a freezer until they could be flown to Beijing, where they cleared customs and were then successfully transported 19,320 km to the Ohio State University. The ice cores, which at this writing are still being analyzed, are believed to contain the oldest ice recovered on earth. Dating well over 200,000 years, these cores may provide extraordinary insight into climatic patterns through the last four ice ages. The Guliya ice cap lies at 35° North latitude—the same distance from the equator as Oklahoma City. Thus, it is located within the range of latitudes where most human beings lived. Early results from these cores indicate an average accumulation of 650 mm H_2O equivalent per year. These results also show significant recent warming, with measurements of average oxygen isotopes for the period of 1986-1992 showing a 2 percent increase over the average for the period of 1935-1985.

Obtaining records of climate changes of the past is essential to our understanding of how the Earth's climate might change in the future. It is important to obtain a broad distribution of records both spatially and with altitude, as the climate of the Earth changes both horizontally and vertically through time. Unfortunately, as a result of the global warming trend, many of the low-altitude, high-elevation glacier systems are retreating and may soon disappear. While complete wastage of these ice masses may not occur for decades, increased free-water flow is destroying their structure, and therefore the valuable paleoclimate information they contain. Thus, there is a pressing need to obtain high-quality ice cores from these glaciers and ice caps as soon as possible.

Summary

1. The seasonal snow cover in the northern hemisphere appears in the Arctic during early autumn, grows in thickness as it expands southward to reach a late-winter maximum, and then retreats rapidly during the spring.
2. The snowline marks the lower limit of perennial snow; its altitude is controlled mainly by precipitation and summer temperature.
3. Glaciers, which constitute the bulk of the ice in the cryosphere, are permanent bodies of moving ice that consist largely of recrystallized snow. They can form only at or above the snowline, which is close to sea level in the polar regions and rises to high altitudes in the tropics.
4. Ice in a temperate glacier is at the pressure melting point, and liquid water exists at the base of the glacier; ice in a polar glacier is below the pressure melting point and is frozen to the rock on which it rests.
5. The mass balance of a glacier is measured in terms of accumulation and ablation. The equilibrium line separates the accumulation area from the ablation area and marks the level on the glacier where net gain is balanced by net loss.
6. Temperate glaciers move as a result of internal flow and basal sliding. In polar glaciers, which are frozen to their bed, motion is much slower and involves only internal flow. Surges involve extremely rapid flow, probably related to excessive amounts of water at the base of a glacier.
7. A fjord glacier with a base below sea level will begin an irreversible calving retreat if its terminus becomes ungrounded and recedes into deepening water upfjord. Retreat ends only when the glacier again terminates in water too shallow for appreciable calving to occur.
8. Ice cores extracted from polar glaciers contain natural archives of changing environmental conditions in the form of changing oxygen isotope ratios, samples of ancient atmospheric gases, dust concentrations, acid volcanic fallout, and organic particles.
9. Sea ice thinly covers vast areas of polar ocean and is highly sensitive to changes in climate and ocean conditions. Perennial sea ice persists for at least several years and reaches thicknesses of 3 to 4 m, whereas seasonal ice that forms and disappears annually is less than 2 m thick. Convergent deformation produces linear pressure ridges where ice is unusually thick.
10. When sea water freezes, dense, cold, saline water is produced that sinks into the deep ocean to form deep water and bottom water. The high albedo of sea ice influences global climate, helping to create a steep pole-to-equator temperature gradient.
11. Permafrost, a common feature of periglacial zones, is mainly confined to areas where annual

air temperature is at least — 5°C. It reaches maximum thicknesses of at least 1500 m and is believed to have formed during glacial ages in subfreezing landscapes not covered by continental ice sheets.

12. Permafrost can present unique engineering problems, for thawing commences when the vegetation cover is broken, leading to collapse and extreme instability of the ground surface.

Important Terms to Remember

ablation (p. 259)
accumulation (p. 259)
albedo (p. 268)
calving (p. 262)
cirque (p. 254)
crevasse (p. 261)
cryosphere (p. 252)
equilibrium line (p. 259)
fjord (p. 255)
glacier (p. 254)
glacier ice (p. 256)
lead (in sea ice) (p. 267)
moraines (p. 263)
mass balance (p. 259)
periglacial (p. 268)
permafrost (p. 269)
polar glacier (p. 258)
pressure melting point (p. 257)
sea ice (p. 265)
snowline (p. 252)
surge (p. 263)
temperate glacier (p. 257)

Questions for Review

1. What is the snowline and how are glaciers related to it?
2. Describe the steps in the conversion of snow to glacier ice.
3. On what characteristics are temperate glaciers distinguished from polar glaciers?
4. Why does the position of the equilibrium line provide a rough estimate of a glacier's mass balance?
5. Why is there a time lag between a change of climate and the response of a glacier's terminus to the change?
6. In what ways does ice temperature influence the way a glacier moves?
7. Describe the unique motions of surging and calving glaciers, and indicate how their fluctuations are related to climate.
8. What factors control the distribution and thickness of sea ice?
9. How does sea ice influence climate? ocean circulation?
10. What is the active layer in permafrost terrain, and how does it form?
11. Where would you expect to find permafrost at latitudes of less than 40°, and why?
12. Describe what potential foundation problems a home builder might encounter in northern Alaska if the contractor were to clear the building site of vegetation and begin construction on the exposed ground surface.

Questions for A Closer Look

1. Why might a large mass of rock debris falling onto the surface of a glacier cause the glacier terminus to advance?

Questions for Discussion

1. Suggest ways in which changes in the solid Earth, such as uplift, earthquakes, and volcanism, might affect the distribution and fluctuations of glaciers?
2. One way to estimate the altitude of a valley glacier's equilibrium line is to calculate the median altitude of the glacier (the difference in altitude between the glacier's head and terminus, divided by 2). Obtain topographic maps from your library (for example, the glacier-clad mountains of southern Alaska, British Columbia, Alberta, or Washington State) and calculate the equilibrium-line altitudes of several valley glaciers. What differences do you obtain for nearby glaciers within a single mountain range? How might the differences be explained?
3. Huge pools of petroleum likely underlie the Arctic continental shelves of Alaska, Canada, and Russia. In exploiting such petroleum resources, what problems might be encountered that are related to the presence of sea ice and permafrost?

CHAPTER 11

The Changing Face of the Land

In 1985, a major eruption of glacier-mantled Nevado del Ruíz volcano in Colombia generated a large, destructive mudflow that descended without warning upon the city of Armero, killing at least 20,000 of its people.

Death in the Night

As the human population increases, settlements and roads expand across the landscape, often in areas where potential landslide hazards are not recognized. Then, when a landslide occurs, the loss of life and property can be devastating.

The high Andes of South America include numerous active volcanoes and rugged peaks thrust up along converging lithospheric plates. The steep, unstable slopes rise above densely populated valleys, a combination spelling potential disaster in a landscape prone to earthquakes and volcanic eruptions. In Colombia, the Andes culminate in a group of lofty active volcanoes lying west of Bogota. One peak, Nevado del Ruiz [5400 m (18,000 ft)], has a history of volcanic activity extending back to at least 1595, when thunderous eruptions spread tephra across the landscape and volcanic mudflows rushed down several valleys.

In late 1984 the volcano began belching clouds of steam and ash, activity that continued through the autumn of 1985. People in the city of Armero, far downvalley from the volcano, grew alarmed. The local authorities appeared unconcerned and reassured them, even though recent geologic studies had disclosed a history of repeated large volcanic mudflows. In early November, when the volcano showed signs of increasing activity, geologists warned that such mudflows could pose a serious danger for Armero in the event of an eruption. At 3 P.M. on November 13, a technical emergency committee urged that Armero be evacuated, but the warning went unheeded. That night, as the local radio station played cheerful music and urged people to be calm, the volcano erupted. Torrents of water released from rapidly melting ice and snow near the summit eroded soil, vegetation, rock, and sediment from the landscape, sending huge waves of muddy debris surging downslope into surrounding valleys. The largest of several mudflows moved headlong in the direction of Armero. Just after 11 P.M., as most of the townspeople were sleeping soundly, a turbulent wall of mud came rushing out of a mountain canyon and inundated the city. At least 20,000 citizens of Armero perished, buried in a tomb of sulfurous volcanic mud. The geologists' prediction, based on careful analysis of the geologic record, proved correct. Had their warning been heeded, the tragedy might have been avoided.

THE EARTH'S VARIED LANDSCAPES

One of the most striking things about our planet is its amazing variety of natural landscapes. Who can fail to be impressed by a view of the majestic snow-covered Himalaya rising abruptly from the plains of India, or the lofty Andes of South America with their array of active volcanic cones? Equally impressive are the vast subtropical deserts of North Africa, Australia, and the Arabian Peninsula; the dense, flat jungle terrain of South America's Amazon basin; and the glaciated landscape of eastern Canada and the north-central United

States with its myriad lakes, streams, and undulating topography.

Even a casual look at the land surface can raise some basic questions in our mind: How can the Earth's varied landscapes be explained? Are landscapes eternal, as many of our forebears once thought, or do they change with time? And what clues do landscapes hold about the history of the Earth's mobile lithosphere and past climates?

Everywhere people live, they are changing the face of the land. The construction of dams across streams, excavations for buildings or waste disposal sites, and the building of highways and cities all modify the landscape. To the untrained observer, however, it may not be immediately apparent that the landscape is also being modified by natural processes. Sometimes these processes act rapidly, and their effects are obvious: a landslide that removes part of a hillside and forces an adjacent stream into a new path is a geologically instantaneous event. Most natural processes, nevertheless, operate at much slower rates, so slow that their effects may be barely discernible over a human lifetime. And yet, these small changes can add up until, over geologically long intervals of time, their effect is dramatic.

WEATHERING AND EROSION

Overwhelming geologic evidence shows that the Earth's surface is constantly changing, although the rate and magnitude of change vary considerably from region to region. Such changes reflect a contest between major tectonic forces that raise the lithosphere and the force of gravity that exerts a downward pull on the same rocks.

Aided by physical and chemical processes that break down rock, and by gravity, various erosional agents transfer rock debris from high places to low places. The net result is the progressive sculpture of the land into a surface of varied **topographic relief** (the difference in altitude between the highest and lowest points on a landscape). Relief varies from one region to another because surface geologic processes operate under different climatic conditions on rocks of different type and structure, and consequently of differing resistance to erosion. (See the "Guest Essay" at the end of this chapter for an interesting comparision of erosion on Mars.)

From Rock to Regolith

Whether rapid or slow, rock is physically broken up and chemically altered throughout the zone where the lithosphere, hydrosphere, biosphere, and atmosphere mix. This zone extends from the ground surface downward to whatever depth air and water penetrate. Within it, the rock constitutes a porous framework full of fractures, cracks, and other openings, some of which are very small but all of which make the rock vulnerable to attack by aqueous solutions. Given sufficient time, the result is conspicuous decomposition and disintegration of the rock, processes known collectively as **weathering.**

We have all seen weathering in action. We may have visited a cemetery and strained to read the inscription on an old marble tombstone so modified by weathering that the characters were barely legible (Fig. 11.1). Or we may have been seated around a

Figure 11.1 Because marble is composed of soluble calcite, this marble tombstone standing in a New England cemetery since the early nineteenth century shows the corrosive effects of the carbonic acid present in rainwater. Over the years the rock surface has been slowly dissolved, making the once sharply chiseled inscription illegible.

roaring campfire and suddenly been struck by flying rock fragments as a rock next to the fire exploded, because it was composed of minerals that expand at different rates when heated. Such examples show that weathering can involve both chemical and physical processes.

Chemical Weathering

The minerals of igneous and metamorphic rocks that have crystallized within the Earth's crust at high pressure and temperature are chemically unstable at the lower temperatures and pressures at the surface. When such rocks are uplifted and eventually exposed, therefore, their mineral components are chemically changed into new, more stable minerals. **Chemical weathering**, then, is the decomposition of rocks and minerals as chemical reactions transform them into new chemical combinations that are stable at or near the Earth's surface.

The principal agent of chemical weathering is a weak solution of carbonic acid (H_2CO_3), formed as falling rainwater dissolves small quantities of atmospheric carbon dioxide (Table 11.1, eq. 1). As the water moves downward and laterally beneath the ground surface, additional carbon dioxide is dissolved from decaying vegetation. Thus, chemical weathering is dependent on the interaction of the atmosphere, the hydrosphere, and the biosphere to produce the weakly acidic solution that attacks the upper part of the lithosphere.

The chemical reaction that decomposes the common rock-forming mineral potassium feldspar provides a good example of chemical weathering (Table 11.1, eq. 2). A molecule of carbonic acid dissociates in water to form a hydrogen ion (H^+) and a bicarbonate ion [(HCO_3^-)]. H^{1+} ions enter the potassium feldspar and replace potassium ions (K^+), which then leave the crystal and pass into solution. Water combines with the remaining aluminum silicate molecule to create *kaolinite*, a new clay mineral not present in the original rock.

Weathering of a rock like basalt, which is composed of feldspars and Fe-Mg bearing minerals, pro-

Table 11.1
Common Chemical Weathering Reactions

1. Production of carbonic acid by solution of carbon dioxide:

H_2O	+	CO_2	®	H_2CO_3	®	H^+	+	$(HCO_3)^-$
Water		Carbon dioxide		Carbonic acid		Hydrogen ion		Bicarbonate ion

2. Hydrolysis of potassium feldspar:

$4KAlSi_3O_8$	+	$4H^+$	+	$2H_2O$	®	$4K^+$	+	$Al_4Si_4O_{10}(OH)_8$	+	$8SiO_2$
Potassium feldspar		Hydrogen ions		Water		Potassium ions		Kaolinite		Silica

3. Oxidation of iron (Fe^{2+}) oxide to form goethite:

$4FeO$	+	$2H_2O$	+	O_2	->	$4FeO \cdot OH$
Iron oxide		Water		Oxygen		Goethite

4, Dissolution of carbonate rock by carbonic acid:

$CaCO_3$	+	H_2CO_3	®	Ca^2+	+	$2(HCO_3)^-$
Calcium carbonate		Carbonic acid		Calcium ion		Bicarbonate ions

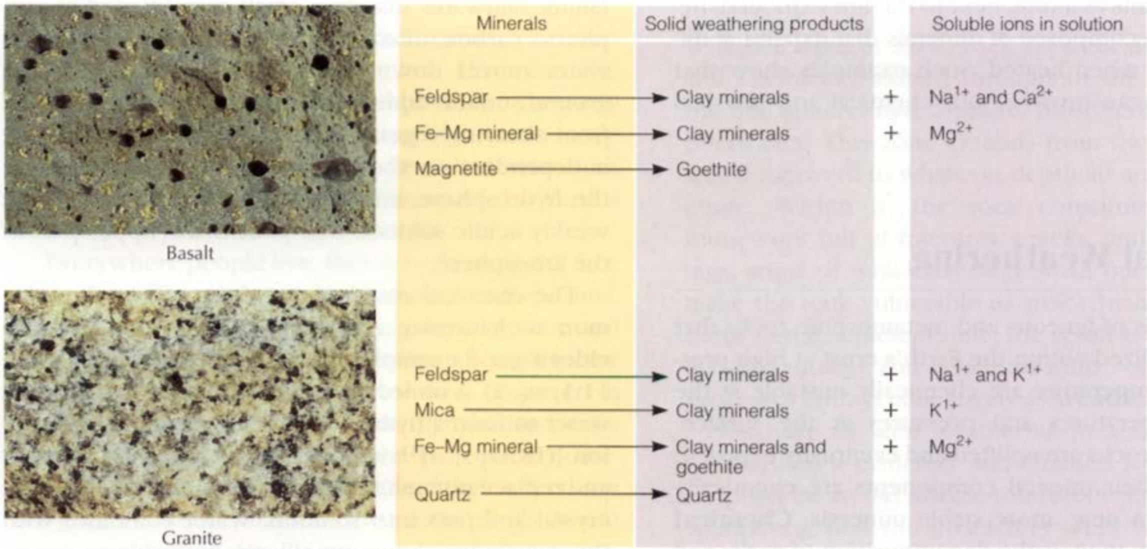


Figure 11.2 When basalt is chemically weathered, its minerals are converted to new clay minerals and to goethite. Soluble ions are carried away in groundwater. When granite weathers, clay minerals, goethite, and soluble ions are also produced, as well as grains of quartz, a mineral that is resistant to chemical decay.

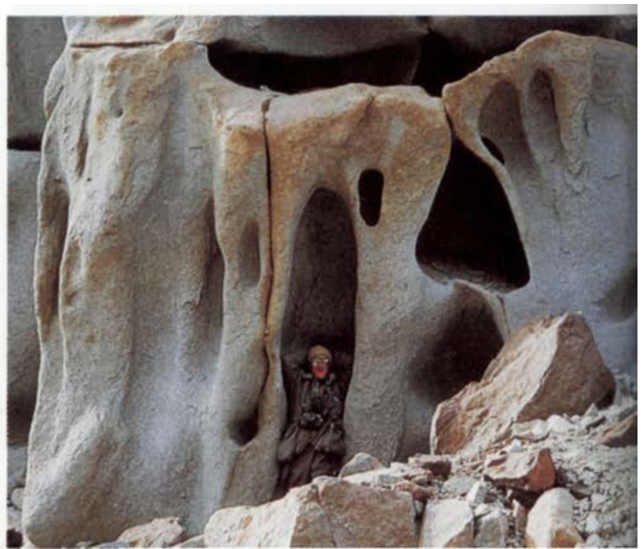
duces clay minerals and *goethite*, a weathering product of magnetite (Table 11.1, eq. 3; Fig. 11.2). When a granite weathers, clay minerals and goethite are also produced from the feldspar and mica it contains; however, the quartz, because it is resistant to chemical weathering, survives unaltered as a weathering product.

Physical Weathering

Sometimes regolith consists of fragments identical to the adjacent bedrock. However, the mineral grains are unweathered or only slightly weathered, indicating little or no evidence of chemical alteration. Instead, the fragments must have experienced **physical weath-**

Figure 11.3A This granite outcrop in Yosemite National Park, California, displays sheetlike Joints, giving a stepped appearance to the mountain slope. The jointing is thought to result from progressive removal of overlying rock, leading to reduced pressure. This causes expansion of the uppermost rock, which fractures along planes parallel to the land surface.

Figure 11.3B Granite on the side of Gondola Ridge in Antarctica is so intensely weathered that it resembles Swiss cheese. Such cavernous weathering is produced by crystallization of salt in small cavities and along grain boundaries.



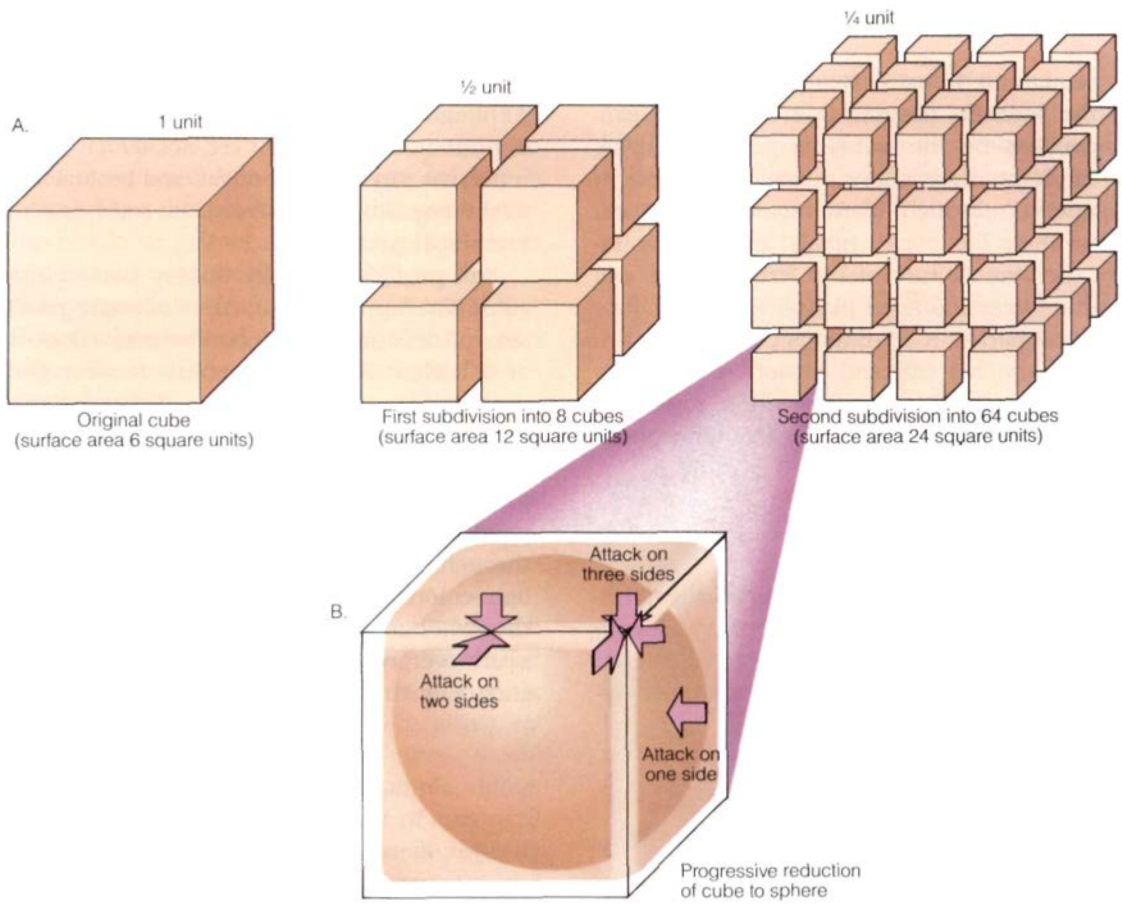


Figure 11.4 Weathering causes progressive subdivision of rocks. A. Each time a cube is subdivided by slicing it through the center of each edge, the aggregate surface area doubles, thereby increasing the effectiveness of chemical attack. B. Solutions moving along joints that separate cube-shaped blocks of rock attack corners, edges, and sides at rates that decline in that order, because the number of corresponding surfaces under attack are 3, 2. and 1, respectively. Corners therefore become rounded, and eventually the blocks are reduced to spheres. Once a spherical form has been reached, chemical attack is distributed over the entire surface, and no further change in form occurs.

ering, which is the disintegration (physical breakup) of rocks.

A variety of natural physical processes are effective in physical weathering: (1) A rock mass buried deep beneath the land surface is subjected to immense confining pressure. However, as erosion gradually removes the overlying rock, the pressure is reduced and the buried rock mass adjusts by expanding upward. In the process, sheetlike fractures develop parallel to the surface (Fig. 11.3A). (2) Ions in the groundwater moving through fractured rock can precipitate out to form salts. The enormous forces exerted by salt crystals growing in cavities or along mineral grain boundaries of a rock can easily lead to rupture or disaggregation (Fig. 11.3B). (3) When water freezes, its volume increases about 9 percent. If water freezes in a confined crack, the resulting stresses can be so great that the rock is wedged apart. (4) Fire, too, can be a very effective agent of weathering. An intense forest or brush fire can overheat the outer part of a rock, causing it to expand, fracture, and break away. Repeated fires can thereby significantly reduce the size of rocks. (5) Finally, plant roots extending along cracks in a rock can slowly wedge the rock apart.

Although physical weathering is distinct from chemical weathering, the two processes generally work hand in hand, and their effects are inseparably blended. The effectiveness of chemical weathering increases as the exposed surface area increases, and surface area increases greatly whenever a large unit is divided into successively smaller units (Fig. 11.4). Repeated subdivision leads to a remarkable result; whereas one cubic centimeter of rock has a surface area of 6 cm^2 (0.9 in^2), when subdivided into particles the size of the smallest clay minerals the total surface area now exposed to weathering increases to nearly 40 million cm^2 (6.2 million in^2).

Soils

The physical and chemical weathering of solid rock is the initial step in soil formation. However, soil also contains organic matter mixed with the mineral component. This organic part is an essential part of the usual definition of **soil:** the part of the regolith that can support rooted plants.

The organic matter in soil is derived from the decay of dead plants and animals. Living plants are nourished by the nutrients released from decaying organisms, as well as by the nutrients released during weathering. Plants draw these nutrients upward, in water solution, through their roots. Therefore, throughout their life cycle, plants are directly involved in the manufacture of the fertilizer that will nourish future generations of plants. These activities are an integral part of a continuous cycling of nutrients through the regolith and biosphere. With its partly mineral, partly organic composition, soil forms an important bridge between the Earth's lithosphere and its teeming biosphere.

The Soil Profile

As bedrock and regolith weather, soil gradually develops from the surface downward, producing an identifiable succession of nearly horizontal weathered zones called **soil horizons.** Each horizon has distinctive physical, chemical, and biological characteristics. Although soil horizons may resemble a sequence of deposits, or layers, they are not strata. Instead, they represent physical, chemical, and biological changes to the regolith. Collectively, the soil horizons constitute a **soil profile** (Fig. 11.5).

Soil profiles generally display two or more horizons. The uppermost horizon of some profiles consists of decomposing organic matter *(O horizon).* If an O horizon is absent, then an *A horizon* generally is the uppermost horizon. Typically, it is dark gray or black because decomposed plant and animal tissues are mixed with the mineral matter. The A horizon has lost some of its original substance through the downward transport of clay particles and, more important, through the chemical removal of soluble minerals. A light-colored *E horizon,* sometimes present beneath the A horizon in acidic soils, is often developed beneath evergreen forests. The *B horizon* underlies the surface horizon(s) and commonly has a brownish or

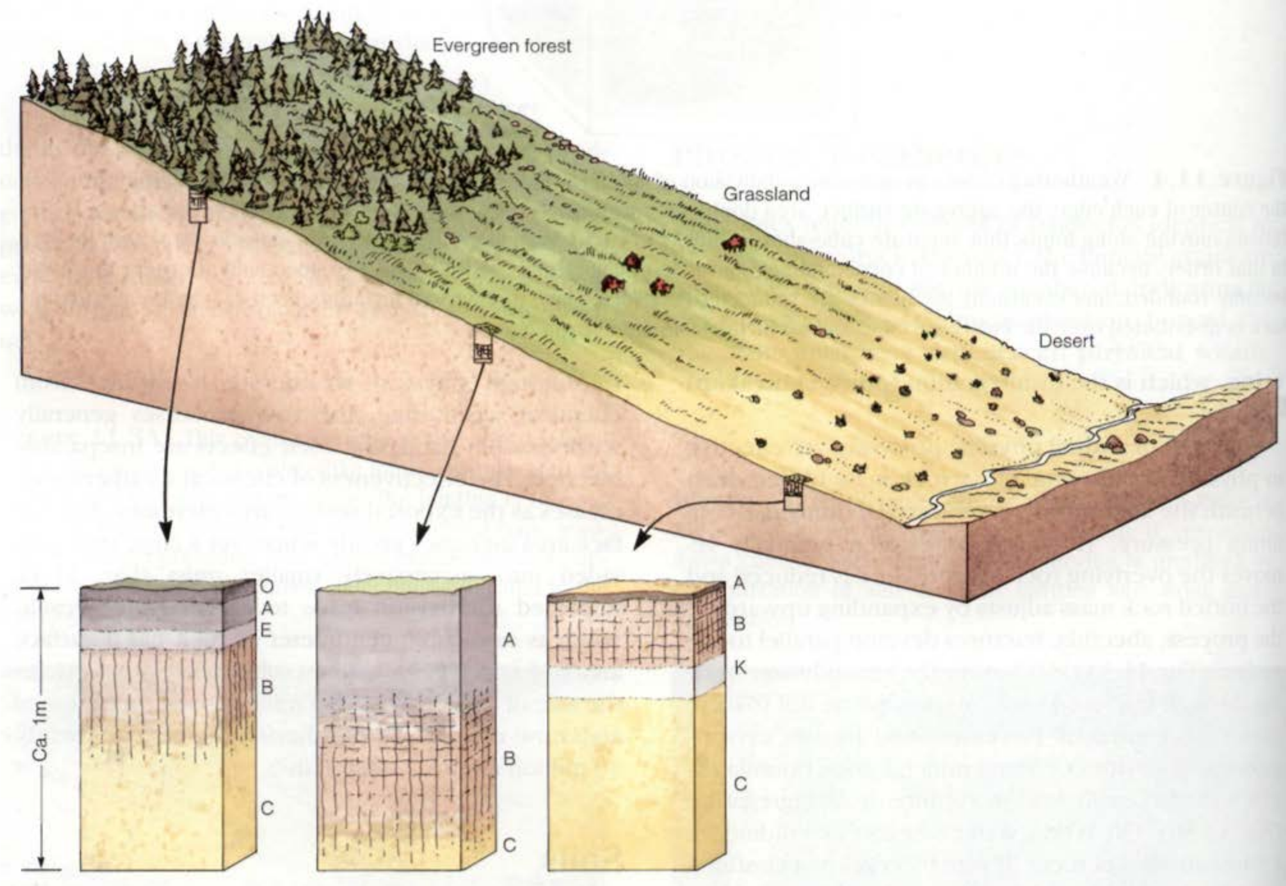


Figure 11.5 Soils vary across the landscape, as shown by this example of three soil profiles from forest, grassland, and desert regions. Differences are explainable in terms of regolith composition, slope steepness, vegetation cover, soil biota, climate, and the time required to develop the profile.

reddish color. This horizon is enriched in clay and/or iron and aluminum hydroxides produced by the weathering of minerals within the horizon and also transported downward from the A horizon. The B horizon often has a distinct structure that causes it to break into blocks or prisms. A *K horizon,* present in some arid-region soils beneath the B horizon, is densely impregnated with calcium carbonate that coats all mineral grains and constitutes up to 50 percent of the volume of the horizon. The *C horizon* is the deepest horizon and constitutes the parent regolith in various stages of weathering, but it lacks the distinctive properties of the A and B horizons. Oxidation in the C horizon generally imparts to it a light yellowish-brown color.

Soil Types

An astute observer traveling across the landscape will note that soils are not everywhere the same (Fig. 11.5). Different soils result from the influence of six factors: climate, vegetation cover, soil organisms, regolith composition, topography, and time. A soil forming under prairie grassland, for example, differs from soil in an evergreen forest or that of a tropical rainforest. The character of a soil may change abruptly as we move from basalt to limestone or from a gentle slope to a steep slope, and it also will change with the passage of time. Such differences make it possible for soil scientists to classify and map soils across the landscape in much the same way that geologists classify and map rocks.

Ancient Soils

A surface soil buried by sediment or lava becomes part of the geologic record. Its top is therefore an unconformity. Buried soils have been identified in the rocks and sediments of many different ages (Fig. 11.27), and distinctive ones have been used by geologists to subdivide, correlate, and date ancient sedimentary deposits.

Soil Erosion

With world population now approaching 5 billion, increasing competition for a finite amount of agricultural land is causing serious erosion of soils. Although soil erosion results from natural changes in topography, climate, or vegetation cover, the results of human activities have overwhelmed natural systems in many parts of the world. Widespread felling of trees has led to accelerated rates of surface runoff and destabilization of soils due to loss of anchoring roots. Soils in the humid tropics, when stripped of their natural vegetation cover and cultivated, quickly lose their fertility (Fig. 11.6). So widespread are the effects of soil erosion and degradation that the problem has been described as "epidemic." Because agriculture is the foundation of the world economy, progressive loss of soil signals a potential crisis that could undermine the economic stability of many countries.

Figure 11.6 Widespread deforestation in Rondonia, Brazil has devastated a formerly luxuriant rain forest and led to accelerated runoff and erosion. Soils on this landscape quickly lose their natural fertility when forest is converted to crops or grazing land, leaving a degraded landscape with little value.

Much of the topsoil eroded from agricultural lands is transported down rivers and deposited along valley floors, in marine deltas, or in reservoirs behind large dams. The resulting impact on society, often unanticipated, can be significant. For example, the designers of a major dam and reservoir in Pakistan projected a life expectancy for the reservoir of at least a century. However, increased population pressure on the region above the dam has resulted in greatly increased soil erosion, leading to such a high rate of sediment production that the reservoir is now expected to be filled with eroded soil within 50 years, making it unusable.

The upper layers of a soil contain most of the organic matter and nutrients that support crops. When the A and B horizons are eroded away, not only the fertility but also the water-holding capacity of a soil diminishes. Because it generally takes between 80 and 400 years to form one centimeter of topsoil, soil erosion, for all practical purposes, is tantamount to mining the soil. It is estimated that farmers in the United States are now losing about 5 tons of soil for every ton of grain they produce, while in India the rate of soil erosion is estimated to be more than twice as high. Worldwide, the most productive soils are being depleted at the rate of 7 percent each decade. One recent estimate projected that, as a result of excessive soil erosion and increasing population, only two-thirds as much topsoil will be available to support each person at the end of the century as was available in 1984.

Although soil erosion and degradation are severely impacting many countries, effective control measures can substantially reduce these adverse trends. One method of reducing soil loss involves crop rotation. A study in Missouri showed that land which lost 49.25 tons of soil per hectare when planted continuously in corn lost only 6.75 tons per hectare when corn, wheat, and clover crops were rotated. In this case, the bare land exposed between rows of corn is far more susceptible to erosion than land planted with a more continuous cover of wheat or clover.

The most serious soil erosion problems occur on steep hillslopes. In Nigeria, land planted with cassava (a staple food source) and having a gentle 1 percent slope lost an average of 3 tons of soil per hectare each year. On a 5 percent slope, however, the annual soil loss increased to 87 tons per hectare. At this rate, a 15 cm (6 in) thickness of topsoil would disappear in a single generation (about 20 years). On a 15 percent slope, the annual erosion rate increased to 221 tons per hectare, a rate that would remove all topsoil within a decade. Despite these grim statistics, steep slopes can be exploited through appropriate terracing.

FROM MOUNTAIN TOP TO OCEAN SHORE

As any mountain climber will affirm, mountains can be dangerous, unstable places. Falling rocks, debris-clad glaciers, and boulder-filled streams provide vivid evidence of the downslope movement of rock debris loosened and broken up by mechanical and chemical weathering. As gravity pulls on the debris, it falls, slides, and tumbles downslope, where it is picked up by glaciers, streams, or wind and carried farther. Battered by constant impact and abrasion, stream sediment is worn down and sorted as it is moved along until by the time it reaches the ocean, the bulk of it consists of sand, silt, and clay.

Mass-wasting

A smooth, vegetated slope may outwardly appear stable and show little obvious evidence of geologic activity. If we were to examine the regolith at and beneath the surface, however, we might find some rock particles derived from bedrock that occurs only farther upslope. We would deduce that these particles have moved downslope. A time-lapse motion picture of such a hillslope, which greatly speeds up any motion, would make the slope appear almost alive and constantly changing. Much of the recorded motion would be the result of **mass-wasting**, the downslope movement of regolith under the pull of gravity. This definition implies that the motive force is gravity rather than a transporting medium such as water, wind, or ice. Mass wasting is not confined to the land. It also occurs on lake floors and over vast areas of the seafloor wherever slopes exist.

Role of Water

Water is almost always present within rocks and regolith near the Earth's surface. Although, by definition, mass-wasting does not involve water as a transporting fluid, it nevertheless plays an important role. Loose sediments behave in different ways depending on whether they are dry or wet, a fact well known to anyone who has constructed a sand castle at the beach. Dry sand is unstable and difficult or impossible to mold, but when some water is added, the sand can be shaped into vertical castle walls. The water and sand grains are drawn together by a force called *capillary attraction.* The attraction results from surface tension, a property of liquids that causes the exposed surface to contract to the smallest possible area. This force tends to hold the wet sand together as a cohesive mass. However, the addition of too much water saturates the sand and turns it into a slurry that easily flows away, as the sand-castle builder sees with dismay when the rising tide on the beach destroys the elaborate work of an afternoon.

On a much larger scale, the same phenomenon happens in nature. Moist or weakly cemented fine-grained sediments, such as fine silt and clay, may be so cohesive that they can stand in near-vertical cliffs. However, if the silt or clay becomes saturated with water and the internal pressure of water trapped between particles rises above some critical limit, these fine-grained sediments may also become unstable and begin to flow.

The movement of some large masses of rock may also involve the effects of increased internal water pressure. If voids along a nearly horizontal surface separating two rock masses are filled with water, and the water is under pressure, a buoying effect can result. Under such conditions, the water pressure may be high enough to support the weight of the overlying rock mass, thereby reducing friction along the contact. Once a critical limit is reached, sudden displacement of the rock may occur. An analogous example can make driving in a heavy rainstorm extremely dangerous: when water is compressed beneath the wheels of a moving car, the increasing fluid pressure can cause the tires to "float" off the roadway. The driver then quickly loses control of the vehicle, a condition known as hydroplaning.

Landslides

Landslide is a general term for a variety of mass-wasting processes that result in the downslope movement of a mass of bedrock or regolith, or a mixture of the two, under the influence of gravity. The composition and texture of the sediment involved, the amount of water and air mixed with the sediment, and the steepness of slope all influence the type and velocity of a landslide (Table 11.2).

One way to distinguish among landslides is to separate them into those involving (1) sudden *slope failures*, the downslope movement of relatively coherent masses of rock or regolith by slumping, falling, or sliding; and (2) *sediment flows*, the downslope flow of sediment mixed with water and air. Among sediment flows, processes are distinguished on the basis of their velocity and the concentration of sediment in the flowing mixture.

Young stratovolcanoes, often rising to heights of more than 1000 m (3300 ft) and with slopes typically exceeding 20°, offer some spectacular examples of slope failures and sediment flows. Such volcanoes consist of unstable piles of jointed lava flows and interstratified loose tephra. When the flank of a volcano collapses, the results can be devastating. The mighty eruption of Mount St. Helens in 1980 illustrated how an active volcano can literally destroy itself by this process (Chapter 5). The massive landslide carried an estimated 2.8 km^3 (0.7 mi^3) of rock, ice, and regolith off the northern slope of the mountain and into the adjacent valley. The landslide, a rapidly moving debris avalanche, and accompanying lateral blast left a gaping hole where the summit had stood and a thick steaming pile of debris along the valley floor. Clear evidence of similar landslides can be found at many other volcanoes. A massive landslide about 300,000 years ago from the northwestern flank of Mount Shasta volcano in northern California was far larger. The chaotic deposit it formed extends nearly 50 km from the mountain and covers an area of at least 675 km^2 (260 mi^2). Its estimated volume of 45 km^3 (11 mi^3) is nearly 16 times that of the Mount St. Helens landslide deposit (Fig. 11.7).

Although landslides may occur for no apparent rea-

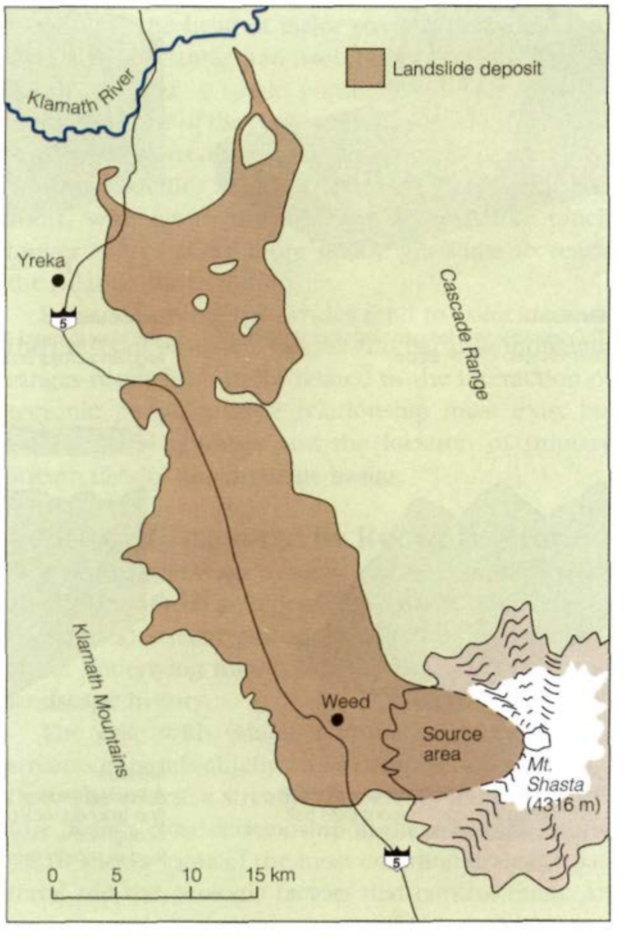
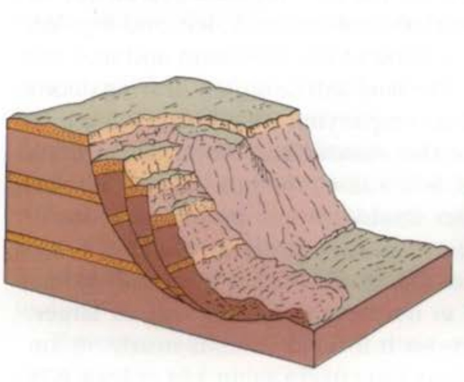
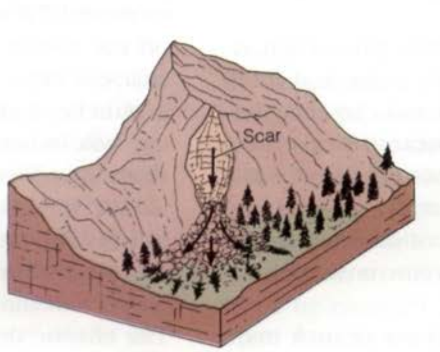


Figure 11.7 A massive landslide that originated on the northwestern flank of Mount Shasta volcano in northern California about 300,000 years ago produced a chaotic deposit of low hills that extends nearly 50 km from the summit and covers some 675 km^2.

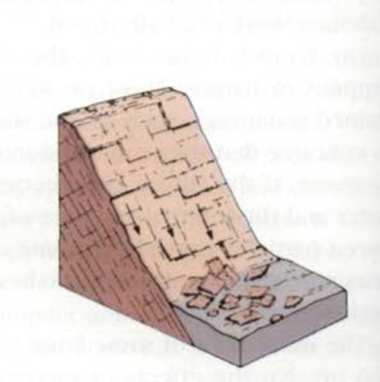
Table 11.2
Types of Landslides



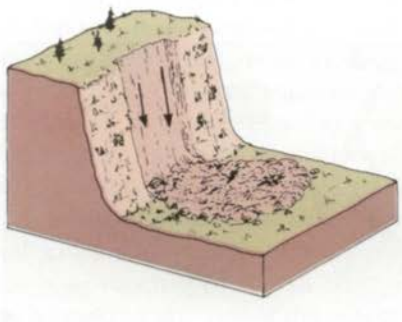
Slump
A slope failure in which a downward and outward rotational movement of rock or regolith occurs along a concave-up surface.



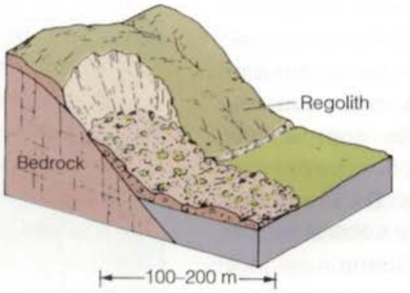
Rockfall
The free falling of detached bodies of bedrock from a cliff or steep slope.



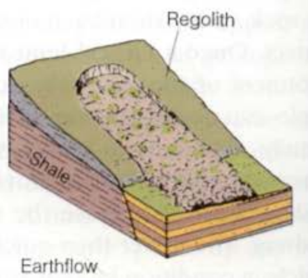
Rockslide
The sudden and rapid downslope movement of detached masses of bedrock across an inclined surface.



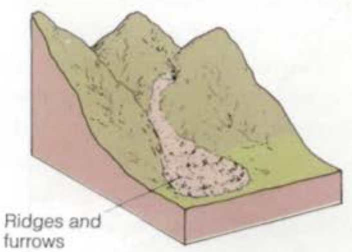
Debris fall
The relatively free fall or collapse of regolith from a steep cliff or slope.



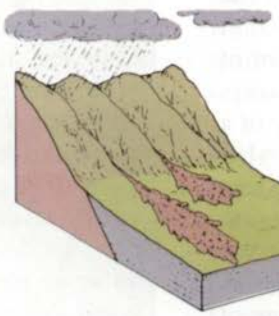
Debris slide
The slow to rapid downslope movement of regolith across an inclined surface.



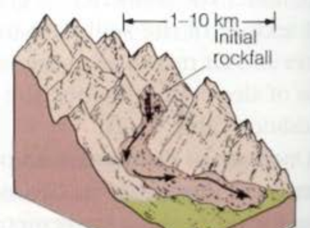
Earthflow
A flow of regolith with a velocity ranging from 10^{-5} to 10^{-1} meters/second.



Debris flow
The downslope movement of a mass of unconsolidated regolith more than half of which is coarser than sand.



Mudflow
A flowing mass of predominantly fine-grained rock debris that has a high enough water content to make it highly fluid (a rapidly moving type of debris flow).



Rock or debris avalanche
Often massive flow of rock or regolith moving at a high velocity (≥10 meters/second).

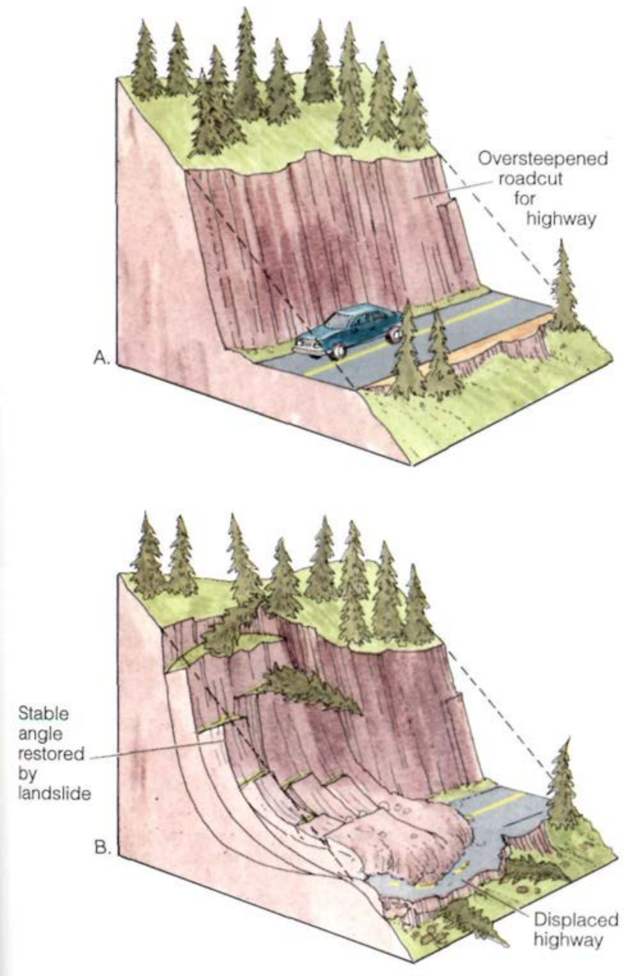


Figure 11.8 When a natural slope is oversteepened in building a road, failure can result. A. A highway cut exceeds the natural angle of the slope, producing an unstable situation. B. The oversteepened slope fails, and a landslide buries the road. The slope angle of the resulting deposit now is similar to the original natural one.

son, many are related to an unusual occurrence. (1) A major earthquake can trigger landslides throughout a large area. (2) Landslides often are associated with heavy or prolonged rains that saturate the ground and make it unstable. (3) A volcanic eruption, like that at Mount St. Helens in 1980, often triggers a variety of landslides, including mudflows that move rapidly into surrounding valleys. (4) Landslides often result when human activities modify natural slopes. Slides frequently occur where road construction significantly oversteepens natural slopes (Fig. 11.8). (5) A stream undercutting its bank can trigger a landslide, and pounding storm surf along a seacoast can also trigger landslides when steep bluffs are undercut.

Landscapes Produced by Running Water

Most of the Earth's land areas show the effects of running water. Except in extremely dry deserts and regions with a continuous ice cover, streams have shaped the land nearly everywhere.

Tectonic Control of Main Divides

All the continents except ice-covered Antarctica can be divided into large regions in which major through-flowing streams enter one of the world's major oceans. The line separating any two such regions is a *continental divide,* one of the major landscape elements of our planet. In North America, continental divides lie at the head of major streams that drain into the Pacific, Atlantic, and Arctic oceans (Fig. 11.9). In South America, a single continental divide extends along the crest of the Andes and divides the continent into two regions of unequal size. Streams draining the western (Pacific) slope of the Andes are steep and short, whereas to the east the streams take much longer routes along more gentle gradients to reach the Atlantic shore.

Because continental divides tend to coincide with the crests of mountain ranges and because mountain ranges result from uplift related to the interaction of tectonic plates, a close relationship must exist between plate tectonics and the location of primary stream divides and drainage basins.

Control of Drainage by Rock Structure

One of the best ways to view stream systems is from an airplane. From an altitude of 8 or 9 kilometers, (5 to 6 miles) stream patterns can tell us a great deal about underlying rock types, geologic structure, and landscape history.

The ease with which a formation is eroded by streams depends chiefly on its composition and structure. The course a stream takes across the land therefore bears a close relationship to these factors. Figure 11.10 shows some of the most common drainage patterns and the geologic factors that control them. An experienced geologist can use these drainage patterns to infer rock type, the orientation of a dipping rock unit, the manner in which the rocks are folded or offset, and the pattern and spacing of joints.

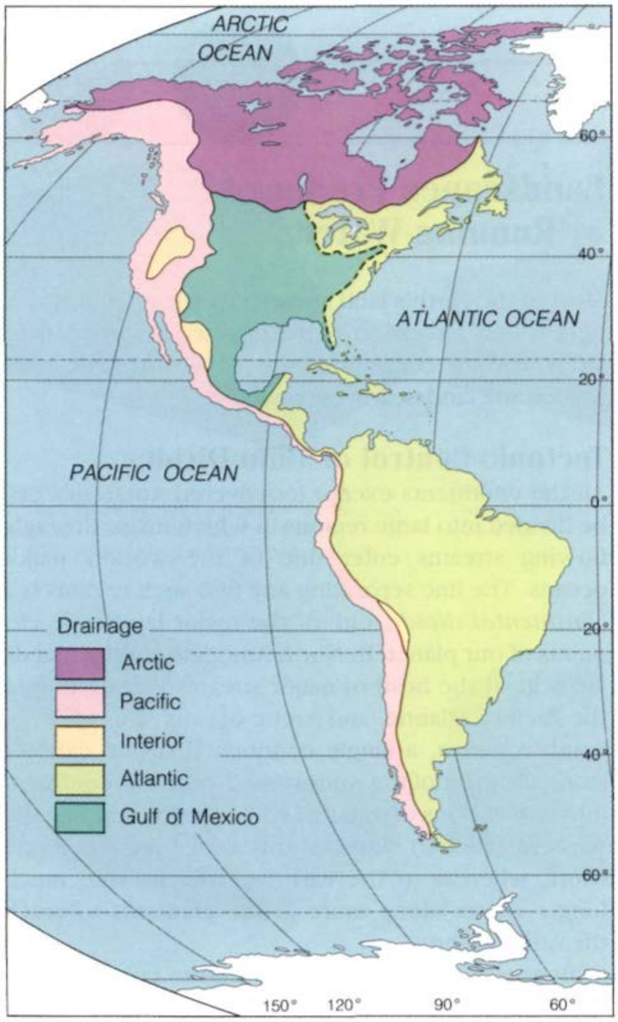


Figure 11.9 Map of the western hemisphere showing the location of major drainage divides. The continental divide separating streams draining to the Pacific, Arctic, and Gulf of Mexico in North America and to the Pacific and Atlantic in South America follows the crest of the high cordillera in both hemispheres. In eastern North America, the divide separating Atlantic and Gulf of Mexico drainage follows the Appalachian Mountains and much of the limit of ice sheet glaciation south of the Great Lakes.

Equally revealing are images of the Earth's surface taken from orbiting satellites, for they show a broad expanse of terrain in considerable detail. In Figure 11.11, the folded sedimentary rocks of the Appalachian Mountains near Harrisburg, Pennsylvania, are seen in striking detail because the alternating shales and sandstone units have been eroded differentially. The sandstones, being more resistant to erosion, stand high as steep linear ridges, whereas the more erodible shales underlie intervening, relatively flat valleys.

Stream Deposits

Distinctive stream deposits form along channel margins, valley floors, mountain fronts, and lake and ocean margins, for these all are places where stream energy changes. The lower Mississippi River and other large, smoothly flowing streams like it typically

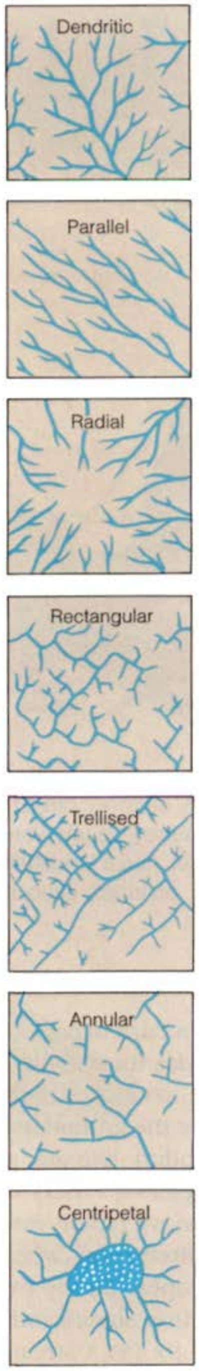


Figure 11.10 From stream patterns, geologists can infer something about the type and configuration of underlying rock and about the structural history of an area.

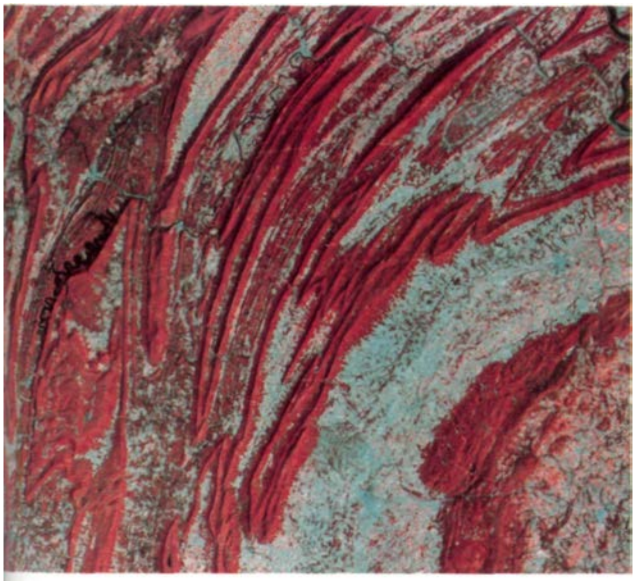


Figure 11.11 A satellite image of the region near Harrisburg, Pennsylvania, reveals a complicated series of northeast-trending ridges and valleys produced by differential erosion of sedimentary rocks. Ridges are underlain by resistant sandstones and conglomerates, while valleys are underlain by more-erodible shales. The folded structure of the rocks is clearly visible due to the pronounced topographic relief between the less-erodible and more-erodible strata.

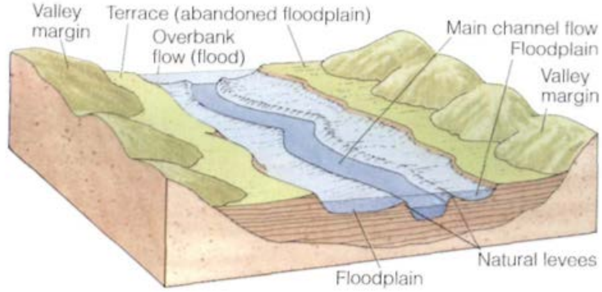


Figure 11.12 The major landforms of an alluvial valley.

deposit well-sorted layers of coarse and fine particles as they swing back and forth across a wide valley. During floods, as sediment-laden water flows out of the completely submerged channel, the depth, velocity, and turbulence of the water decrease abruptly at the channel margins, where the coarsest part of the suspended load is deposited to form a *natural levee* (Fig. 11.12). Farther away, finer silt and clay settle out across the stream's floodplain. a relatively flat region of valley floor that is periodically inundated by floodwater.

Most stream valleys contain terraces, which are floodplains abandoned when the stream cut downward to a lower level (Figs. 11.12 and 11.13). Typically, such downcutting occurs in response to tectonism or to a change in discharge, load, or gradient. In many stream valleys, terraces lying at various levels record a complex history of alternating deposition and erosion.

A large, swift stream flowing down a steep mountain valley can transport an abundant load of coarse sediment, but on leaving the valley the stream loses

Figure 11.13 Alluvial terraces adjacent to Cave Stream, South Island, New Zealand, record former floodplains that were abandoned when the stream incised its channel and reached a new level.

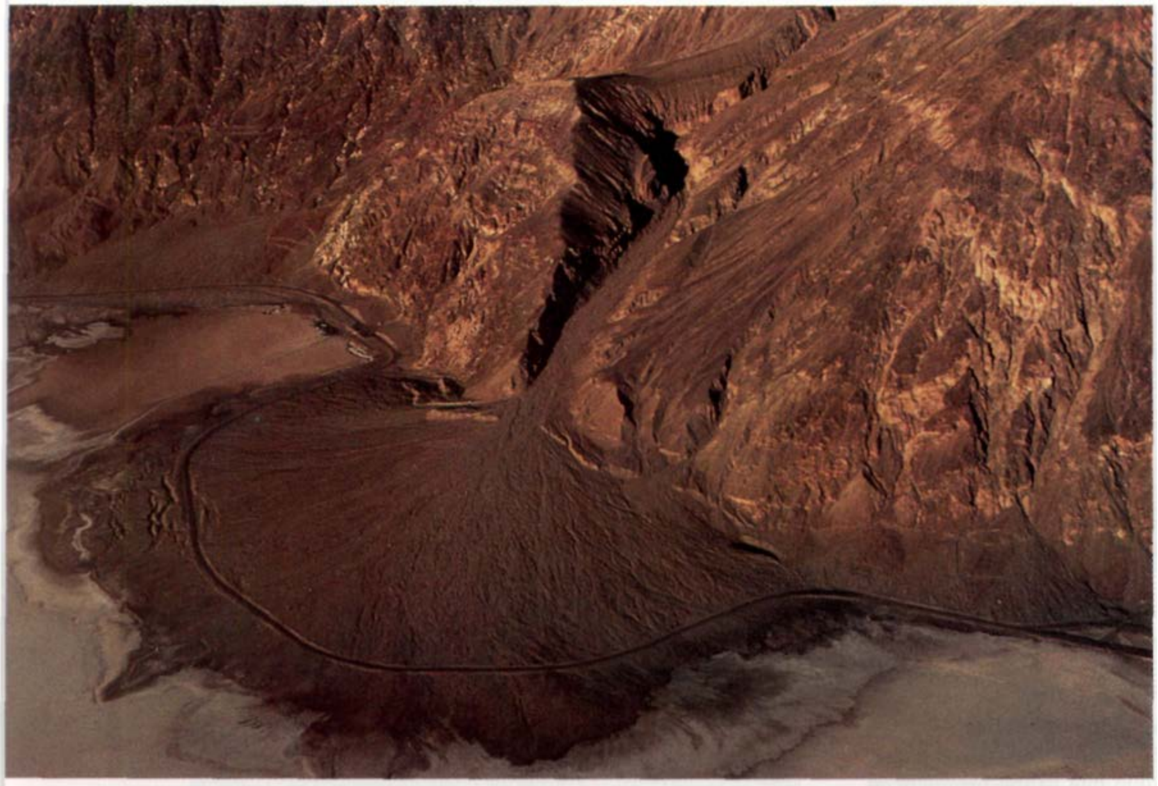


Figure 11.14 A symmetrical alluvial fan has formed at the margin of Death Valley, California, where a stream channel emerges from a steep mountain canyon.

Figure 11.15 Delta of Nile River, along the Mediterranean coast of Egypt. The reddish color in this vertical satellite image denotes vegetation growing on the fertile delta sediments. The delta and Nile River are bounded by a desert landscape of bare rock and shifting sands.

energy, usually because of a change in gradient, velocity, or discharge. Its transporting power therefore decreases, and it deposits part of its sediment load. No longer constrained by valley walls, the stream can shift laterally back and forth across more gentle terrain. The resulting deposit, an **alluvial fan,** is a fan-shaped body of alluvium at the base of an upland (Fig. 11.14). Alluvial fans are common along the base of most arid and semi-arid mountain ranges. Some fans are so large and closely spaced that they merge to form a broad piedmont surface that slopes away from the base of the mountains.

When stream water enters the standing water of the sea or a lake, its speed drops rapidly, decreasing its ability to transport sediment. The water deposits its load in the form of a **delta,** so named because the deposit may develop a crudely triangular shape that resembles the Greek letter delta (A) (Fig. 11.15). Many of the world's largest streams, among them the Nile, the Ganges-Brahmaputra, the Huang He, the Amazon, and the Mississippi, have built massive deltas at their mouths. Each delta has its own peculiarities, determined by such factors as the stream's discharge, the character and volume of its sediment load, the shape of the adjacent bedrock coastline, the offshore topography, and the strength and direction of currents and waves.

The Work of Groundwater

Slowly moving groundwater has the capacity to perform a prodigious amount of geologic work. In regions underlain by rocks that are highly susceptible to chemical weathering, groundwater creates distinctive landscapes that are among the most unusual on our planet.

Dissolution of Carbonate Rocks

As soon as rainwater infiltrates the ground, it begins to react with minerals in regolith and bedrock and weathers them chemically. An important part of chemical weathering involves minerals and rock materials passing directly into solution, a process known as **dissolution.** Limestone, dolostone, and marble—the common carbonate rocks—are most readily attacked by dissolution. Although carbonate minerals are nearly insoluble in pure water, they are readily dissolved by rainwater charged with CO_2, which is a dilute solution of carbonic acid (Table 11.1, eq. 4). The result is impressive. When carbonate rocks weather, nearly all their volume can be dissolved away in slowly moving groundwater.

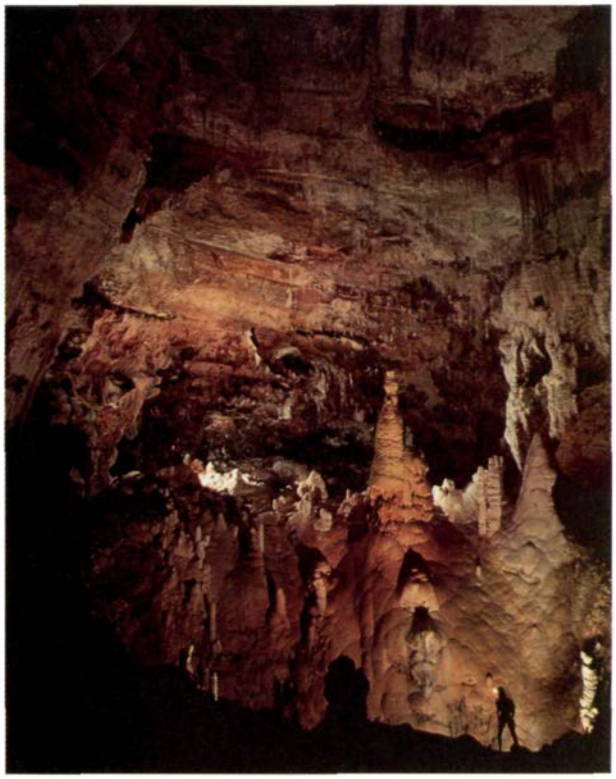


Figure 11.16 An explorer in Lechuguilla Cave, a limestone cave in the Carlsbad Caverns region of New Mexico, examines the bizarre formations produced as carbonate precipitated from dripping and flowing water during past millennia.

Caves and Sinkholes

Carbonate caves come in many sizes and shapes, and they often contain spectacular formations on their walls, ceilings, and floors (Fig. 11.16). Although most are small, some are of exceptional size. The Carlsbad Caverns in southeastern New Mexico include one chamber 1200 m long, 190 m wide, and 100 m (3940 ft, 625 ft, and 330 ft respectively) high. Mammoth Cave, in Kentucky, consists of interconnected chambers with an aggregate length of at least 48 km (30 mi).

Caves form as circulating groundwater slowly dissolves carbonate rock. The usual sequence of development is thought to involve (1) initial dissolution along a system of interconnected open joints and bedding planes by percolating groundwater, (2) enlargement of a cave passage along the most favorable flow route by water that fully occupies the opening, (3) deposition of carbonate formations on the cave walls while a stream occupies the cave floor, and (4) continued deposition of carbonate on the walls and floor of the cave after the stream has stopped flowing. Although geologists have argued for years as to whether caves form in the zone of aeration or in the saturated zone, available evidence favors the idea that most caves are excavated in the shallowest part of the saturated zone, along a seasonally fluctuating water table.

In contrast to a cave, a **sinkhole** is a large dissolution cavity that is open to the sky. Some sinkholes are caves whose roofs have collapsed (Fig. 11.17); others are formed at the surface. Those produced by cave collapse can form abruptly. As a result, they pose a potential hazard for people whose houses or property may suddenly disappear into a widening conical depression tens of meters across.

Karst Landscapes

In some regions of exceptionally soluble rocks, sinkholes and caves are so numerous that they combine to form a distinctive topography characterized by many small, closed basins and intervening ridges or pinnacles. In this kind of landscape, streams disappear into the ground and eventually reappear elsewhere as large springs. Such terrain is called **karst topogra-**

Figure 11.17 Most of a city block in Winter Park, Florida disappeared into a widening crater as this sinkhole formed in underlying carbonate bedrock.

phy after the Karst region of former Yugoslavia, where a remarkable landscape of closely spaced sinkholes has resulted from dissolution of the bedrock.

Several factors control the development of karst landscapes. The topography must produce a hydraulic gradient steep enough to permit the flow of groundwater through soluble rock under the pull of gravity. Precipitation must be adequate to supply the groundwater system, soil and plant cover must supply an adequate amount of carbon dioxide, and temperature must be high enough to promote dissolution. Although karst terrain is found throughout a wide range of latitudes and at various altitudes, it often is best developed in moist temperate to tropical regions underlain by thick and widespread soluble rocks.

One of the most famous and distinctive of the world's karst regions lies in southeastern China, where vertical-sided limestone peaks stand up to 200

Figure 11.18 Steep limestone pinnacles up to 200 m high, surrounded by flat expanses of alluvium, form a spectacular karst landscape around the Li River near Guilin, China.

m (660 ft) high (Fig. 11.18). This dramatic landscape has inspired both classical Chinese painters and present-day photographers.

Glaciated Landscapes

Skiers racing down the steep slopes at Alta, Mammoth, or Whistler and rock climbers inching their way up the cliffs of Yosemite Valley, the granite spires of Mont Blanc, or the icy monoliths of the southern Andes owe a debt to the ancient glaciers that carved these mountain playgrounds. The scenic splendor of these and most of the world's other high mountains is the direct result of glacial sculpturing. Over other vast areas of central North America and northern Europe, farmers gain their livelihood from productive soils developed on widespread glacial sediments left by former continental ice sheets. In all, fully 30 percent of the Earth's land area has been shaped by glaciers in the recent past.

In shaping the land surface over which it moves, a glacier acts like a plow, a file, and a sled. As a plow, it scrapes up weathered rock and soil and plucks out blocks of bedrock; as a file, it rasps away firm rock; and as a sled, it carries away the sediment acquired by plowing and filing, along with rock debris that falls from adjacent slopes.

Unlike a stream, part of a glacier's coarse load can be carried at its sides and even on its surface. A glacier can carry very large rocks and can transport large and small pieces side by side without segregating them according to size and density into a bed load and a suspended load. Thus, sediments deposited directly by a glacier are neither sorted nor stratified.

The load of a glacier typically is concentrated at its base and sides because these are the areas where glacier and bedrock are in contact. The coarse fraction of the load is derived partly from fragments of rock plucked from the lee side of outcrops over which the ice flows. Generally, such fragments are bounded by joints along which the rock has fractured. A significant component of the basal load of a glacier consists of very fine sand and silt grains informally called *rock flour*. If we examine such particles under a microscope, we find that they have sharp, angular surfaces that are produced by crushing and grinding.

Small rock fragments embedded in the basal ice scrape away at the underlying bedrock and produce long, nearly parallel scratches called **striations** (Fig. 11.19). Larger rock fragments that the ice drags across a bedrock surface abrade *glacial grooves,* aligned in the direction of ice flow. Rock flour in the basal ice acts like sandpaper and can polish the rock until it has a smooth, reflective surface.

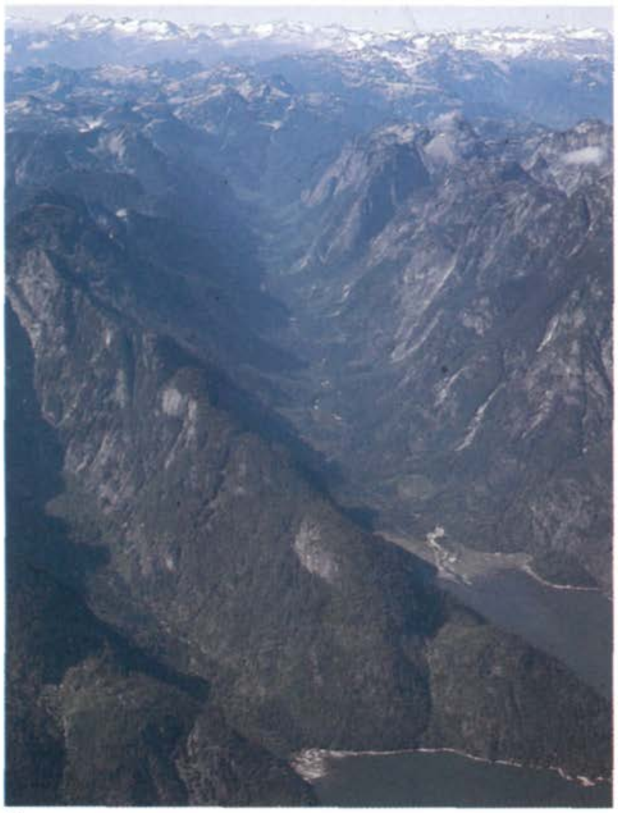
Figure 11.19 A deglaciated bedrock surface beyond Findelen Glacier in the Swiss Alps displays grooves and striations etched by rocky debris in the base of the moving glacier when it overlay this site. In the background rises the Mattcrhorn, a glacial horn sculpted by glaciers that surround its flanks.

The bulk of the rock debris visible on the surface of valley glaciers arrived there by rockfalls from adjacent cliffs. If a rockfall reaches the accumulation area, the flow paths of the ice (Fig. 10.12) will carry the debris downward into the glacier and then upward to the surface in the ablation area. If rocks fall onto the ablation area, the debris will remain at the surface and be carried along by the moving ice. Where two glaciers join, rocky debris at their margins merges to form a medial moraine (Figs. 10.5B and 10.12).

Glacial Sculpture

In mountainous regions, cirques are among the most common and distinctive landforms produced by

A.



B.

Figure 11.20 Typical landforms of glaciated mountains. A. This cirque, carved in sedimentary rocks of the Brooks Range, in northern Alaska, was the site of a former glacier that built the large bouldry end moraine on the floor of the cirque. B. A deep U-shaped valley in the southern Coast Range of British Columbia, Canada, was carved during repeated invasions of ice-age glaciers that left the valley walls smoothed and abraded to a height of nearly 2 km above the valley floor.

glacial erosion (Fig. 11.20A). The characteristic bowl-like shape of a cirque is the result of frost-wedging, combined with plucking and abrasion at the glacier bed. As cirques on opposite sides of a mountain grow larger, they intersect to produce sharp-crested ridges. Where three or more cirques have sculptured a mountain mass, the result can be a high, sharp-pointed peak, a classic example of which is the Matterhorn in the Swiss/Italian Alps (Fig. 11.19).

A valley that has been shaped by glaciers differs from ordinary stream valleys in having a distinctive U-shaped cross profile and a floor that often lies below the floors of tributary valleys (Fig. 11.20B). Streams commonly descend as waterfalls, or cascades, as they flow from the tributary valleys into the main valley; the glaciated Cascade Range of western United States derives its name from such streams. The long profile of a glaciated valley floor may possess steplike irregularities and shallow basins related to the spacing of joints in the rock, which influences the ease of glacial plucking, or to changes in rock type along the valley. Finally, the valley typically heads in a cirque or group of cirques.

Fjords deeply indent the mountainous, west-facing coasts of Norway, Alaska, British Columbia, Chile, and New Zealand (Fig. 11.21). Typically shallow at their seaward end, fjords become deeper inland, implying deep glacial erosion. Sognefjord in Norway, for example, reaches a depth of 1300 m (4260 ft), yet near its seaward end the water depth is only about 150 m (495 ft).

Glacial erosion is also responsible for countless lakes that lie inside the limit of the last glaciation. Among the largest are the huge lakes that form an arc across southern and western Canada and include the Great Lakes, Lake Winnipeg, Lake Athabaska, Great Slave Lake, and Great Bear Lake.

Eroding ice sheets sometimes mold smooth, nearly parallel ridges of till or bedrock, called *drumlins,* which are elongated parallel to the direction of ice flow (Fig. 11.22). Drumlins, like the streamlined bodies of supersonic airplanes that are designed to reduce air resistance, offer minimum resistance to glacier ice flowing over and around them.

Glacial Deposits

A moving glacier carries with it rock debris eroded from the land over which it is passing or dropped on the glacier surface from adjacent cliffs. As the debris is transported past the equilibrium line and ablation reduces ice thickness, the debris begins to be deposited. Some of the basal debris is plastered directly onto the ground as till (Chapter 7). Some also reaches the glacier margin, where it is released by the melting

Figure 11.21 Trekkers atop the Pulpit, a spectacular vantage point far above Lysefjord, can look far inland toward the source region of the glacier that carved this fjord, typical of numerous others that indent the rocky western coast of Norway.

Figure 11.22 A field of drumlins in Dodge County, Wisconsin, each shaped like the inverted hull of a ship, are aligned parallel to the flow direction of the continental ice sheet that shaped them during the last glaciation.

Figure 11.23 Lobuche Glacier, which flows out of a high cirque near Mount Everest in the Himalaya, has retreated upslope from a terminal moraine it deposited on the margin of Khumbu valley during the nineteenth century.

ice and either accumulates there or is reworked by meltwater that transports it beyond the terminus where it is deposited as *outwash.*

A ridgelike accumulation of sediment built up along the margin of a glacier is an **end moraine** (Fig. 11.23). An end moraine built at the terminus of a glacier is a *terminal moraine,* and one constructed along the side of a mountain glacier is a *lateral moraine.* End moraines form as sediment is bulldozed by a glacier advancing across the land, as loose surface debris on a glacier slides off and piles up along the glacier margin, or as debris melts out of ice and accumulates along the edge of a glacier. They range in height from a few meters to hundreds of meters. The great thickness of some end moraines results from the repeated accretion of sediment from debris-covered glaciers during successive ice advances.

When rapid melting greatly reduces a glacier's thickness in its ablation area, ice flow may virtually cease. Sediment deposited by meltwater streams flowing over or beside such immobile ice will slump and collapse as the supporting ice slowly melts away, leaving a hilly, often chaotic surface topography. Among the landforms associated with such terrain are *kames,* small hills of stratified sediment, and *kettles,* closed basins created by the melting away of a mass of underlying glacier ice. Landscapes marked by numerous kettles and kames are clear evidence of stagnant-ice conditions (Fig. 11.24).

Figure 11.24 Lake-filled kettles are scattered over the surface of an end-moraine complex in the lake district of central Chile that formed at the end of the last glaciation when debris-covered stagnant ice slowly melted away.

The Work of the Wind

Wind is an important agent of erosion wherever it is strong and persistent and wherever either the land is too dry to support vegetation or the influx of airborne sediment is so rapid that vegetation cannot gain a foothold and stabilize the ground surface. Because the density of air at sea level (1.22 kg/m^3) is far less than that of water (1000 kg/m^3), air cannot move as large a particle as water can flowing at the same velocity. In extraordinary wind storms, when wind speeds locally reach or exceed 300 km/h (190 mi/h), coarse rock particles up to several cm in diameter can be lifted to heights of a meter or more. In most regions, however, wind speed rarely exceeds 50 km/h (30 mi/h). At this velocity, the largest particles of sediment that can be suspended in air are grains of sand. At lower wind speeds, sand moves along close to the ground surface, and only finer grains of dust move in suspension.

Wind Erosion

Wind erosion on a large scale takes place only where little or no vegetation exists and where loose rock particles are fine enough to be picked up by the wind. Areas of significant wind erosion are found mainly in deserts; nondesert sites include ocean beaches, the shores of large lakes, and the floodplains of large glacial streams. Of greatest economic importance, however, are bare plowed fields, which are especially susceptible to wind erosion during times of drought. In the dry 1930s soil loss due to wind erosion in parts of the western United States amounted to 1 m or more within only a few years. By contrast, the long-term rate of erosion for the region as a whole is only a few centimeters per thousand years.

Small saucer- or trough-shaped hollows and larger basins are among the most conspicvious evidence of wind erosion. Where sediments are particularly susceptible to erosion, basins 50 m (165 ft) or more deep can be excavated. However, once an eroding basin floor reaches the water table, the surface soil becomes moist, encouraging the growth of vegetation that inhibits further erosion.

Wind-Blown Sand

If a wind is strong enough, it can start a grain rolling along the ground where it may impact another grain and knock it into the air. As this second grain falls to the ground, it will impact other grains, some of which are thrown upward into the air. Within a very short time the air close to the ground may contain a very large number of sand grains. They are moving by saltation, the same process that operates on a sandy stream bed (Chapter 9). However, even in strong winds saltating sand grains seldom rise far off the ground, as shown by abrasion marks on utility poles and fence posts that are sandblasted up to a height of about a meter.

Where some minor surface irregularity or obstacle distorts the flow of air, wind-blown sand may pile up to form a hill or ridge called a **dune** (Fig. 11.25). Dunes assume distinctive shapes that depend on wind strength and direction, and on local moisture and vegetation conditions. Five different dune types are shown in Table 11.3-

Many dunes grow to heights of 30 to 100 m (100 to 330 ft), and some massive desert dunes in western China reach heights of 500 m (1640 ft) or more. The height to which any dune can grow probably is determined by the maximum wind velocity, which increases above the land surface. At some height, the wind will reach a velocity great enough to carry the sand grains up into suspension off the top of a dune as

Figure 11.25 Coalescing barchan dunes migrate slowly across a vast plain in the Namibian Desert of southwest Africa. The steep slopes of the dunes descend in the direction toward which the prevailing wind blows and show that the dunes are moving in the direction of the photographer.

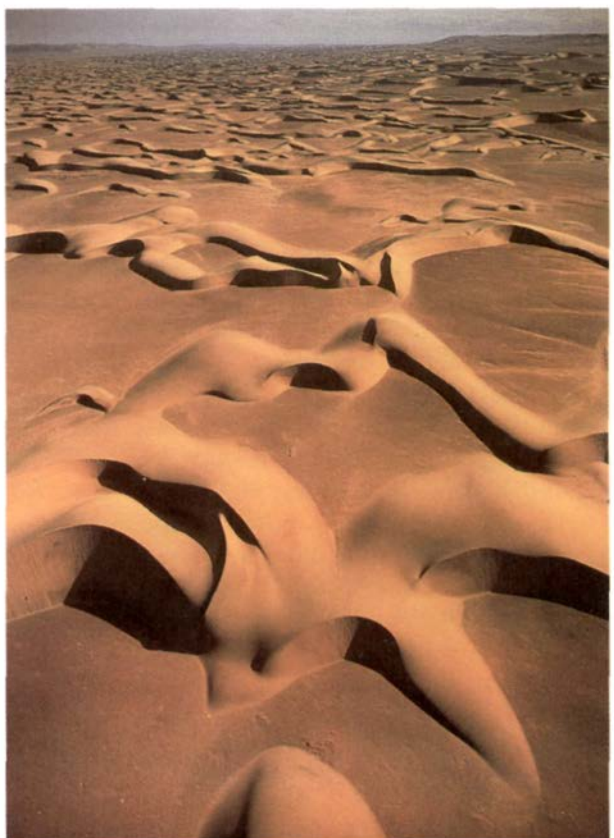
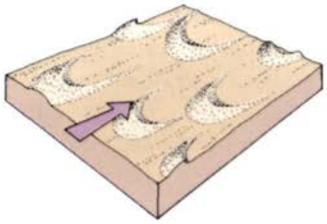
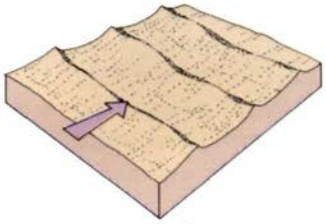
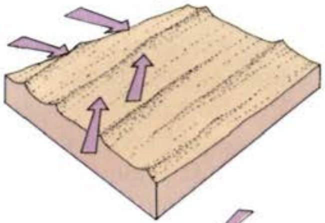
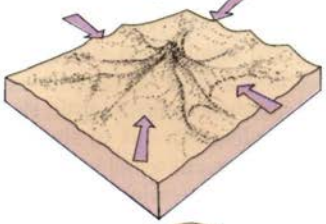
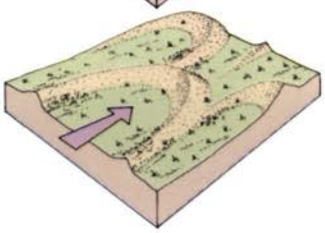


Table 11.3
Major Dune Types

	Dune Type	Description and Occurrence
	Barchan dune	Crescent-shaped dune formed where sand is limited, lack of moisture inhibits growth of vegetation, and strong winds blow from one direction. Tails of dune point downwind. These dunes can migrate over great distances without much change in form.
	Transverse dune	Where sand supply is large, individual barchans may merge to form a **transverse dune** having a sinuous crest oriented perpendicular to the strongest wind direction.
	Linear dune	Long, relatively straight dune that tends to occur in rather regularly spaced groups, mainly in areas of limited sand supply and variable (often bidirectional) winds.
	Star dune	A pyramidal dune with sinuous radiating arms formed by wind blowing from all directions. Although averaging 50 to 150 m high, exceptional star dunes more than 300 m high are known.
	Parabolic dune	A dune shaped like a U or a V, with two trailing arms, generally vegetated, pointing upwind. These dunes are common in coastal dune fields where relatively constant wind off the ocean creates a moist environment that allows vegetation to grow. Parabolic dunes almost always develop where local disturbance of vegetation permits sand to accumulate and may develop multiple crests and slip faces.

fast as they move up the windward slope by saltation. Thus, the dune can grow higher only as long as the rate at which sand is supplied to the crest exceeds the rate of removal by the wind.

A typical isolated sand dune has a gentle windward slope (facing the wind) with an average maximum angle of about 12° and a steep lee face (away from the wind) that stands at the angle of 33-34° (Fig. 11.26). Sand grains move up the windward slope by saltation to reach the crest of the dune and then generally fall onto the lee face near its top. When the accumulating sand becomes unstable, it avalanches (slips) downward, spreading the grains down the lee face (also called the *slip face)* and producing distinctive cross bedding.

Transfer of sand from the windward to the lee side of an active dune causes the whole dune to migrate slowly downwind. Measurements of barchan dunes show rates of migration as great as 25 m/yr (82 ft/yr). The migration of dunes, particularly along sandy coasts and across desert oases, has been known to bury houses and farmer's fields (Fig. 19.4), fill in canals, and even threaten the existence of towns.

Some large deserts contain vast tracks of shifting sand known as **sand seas**, that contain a variety of dune forms, ranging from low mounds of sand to

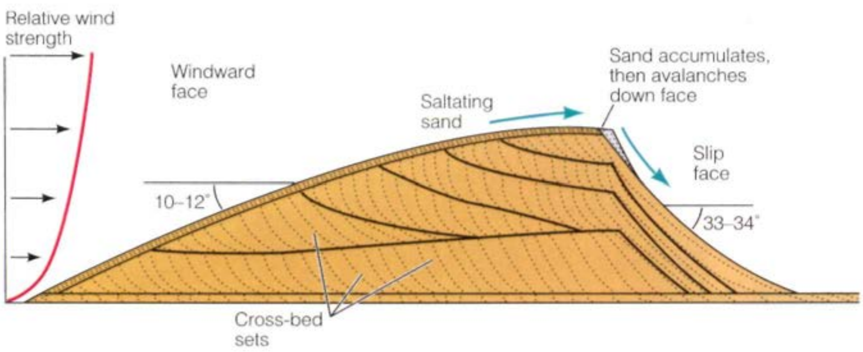


Figure 11.26 Cross section through a barchan dune showing the typical gentle windward slope and steep slip face. Sand grains saltate up the windward slope to the top of the slip face where they accumulate and then avalanche downward. Cross-bedded strata inside the dune represent old slip faces.

barchans, transverse dunes, and star dunes. In a typical sand sea, huge dune complexes form a seemingly endless and monotonous landscape.

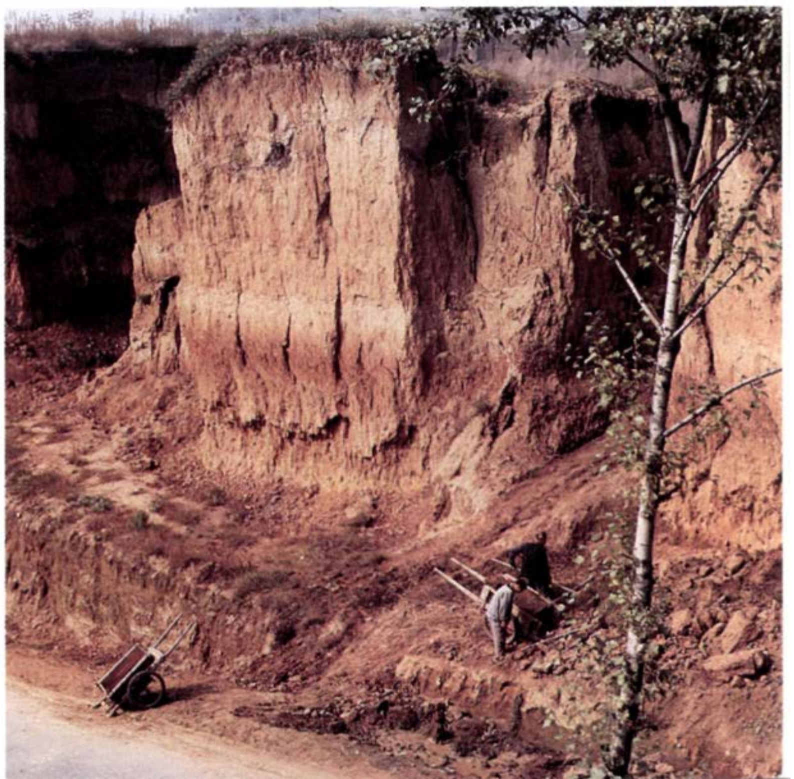
Wind-Blown Dust

Fine particles of dust (silt- and clay-size sediment) travel faster, longer, and much farther than sand grains before settling to the ground. As a result of frictional drag, the velocity of moving air decreases sharply near the ground surface. Right at the surface lies a layer of quiet air less than 1 mm (0.04 in) thick (See Figure 13.16 for an illustration of this phenomenon.) Sand grains that protrude above this layer of quiet air can be swept aloft by rising turbulent eddies. Dust particles, however, are so small and often so closely packed that they form a very smooth surface that does not protrude above the quiet air. Mobilization of the fine sediment may require the impact of bouncing sand grains or other physical disruption of the smooth surface. Once in the air, dust grains are continually tossed about by eddies, like particles in a stream of turbulent water, while gravity tends to pull them toward the ground. Meanwhile, the wind carries the dust forward. Although in most cases suspended dust is deposited fairly near its place of origin, strong winds associated with large dust storms are known to carry very fine dust into the upper atmosphere, where it can be transported thousands of kilometers.

Although most regolith contains a small proportion of wind-laid dust, the dust is thoroughly mixed with other fine sediments making it indistinguishable from them. However, in some regions wind-laid dust is so thick and uniform that it forms a distinctive deposit and controls the character of the landscape. Known as loess (German for *loose;* pronounced *lurs),* such deposits consist largely of silt. They have two characteristics that indicate deposition by wind rather than by streams, seawater, or lake water: loess forms a rather uniform blanket, mantling hills and valleys alike through a wide range of altitudes, and it contains fossils of land plants and air-breathing animals. Loess typically is homogeneous, lacks stratification, and, where exposed, stands at such a steep angle that it forms vertical cliffs, just as though it were firmly cemented rock (Fig. 11.27).

The main sources of wind-blown dust are deserts and the floodplains of glacial meltwater streams. Loess covering vast areas of central China is 100 to 400 m (330-1300 ft) thick and originated on the floors of large desert basins in central Asia that are surrounded by high glaciated mountains.

Figure 11.27 A steep cliff of loess rises above a road near Xi'an on the Loess Plateau of central China. The bulk of this loess accumulated during the last glacial age when cold, dry winds blowing across desert basins of central Asia swept up fine dust, transported it eastward, and deposited it like a thick blanket across the landscape. The reddish-brown band near the base of the exposure is an ancient soil that formed during an interval of moist, warm climate.



Where Land and Ocean Meet

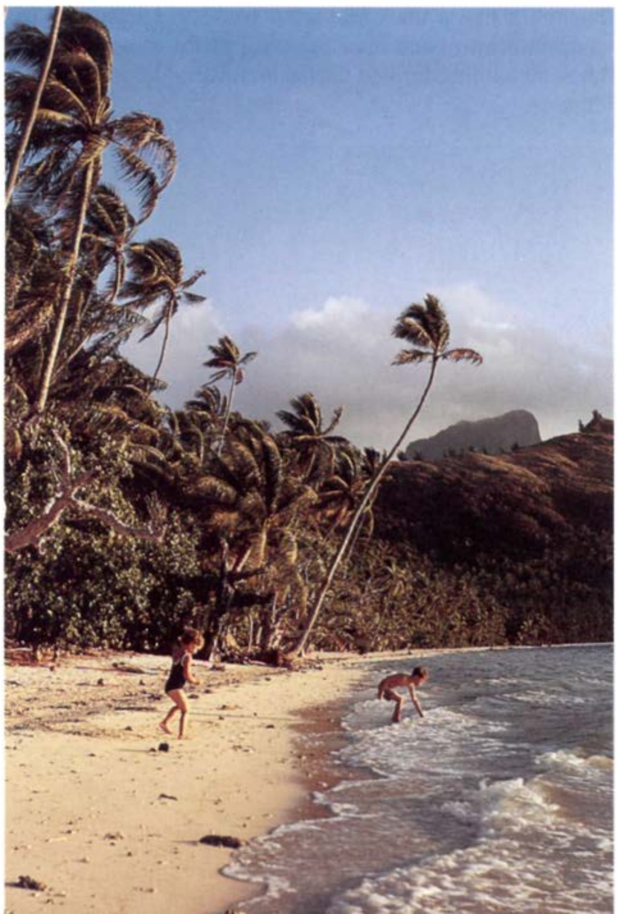
The oceans meet the land in a zone of dynamic activity marked by erosion and the creation, transport, and deposition of sediment. At a coast, waves that may have traveled unimpeded across hundreds or thousands of kilometers of open ocean encounter an obstruction to further progress. They dash against the shore, erode rock and sediment, and move the resulting particles about. Over time, the net effect is substantial: the form of a coast changes, often slowly, but at times very rapidly. At any given moment, the geometry of the shoreline represents a compromise among constructive and destructive forces.

The Impact of Waves and Surf

Ocean waves typically break at depths that range between wave height and 1.5 times wave height (Chapter 10). Because waves are seldom more than 6 m (17 ft) high, the depth of vigorous erosion by surf should be limited to 6 m times 1.5, or 9 m (30 ft) below sea level. This theoretical limit is confirmed by observation of breakwaters and other coastal structures, which are only rarely affected by surf to depths of more than 7 m (23 ft).

In the surf zone joint-bounded blocks of bedrock are plucked out and carried away. At the same time, continuous rubbing and grinding of moving rock particles in the turbulent surf wears away solid rock. In effect, the surf acts like a knife or saw, cutting horizontally into the land.

Figure 11.28 A sandy beach along the shore of Bora Bora, a volcanic island in French Polynesia, consists of coral and shell debris carried landward by wave action and mixed with lava fragments from the eroding volcano.



Beaches and Other Coastal Deposits

Beaches are a primary landform of most coasts (Fig. 11.28); even coasts dominated by steep, rocky cliffs generally have beaches interspersed with rocky headlands. A **beach** consists of wave-washed sediment along a coast, including sediment in the surf zone that is in constant motion. Although some beach sediment is derived from erosion of adjacent cliffs or older beach deposits, most sediment reaches beaches by rivers that enter the sea.

During storms, powerful surf erodes the exposed part of a beach and makes it narrower. In calm weather, the exposed beach is likely to receive more sediment than it loses and therefore becomes wider. Because storminess tends to be seasonal, beaches also change character seasonally.

Where surf and currents are inadequate to erode all new sediment carried to the sea by a large stream, the sediment builds outward as a marine delta (Fig. 11.15). The size and shape of a delta reflect the balance reached between sedimentation and erosion at the coast. Some deltas, such as that of the Mississippi River, consist of a complex of subdeltas of different ages, indicating a long and complicated history.

Elongate ridges of sand or gravel, called *spits,* that project from land and end in open water are another conspicuous coastal landform. Most spits are merely seaward continuations of beaches (Fig. 11.29A). Many spits are built of sediment moved by longshore currents and dropped at the mouth of a bay, where the current encounters deeper water and its velocity decreases.

Barrier islands, which are long, narrow sandy islands lying parallel to a coast and separated from it by a lagoon, are found along most of the world's lowland coasts (Fig. 11.29B). Some barrier islands, like those off the North Carolina coast, occasionally receive the full fury of destructive hurricanes, which erode and reshape these ephemeral landforms.

A. B.

Figure 11.29 Coastal landforms. A. The long, curved spit of Cape Cod, Massachusetts, has been built by longshore currents that rework glacial deposits forming the peninsula southeast of Cape Cod Bay. B. Barrier islands off Corpus Christi, Texas (along south side of large bay) seen from an orbiting satellite. To the right is the Gulf of Mexico. Padre Island National Seashore occupies the barrier island extending south from Corpus Christi Bay.

Reefs and Atolls

Not all coasts display the effects of erosion; some coasts are constructional. Many of the world's tropical coastlines consist of limestone reefs built by vast colonies of corals and other carbonate-secreting organisms. Three principal reef types are recognized: a *fringing reef* is either attached to or closely borders the adjacent land and lacks a lagoon (Fig. 11.30A); a *barrier reef* is a reef separated from the land by a lagoon and may be of considerable length and width (Fig. 11.30B and cover illustration); and an *atoll* is a roughly circular coral reef enclosing a shallow lagoon (Fig. 11.30C), formed when a tropical volcanic island with a fringing reef slowly subsides. Charles Darwin was the first to deduce, during his voyage on the H.M.S. *Beagle* in the 1830s, that slow subsidence forces reef organisms to grow upward so that they can survive near sea level. As an island subsides, the fringing reef is transformed into an offshore barrier reef and eventually, as the last remnants of the volcanic island disappear beneath the sea, into an atoll. Atolls generally lie in deep water in the open ocean and are as large as 130 km (80 mi) in diameter. Darwin's hypothesis was confirmed a century after he proposed it by drill holes on atolls that reached volcanic rock after penetrating thick sections of ancient reef rock.

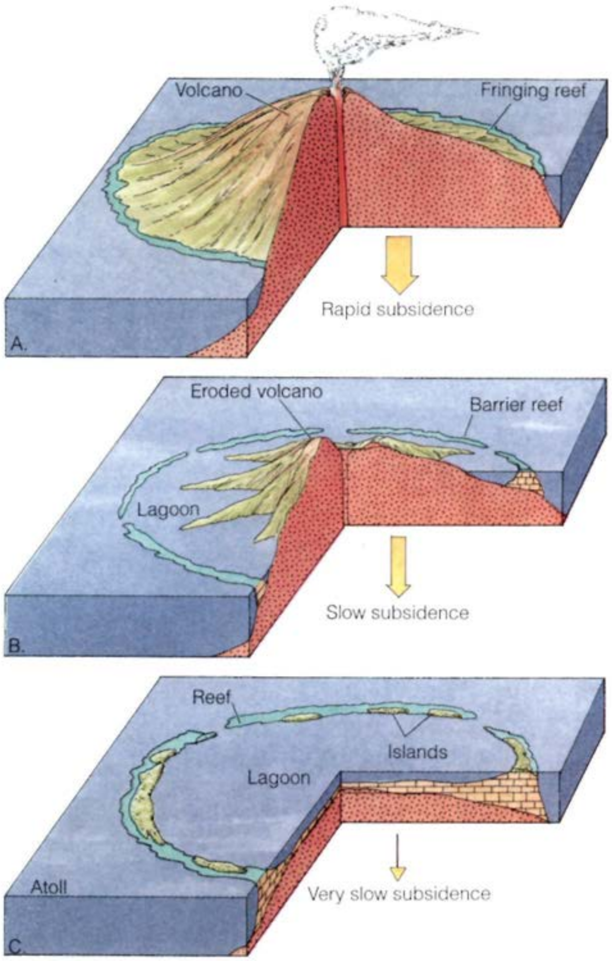


Figure 11.30 Evolution of an atoll from a subsiding oceanic volcano. A. Rapid extrusion of lava builds a shield volcano that begins to subside as the ocean crust is loaded by the growing volcanic pile; a fringing reef grows upward, keeping pace with subsidence. B. As subsidence continues, the fringing reef becomes a barrier reef, separated from the eroded volcano by a lagoon. C. With continuing subsidence and upward reef growth, the last remnants of volcanic rock are submerged, leaving an atoll reef surrounding a central lagoon.

A Variety of Coasts

The Pacific and Atlantic/Gulf coasts of North America represent two extremes of coastal landscapes. Each owes its general character to its structural setting. The rugged and mountainous Pacific coast lies along the margin of the American lithospheric plate, which is continuously being deformed where it interacts with adjacent plates to the west. Uplifted and faulted shorelines are common features along parts of this coast (for example, in southern Alaska, Oregon, and California). By contrast, the eastern continental margin lies within the same lithospheric plate as the adjacent ocean floor, but in a zone that is tectonically passive. The old bedrock has low relief, and much of the coastal zone borders young sedimentary deposits of the Atlantic and Gulf coastal plains. Sandy beaches and barrier islands are common from Long Island to Florida and beyond into northern Mexico.

The scenic coasts of northern New England, maritime Canada, southern Alaska, and British Columbia owe their special character to repeated glaciation. Each overriding ice sheet eroded bedrock in the coastal zone and depressed the land below sea level. With retreat of the ice, both the land and world sea level rose (although not always at the same rate). The result is embayed, rocky coastlines that show the effects both of differential glacial erosion and drowning of the land by the most recent rise in sea level (Fig. 11.21).

Relative Movements of Land and Sea

Nearly all coasts have experienced recent **submergence,** a rise of water level relative to the land that resulted when water from melting ice age glaciers returned to the oceans. Geologic evidence of lower, glacial-age sea levels is almost universally found seaward of the present coastlines and to depths of 100 m (330 ft) or more. Evidence of past higher sea levels, on the other hand, is related mainly to past interglacial ages when climates were warmer than now, glaciers were smaller, and sea level was therefore higher. The position of interglacial shoreline features above present sea level points to **emergence,** a lowering of the water level relative to the land.

The rise and fall of sea level are global movements, affecting all parts of the world's oceans at the same time. By contrast, uplift and subsidence of the land, which cause emergence or submergence along a coast, generally involve only parts of landmasses. Nevertheless, such movements can cause rapid relative changes in sea level. Vertical tectonic movements at the boundary of converging lithospheric plates have uplifted beaches and tropical reefs to positions far above sea level (Fig. 8.27A). Because movements of land and sea level may occur simultaneously, either in the same or opposite directions, unraveling the history of sea-level fluctuations along a coast can be a difficult and challenging exercise.

LANDSCAPE CHANGES THROUGH TIME

Mountain ranges and high plateaus are major landforms of the Earth's crust, and it is natural to ask: How long can such features persist? As soon as mountain

A Closer Look

Calculating Uplift and Denudation Rates

One way geologists attack the problem of calculating uplift rates is to measure how much local uplift occurs during historic large earthquakes, estimate the recurrence interval of such earthquakes, and then extrapolate the recent rates of uplift back in time. This approach assumes that the brief historic record is representative of longer intervals of geologic time, an assumption that may not be valid.

A second approach is to measure the warping, or vertical dislocation, of originally horizontal geologic surfaces of known age. Examples are flood basalts that have been deformed since extrusion and uplifted coral reefs along a tectonically active coast. In each case, we need both the difference in altitude between the present position of the surface and that at the time of formation, as well as a radiometric age for the rock forming the land surface.

Stream incision of a mountain range or plateau may produce a series of terraces that record uplift events. If the age of the terraces can be found, then uplift rates can be calculated. However, because terracing can also result from changes in stream activity caused by world sea-level fluctuations, changes of climate, or variations in the structure of rocks across which a stream flows, interpretation of river terraces is not always easy.

Finally, rocks that formed deep within the crust and subsequently were exposed at the surface by uplift and erosion can provide uplift rates. When the mineral zircon crystallizes in a plutonic rock, the subsequent decay of radioactive isotopes trapped in the mineral damages the internal arrangement of atoms, leaving tiny tracks (called *fission tracks)* that can be detected under a microscope. At high temperatures, fission tracks can form but they quickly anneal and disappear, leaving no record of their former existence. Only when the cooling mineral falls below a certain critical temperature (its *closure temperature)* is the annealing rate so slow that fission tracks will be retained. The number of tracks forming in a mineral increases with time, and so track density can be used to date the time elapsed since a rock containing the mineral cooled below the closure temperature. In the example shown in Figure C11.1, we assume that the closure temperature for the mineral zircon is 240°C (464°F). By measuring the geothermal gradient (we'll assume in the example it is 40 C7km), we then calculate that the 240°C isotherm lies at a depth of 6000 m (3.7 mi). The average rate of uplift can then be calculated because we know the fission track age of the rock and the total uplift since the rock first acquired fission tracks (i.e., the depth beneath the sample site at which the 240°C isotherm lies).

The calculation of long-term denudation rates re-

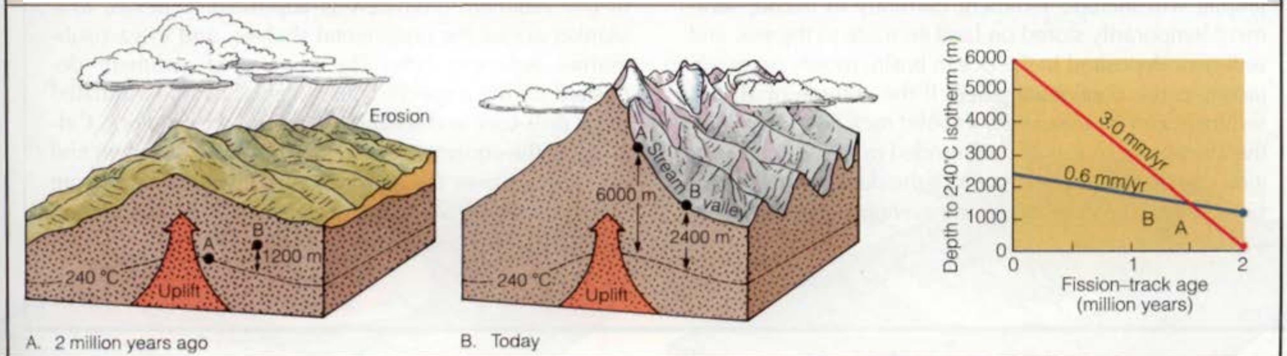


Figure C11.1 Uplift rate across a mountain range calculated using fission-track ages. A. Two million years ago, a zircon crystal (A) in a cooling pluton passes the closure isotherm of 240°C and begins to acquire fission tracks. Another crystal (B) began acquiring tracks 2 million years earlier and since then has been uplifted 1200 m above the 240°C isotherm. B. Rock samples containing zircon crystals A and B collected from a stream valley eroded into the rising mountain range have fission track ages of 2 and 4 million years, respectively, and lie 6000 and 2400 m, respectively, above the closure isotherm of 240°C. C. By using these data, the average uplift of samples A and B are calculated as 2.5 and 0.6 mm/year, respectively.

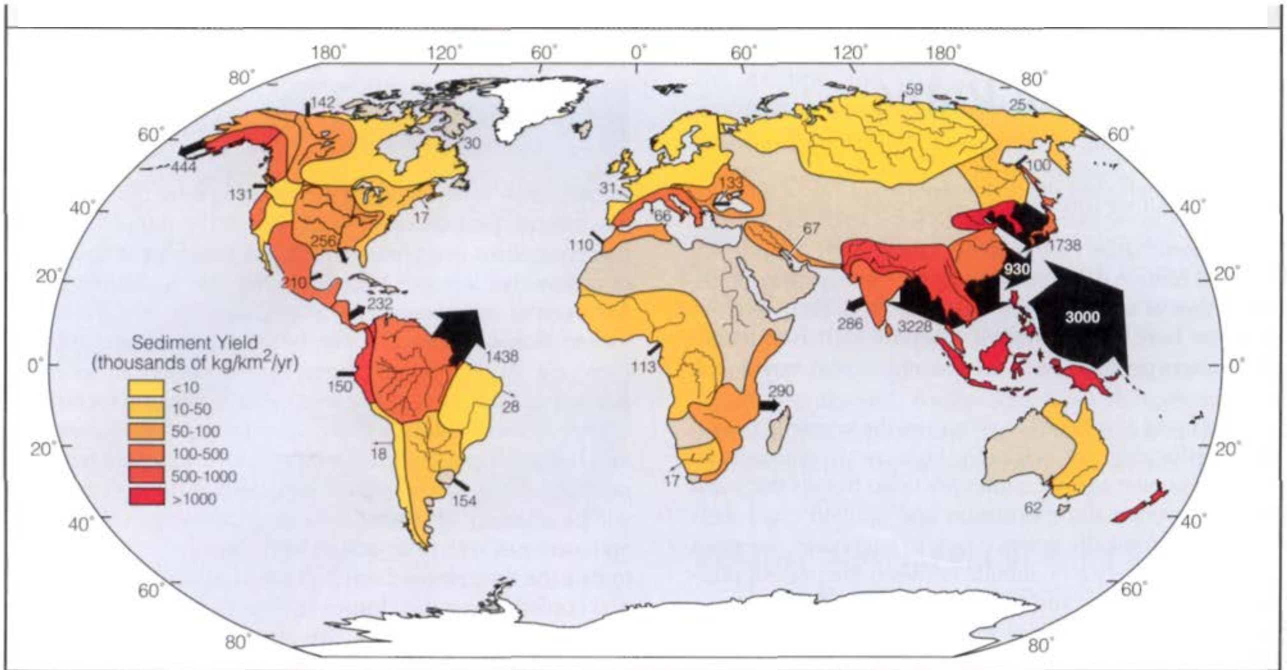


Figure C11.2 Average yearly discharge by streams of suspended sedimenlto the oceans. The width of each arrow is related to the amount of sediment entering the sea from a drainage basin (numbers on the map, in billions of kilograms). Colored land areas show the average amount of sediment contributed per square kilometer of each drainage basin. The areas colored pale green contribute minimal sediment to the oceans.

quires knowledge of how much rock debris has been removed from an area over a specified length of time. The total sediment removed from a mountain range, for example, will include sediment currently in transit, sediment temporarily stored on land en route to the sea, and sediment deposited in the ocean basin, mostly on or adjacent to the continental shelf. If the volume of all this sediment can be measured, its solid rock equivalent and the average thickness of rock eroded from the source region can be calculated. Finally, if the duration of the erosional interval can be dated, the average denudation rate can be determined.

For areas drained by major through-flowing streams, the volume of sediment reaching the ocean each year is a measure of the modern erosion rate (Fig. C11.2). Most of this sediment ultimately is deposited in deltas, as a blanket across the continental shelves, and in vast submarine sediment fans. The volume of sediment deposited during a specific time interval can be estimated using drill-core and seismic records of the sea floor. Calculating the equivalent rock volume of this sediment and averaging it over the area of the drainage basin(s) from which it was derived then gives an average denudation rate for the source region.

building commences, the forces of erosion begin to attack the rising land. The average altitude of the land at any time, therefore, should reflect this contest between uplift and erosion. If uplift is the winner, the land becomes higher, but if erosion wins out, the average altitude of the land surface declines. If uplift and erosion are balanced, then the gross form of the land may change relatively little with time.

While such deductions appear straightforward, obtaining reliable geologic information about long-term rates of landscape change is not so simple. First, we must be able to determine how rapidly a land mass is rising and how rates of uplift may have changed through time. Secondly, we must be able to calculate changing rates of **denudation,** which represents the combined destructive effects of weathering, mass-wasting, and erosion. Knowing uplift and denudation rates for a region, we can then infer something about how the landscape may have evolved. (See "A Closer Look: Calculating Uplift and Denudation Rates.")

Uplift versus Denudation

The history of uplift and erosion of the Himalaya and adjacent Tibetan Plateau in southern Asia shows how closely land, ocean, and atmospheric records are linked. Chinese geologists have suggested that these highlands did not exist as major topographic features before about 15 to 20 million years ago. They base their conclusion in part on evidence of plant fossils collected at altitudes of 4000 to 6000 m (13,000 to 20,000 ft), including many subtropical forms that exist today only at altitudes below 2000 m. Sediment cores from the northern Indian Ocean show low rates of sedimentation, implying reduced erosion rates and low-altitude source areas, until about 10 million years ago when sediment accumulation rates rose sharply. Two peaks in sediment supply (9 to 6 and 4 to 2 million years ago), seen both in the deep-sea cores and in alluvium on land, are interpreted as representing major intervals of uplift.

The current local uplift rate for one high region of the western Himalaya, based on fission-track measurements, is as high as 5 mm/year (0.2 in/yr). If sustained for only 2 million years, total uplift would be 10 km (33,000 ft). Such a high rate of uplift is generally associated with steep slopes and high relief. This is certainly true in the Himalaya where glacially eroded mountain slopes near the crest of the range typically exceed angles of 30° and local relief can exceed 5 km (16,500 ft). At lower altitudes; streams are the dominant erosional force, slopes generally decline to between 15 and 20°. Erosion rates in the high mountains

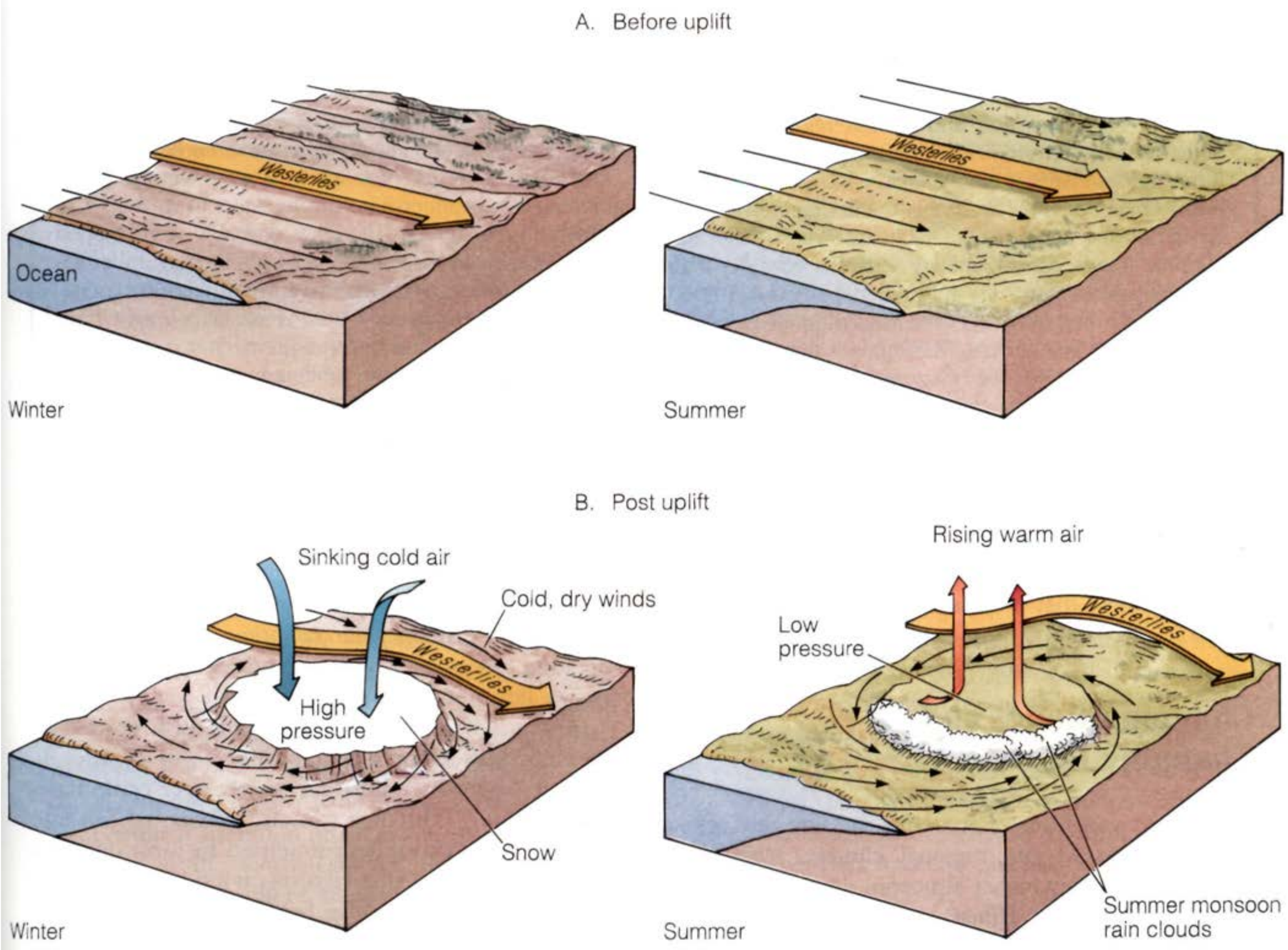


Figure 11.31 Development of a monsoon climate system. Before uplift, westerly winds pass directly across a landscape, the axis of flow shifting seasonally north and south. As a large plateau is uplifted, the westerly flow is diverted around the upland, a winter high-pressure system develops over the high, snow-covered plateau, and cold, dry winds blow clockwise off the upland region. In summer, as the plateau surface warms, the warm air rises, drawing moist marine air inland. The summer monsoon clouds move inland toward the plateau margin and lose their moisture as heavy rainfall.

Guest Essay

Erosion on Mars

Wes Ward is a geologist with the U.S. Geological Survey in Flagstaff, Arizona. He holds a Ph.D. in geological sciences from the University of Washington. Projects he has worked on include wind features on Mars and the deserts of Arizona and California, the geology of the San Juan River basin in New Mexico, and volcanoes in Arizona and Washington.

Terrestrial geomorphologists agree overwhelmingly that running water accomplishes most of the work of erosion on Earth. On Mars, the dominance of any one erosional agent is not so obvious. Surface water is not stable in the current Martian climate, polar and ground ice are restricted in latitude, and wind may have the energy and abrasive particles to attack only the most loosely consolidated of rock units. Yet geologists have found evidence that the same erosive agents operating on Earth (except for organisms) have been at work on the Red Planet.

Surface Water

Since the Mariner 9 mission in the early 1970s, we have recognized the presence of many huge channels that emanate from fractured and faulted regions of Mars called chaotic terrain. These channels are gigantic by terrestrial standards—up to 100 km wide, 2000 km long, and 3 km deep. They are similar to the Channeled Scablands of Washington State, which were caused by catastrophic floods. Most planetary scientists consider that, perhaps several times in the past, huge floods 10,000 times larger than the flow of the Mississippi River roared across the surface of Mars. These floods may have been caused by ground ice that melted during various episodes of volcanism, causing water to burst onto the surface.

Many investigators regard at least a dozen lowland areas on Mars as former lake and sea basins. Some of these were up to 2 million km^2 in area and 2 km deep. Evidence for these seas includes small drainage channels that lead from martian highlands to giant craters and lowlands, and terraces and shorelines in those craters and lowlands. The lakes are thought to have been present relatively recently (a few hundred million years ago). They, too, perhaps were created by the melting of subsurface ice, causing water to work its way slowly to the surface. The lakes probably were capped by thick ice covers, in order not to boil away under the thin atmosphere, so waves and currents were not as significant as they are on Earth. Some lakes might have eventually frozen and become permanent ice deposits, whereas others may have sublimed into the atmosphere and eventually formed polar ice deposits.

are estimated to be about 3 to 4 mm/year (0.1 to 0.2 in/yr), meaning that erosion and uplift are nearly balanced.

Changing Relief and Changing Climate

The height and form of the land can strongly influence both world and regional climate. Computer modeling has shown that atmospheric circulation in the absence of the Himalaya and Tibetan Plateau would be very different from what it is today (Fig. 11.31): the modern monsoon climate would not exist. The high mountains and plateau now divert the flow of westerly winds and are a primary factor in the present monsoon circulation. Summer heating of the high plateau causes warm air to rise, and this creates a low-pressure region that draws moist air inland from the adjacent ocean to produce the summer monsoon rainy season. In winter, the snow-covered surface of the plateau reflects solar radiation and cools down, creating a region of high pressure from which cold, dry air flows outward and downward.

About 2.5 million years ago, fine dust, eroded by strong winds from desert basins north of the Himalaya, began accumulating widely over central and eastern China. The dust was also carried far out into the Pacific Ocean where it settled to form deep-sea clay. The onset of this vigorous wind erosion apparently coincided with the beginning of the present monsoon climate of Asia as well as with the first widespread glaciation of the Asian highlands, both associated with strong regional uplift. Whereas the rising uplands initially were attacked by mass-wasting and stream erosion, eventually glacial and wind erosion were added to the list of processes that were tearing away at the land.

Ice

Martian continental glaciation may have occurred in plains regions where long and sinuous esker-like ridges intersect other ridges that resemble moraines. Alpine glaciation is indicated by large pits and hollows atop the highest martian volcanoes; these are similar to glacial features seen on Hawaiian volcanoes. Mountains forming the rims of the largest impact basins near the South Pole also seem to have cirque-like features sculpted into them. In addition, the north and south polar regions of Mars have thick sequences of light and dark layered sediments in a frost-covered, ridge-and-valley terrain. We think the banded layers consist of various mixtures of ice and dust. Our spacecraft observations indicate that the surface frost is water-ice (H_2O) at the North Pole but "dry" ice (CO_2) in the South.

Mars also has a unique type of crater, called a rampart crater because of its unusually high and steep-walled ejecta deposits. These are though, to indicate that ground ice was melted or vaporized during meteorite impact, producing a muddy slurry that eventually formed the ejecta ramparts. Other features that may have resulted from the former occurrence of ground ice are polygonal and patterned ground, and pingoes (ground ice hills).

Wind

Mars has light and dark surface-coloration patterns that change after the mighty planetwide dust storms that we see even with Earth-based telescopes, as well as vast dune fields, deflation hollows, and streamlined hills that we have seen from martian orbit, and even faceted boulders and drifts at the Viking landing sites. Yardangs (wind-carved hills that look like inverted boat hulls) up to 50 km long are found in several regions on Mars. We think that martian yardangs are carved into rocks that are relatively easily eroded, such as sandstones or volcanic ash deposits, as opposed to massive crystalline rocks, such as granite and basalt flows that could be difficult for wind alone to erode. Dunes found in many large martian canyons may represent sediment produced or delivered by water and mass wasting and then shaped by wind. One vexing problem, however, is that, at least near both Viking landings, we find no sand grains. The sediment that we can see seems to be fine dust that was deposited in the lees of rocks and boulders. This casts some doubt on the effectiveness of wind erosion on Mars, because of the predominance and effectiveness of sand erosion on Earth. We also do not understand how the giant dunes could have been formed. Such large features occur on Earth only with sand grains. Perhaps sand on Mars is concentrated in only a few regions, which we have not landed in yet, or perhaps dust particles can, under certain conditions, clot to form sand grains and build those large dunes. In areas where yardangs are found, perhaps the low mass of dust grains is compensated for in erosion by ultra-high-velocity winds. Regardless of how this dilemma is resolved, many scientists presently think that wind erosion and deposition are currently the most significant surface processes on Mars.

Mars, like Earth, is a fascinating place. We have been studying it closely for only a few years, and we have only theories to explain what goes on up there. Perhaps one of you readers will have a chance to test some of these ideas in the field.

Summary

1. Natural processes are constantly changing the character of the landscape, but so, too, are human activities. Although generally operating at slow rates, natural processes can produce dramatic changes on geologic time scales.
2. Weathering involves the physical breakup and chemical alteration of rock, and it occurs in the zone at the Earth's surface where the lithosphere, hydrosphere, biosphere, and atmosphere interact.
3. Physical and chemical weathering processes generally operate together, and their effects are inseparably mixed.
4. Soil is weathered regolith capable of supporting plants. A soil profile consists of successive horizons that develop from the surface downward.
5. Differences in soils result from variations in climate, vegetation cover, soil organisms, composition of the regolith, topography, and the length of time during which a profile has developed.
6. Soil erosion and degradation have become "epidemic," largely as a result of human activities. Because soils are essentially a nonrenewable resource, they must be carefully utilized and preserved.
7. Mass-wasting causes rock debris to move downslope under the pull of gravity without a transporting medium. The composition and texture of debris, the amount of entrapped air and water, and the steepness of the terrain influence the character and velocity of different types of slope movements.

8. The Earth's surface topography reflects sculpture by different erosional processes and deposition of resulting sediment. Water, ice, and wind are the principal agents that erode the land and transfer sediment toward the ocean basins.
9. A close relationship exists between rock structure—resulting from the movement of crustal plates—and the location of major drainage basins and continental divides.
10. The character of seacoasts is strongly influenced by fluctuations in world sea level that result in the submergence and emergence of coastal lands on glacial-interglacial time scales.
11. Landscapes evolve through time as tectonic forces raise crustal rocks and erosional agents wear them away. Rates of denudation in some mountain areas appear to be nearly equal to rates of uplift.

Important Terms to Remember

alluvial fan (p. 290)
beach (p. 300)
chemical weathering (p. 279)
delta (p. 290)
denudation (p. 304)
dissolution (p. 291)
dune (p. 297)
emergence (p. 302)
end moraine (p. 296)
floodplain (p. 289)
karst topography (p. 291)
landslide (p. 285)
loess (p. 299)
mass-wasting (p. 284)
physical weathering (p. 280)
relief (topographic) (p. 278)
sand seas (p. 298)
sinkhole (p. 291)
soil (p. 281)
soil horizon (p. 282)
soil profile (p. 282)
striations (p. 293)
submergence (p. 302)
terrace (p. 289)
topographic relief (p. 278)
weathering (p. 278)

Questions for Review

1. Why does the physical breakup of a rock increase the effectiveness of chemical weathering?
2. How and why does chemically weathered regolith developed on limestone differ from that developed on granite?
3. Explain why a soil profile formed in a semi-arid grassland differs from one developed in a wet tropical forest.
4. Why is the top of a buried soil an unconformity?
5. In what way does mass-wasting differ from stream erosion?
6. Name three geologic factors that make high mountain regions especially prone to landslides.
7. Make a list of the distinctive erosional and depositional features that would enable you to differentiate a landscape primarily shaped by streams from one shaped by glaciers.
8. Why is karst terrain not well developed in high mountains or desert landscapes underlain by carbonate rocks?
9. What internal sedimentary evidence shows that sand dunes migrate across the land?
10. What physical and biological characteristics would you look for in trying to determine whether a widespread body of silt mantling a landscape is lake sediment or loess?
11. What distinctive features could tell you whether a stretch of coast had experienced recent submergence or emergence?

Questions for A Closer Look

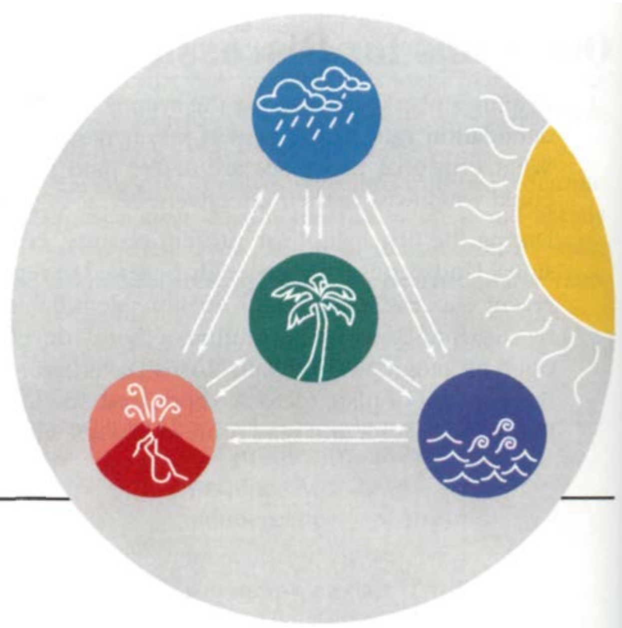
1. How might one calculate the volume of sediment removed from the land surface during the last 2 million years by a major stream that has built a large delta where it enters the ocean?
2 In what ways can the interpretation of river terrances pose problems in calculating uplift rates?

Questions for Discussion

1. Outline a plan for measuring the average present denudation rate in the region where you live. What erosional processes are active, and how could you measure their effectiveness?
2. During the first half of the present century, erosional landscapes were generally believed to represent successive cycles of erosion identifiable by stages of landscape evolution, a theory developed by geographer William Morris Davis before the concept of plate tectonics appeared. In your library, search for and read some of Davis's work and discuss whether his theory is consistent with our present understanding of plate tectonics.
3. Find a cemetery at least 100 years old and carefully examine the surfaces of gravestones in it for evidence of weathering. Classify the gravestones according to rock type and date of emplacement. Which rock types weather most rapidly and least rapidly? What factors likely control the degree of weathering that you can see?

PART FOUR

The Earth's Gaseous Envelope



Evolution of the Atmosphere

Two obvious differences separate the Earth from the other planets—the composition of its atmosphere and the presence of a biosphere. Other planets have atmospheres, but none has an atmosphere that nurtures a biosphere.

The Earth has not always had a biosphere-friendly atmosphere. Today's atmosphere has developed through the geological ages in response to slow changes in other parts of the Earth system. The primordial atmosphere consisted of the gases contained in the solar nebula from which the Earth formed. That atmosphere is no longer present. Astronomers hypothesize that the remains of the solar nebula were either blown away by protons streaming out of the Sun or blasted away during a particularly violent meteorite bombardment. Whatever the mechanism, the Earth's first atmosphere is thought to have been lost about 4.5 billion years ago, almost as soon as the Earth had grown to its present size.

The beginnings of today's atmosphere were volcanic gases. Evidence of ancient volcanism can be found in the most ancient rocks on the Earth, all of which were originally *igneous* (many have been metamorphozed), as well as in the most ancient rocks (also igneous) on the other terrestrial planets. The evidence is compelling that volcanism started very early in the life of each terrestrial planet. Volcanoes spew forth huge quantities of gas as well as lava, and then, as now, the principal volcanic gases were water vapor (H_2O), nitrogen (N_2), ammonia (NH_3), methane (CH_4), and carbon dioxide (CO_2). The early volcano-derived atmosphere was devoid of oxygen (O_2).

Sometime between 4.0 and 4.5 billion years ago, when the surface of the Earth had cooled sufficiently, water vapor in the atmosphere started to condense and rain began. When it started raining, the hydrosphere was born. This ancient Earth must have been a hot, steamy, oppressive place. The water temperature must have been so high that the primordial ocean was close to boiling, and the CO_2 level in the atmosphere was probably at least 100 times today's level.

Despite the hostile environment, primitive microscopic life eventually appeared on the Earth, most probably somewhere in the ocean, since there it was sheltered from the hostile atmosphere. (The appearance of life marked the origin of the biosphere.) When, where, and how life started is still a matter of intense research. Microscopic fossils provide evi-

Photograph taken from the space shuttle Columbia high above the atmosphere. The atmosphere can be seen as the thin bluish colored layer above the curved horizon. The view is looking north across the eastern end of the Mediterranean Sea. The dark band in the lower left is the Nile valley; the larger dark strip starting in the lower right corner is the Gulf of Suez, the smaller dark strip is the Gulf of Aqaba.

dence that life existed in the sea at least 3.5 billion years ago, and there is additional, though somewhat ambiguous, chemical evidence that life was present as much as 3.8 billion years ago. Whenever the momentous life-event occurred, the biosphere slowly started to change the atmosphere in ways that made it friendlier for the biosphere to grow ever larger. Indeed, the atmosphere eventually became so friendly that, by several hundred million years ago, it was possible for life to leave the sea, stay continuously in contact with the atmosphere, and spread to the land.

The biosphere changed the atmosphere in two ways. First, through the process of photosynthesis (by which plants combine CO_2 and H_2O to form organic matter and O_2), the biosphere added oxygen to the atmosphere. Second, by removal of carbon from the atmosphere to form organic matter and limestone, the biosphere lowered the CO_2 content; as a result, the temperature declined.

Volcanoes still give off gases, and the biosphere still pumps oxygen into the atmosphere. The Earth system that came into being with the birth of the biosphere is still evolving, and the part of the Earth system that responds most sensitively to change is still the atmosphere. In Part IV of the text, therefore, we turn our attention to today's atmosphere and to two of its most important roles in the Earth system: the weather and the climate.

CHAPTER 12

Composition and Structure of the Atmosphere

A climber working near the limit of human capacity, on the southwest face of Mount Everest, the world's highest mountain. At the elevation of Mount Everest, the air pressure is so reduced a climber gets insufficient oxygen with each breath and has to supplement the air with bottled oxygen.

The Air We Breathe

One very important criterion must be met if a planet is to be habitable: the air must be breathable, and the most essential ingredient of a breathable atmosphere is oxygen.

Where oxygen is concerned, the human body must rely entirely on the atmosphere to supply its needs. Although the body has some capacity to adjust for changes in the amount of oxygen available, the range of adjustment is limited. A measure of the lower limit of this range is provided by people who live, work, or visit at high altitude. Miners in the Andes Mountains of Chile and Peru, for instance, work at elevations as great as 5300 m (17,400 ft); at this altitude, a lungful of air contains only 50 percent of the amount of oxygen contained in a lungful at sea level. Despite the small amount of oxygen available in each breath, miners who grow up in the mountains become acclimatized and lead active working lives.

The absolutely lowest level of oxygen intake a body can handle probably varies from person to person, however, and so what a Peruvian miner can do may not be possible for everyone. For at least limited times, visitors to high altitudes can handle levels of oxygen even lower than the 50 percent value quoted above. Mountain climbers in the Himalaya, for example, have camped and climbed for weeks on end from bases as high as 7000 m (23,000 ft), at which height a lungful of air contains only 44 percent of the oxygen in a sea level lungful. This amount of oxygen is probably very close to the lower human limit because balloonists who have attempted long-time flights at heights in excess of 7000 m have found it necessary to breathe bottled oxygen; some who attempted to fly without extra oxygen perished in the attempt.

By experiments on divers and in hospitals, the upper limit of oxygen has been found to be 55 percent above the amount found in sea-level air. Beyond this limit, oxygen becomes toxic because the body responds, in effect, by starting to burn up. For safety, the upper limit of oxygen used in hospitals for patients having breathing difficulties is a 40 percent increase above normal sea-level air.

Therefore, we can conclude that, to be habitable, a planet must have an oxygen level ranging from 40 percent above to 44 percent below the level found in today's air. As far as we can tell from the geological record, the oxygen content has varied, but it has not moved outside the habitable range for several hundred million years. This means that, if a time machine really were possible, we could turn the clock back and visit the dinosaurs and breathe their air quite comfortably. However, the fact that the oxygen content of air has varied may have had major consequences for the biosphere in ages past. For example, samples of air trapped in amber (a fossil tree resin) suggest that, 100 million years ago, during the Cretaceous Period, the oxygen content of the atmosphere was 40 percent higher than in today's atmosphere. Although the dinosaurs were sent to extinction at the end of the Cretaceous, probably by a great meteorite impact, their numbers had been declining for a long time before the impact. As the Cretaceous came to a close, the oxygen level started to decline. One hypothesis for the decline of the dinosaurs, which occurred at this time, is that they had small lungs because the air contained so much oxygen. (Research on fossil skeletons from dinosaurs of this period suggests not only that the lungs were small but that the nasal capacity was also small and incapable of supplying a large quantity of air to the lungs.) As the oxygen level dropped, their small lungs could not adjust, and so, like high-flying balloonists, they died out as a result of respiratory stress.

WEATHER AND CLIMATE

The weather is said to be the most popular topic of conversation for all cultures. If this is true, the weather deserves its high ranking because it plays such an important role in our daily lives—from the clothes we wear to the activities we pursue. Even though we talk a lot about the weather, however, we are sometimes a bit confused about what is actually being discussed. To avoid confusion, we follow the lead of *meteorologists* (weather scientists) who have a formal definition for **weather**; they define it as the state of the atmosphere at a given time and place. The five variables that meteorologists measure in order to determine the state of the atmosphere are:

1. Temperature
2. Air pressure
3. Humidity
4. Cloudiness
5. Wind speed and direction

These five weather variables are the climate variables, but there is an important difference between weather and climate. Weather is a short-term event, whereas climate is a long-term one. Weather can change over a short time span. For example, the weather can be cold and wet in the morning but warm and dry in the afternoon. Climate, on the other hand, can be measured only over a period of years because **climate** is the *average* weather condition of a place. The climate of northern Canada is cold and wet, for example, even though this area enjoys many warm, dry days. Over a period of years, the number of cold, wet days is far larger than the number of warm, dry days, however, and so the average weather (that is, the climate) of northern Canada is cold and wet regardless of what the weather may be on a given day or even during a given week or month. The opposite condition is found in the Sahara of northern Africa; there, hot, dry days are far more common than cold, wet ones, and so the climate of the Sahara is classified as hot and dry.

Weather and climate are especially sensitive indicators of changes in the Earth system. This is so because a rapid change to the atmosphere can quickly change weather; for example, a volcanic eruption can affect weather around the world in a matter of days. Because the atmosphere is such an important sensor in the Earth system, Chapters 12 and 13 are dedicated to the properties of the atmosphere and to how changes in these properties affect the weather. Chapter 12 focuses primarily on the composition, structure, and physical properties of the atmosphere, and then Chapter 13 addresses motions in the atmosphere. Note that these two chapters are concerned primarily with weather rather than climate. Because climate can be affected by long-term changes in the solid Earth, the hydrosphere, and the biosphere, as well as in the atmosphere, the controls on climate and climatic changes are addressed separately in Chapter 14, and potential future human influences on the climate are discussed in Chapter 18.

COMPOSITION AND STRUCTURE OF THE ATMOSPHERE

Two things energize the atmosphere: the Sun's heat and the Earth's rotation. As discussed in Chapter 2, energy from the Sun reaches the Earth in the form of electromagnetic radiation. Solar radiation is the energy source responsible for such things as clouds, rain, snowstorms, and much of the local weather. By contrast, the Earth's rotation is mainly responsible for large-scale effects in the atmosphere—effects such as the west-to-east movement of weather patterns, the jet stream, and global wind systems. Although the two major energy sources always operate in concert, they can nevertheless be separated for discussion, and that is the basis on which the topics in Chapters 12 and 13 are separated.

To appreciate how the two energizing sources play the roles they do, we first have to understand the structure of the atmosphere. It may seem strange to say "structure" when we discuss a gaseous layer, but measurements show that, with increasing altitude, there are distinct variations in such factors as the composition, temperature, pressure, and humidity of the atmosphere. We humans live at the bottom of the atmosphere, and so most of the variations are far above our heads. What is known about the upper reaches of the atmosphere, like our knowledge of the solid Earth beneath our feet, comes mainly from instrument probes of one kind or another.

Composition

There is a subtle difference between the words *atmosphere* and *air*. An *atmosphere* is the gaseous envelope that surrounds a planet or any other celestial body. **Air**, by contrast, is the invisible, odorless mixture of gases and suspended particles that surrounds one special planet, the Earth. In other words, air is the Earth's atmosphere; it is also our most precious commodity. Denied access to air, the human body dies within a few minutes.

Considering our total dependence on air, you might think that everyone would know and care a lot about the air we breathe. Unfortunately, that does not seem to be the case. All too often, people take this most precious commodity for granted and foolishly treat it as if it had an endless capacity to absorb pollutants.

Air is a complex mixture of gases and tiny suspended particles. Because air pressure decreases with altitude, the amount of air per unit volume (that is, the density) also varies with altitude. In order to separate changes caused by composition from those caused by density, the composition of air is always discussed in terms of the *relative* rather than the absolute amounts of the different constituents present.

The relative composition, it turns out, varies somewhat from place to place on the surface of the Earth and even from time to time in the same place. There are two reasons for the variations—the presence of aerosols and the presence of water vapor—both of which vary widely in amount:

1. **Aerosols** are tiny liquid droplets or tiny solid particles that are so small they remain suspended in the air (Fig. 12.1). Common liquid aerosols, such as the water droplets in fogs, are familiar to all of us. Solid aerosols such as tiny ice crystals, smoke particles from fires, sea-salt crystals from ocean spray, dust stirred up by winds, volcanic emissions, and pollutants from industrial activities are less familiar but nevertheless widespread. Aerosols are everywhere in the atmosphere, particularly in the air nearest the ground. We breathe them in all the time, but fortunately the human respiratory system is designed to prevent them from causing harm to our lungs.

2. Because the Earth has a hydrosphere from which water evaporates, there is always water vapor in the atmosphere. The amount of water vapor in the air, for which the term **humidity** is used, is quite variable. On a hot, humid day in the tropics, as much as 4 percent of the air by volume may be water vapor, whereas on a crisp, cold day, less than 0.3 percent water vapor may be present.

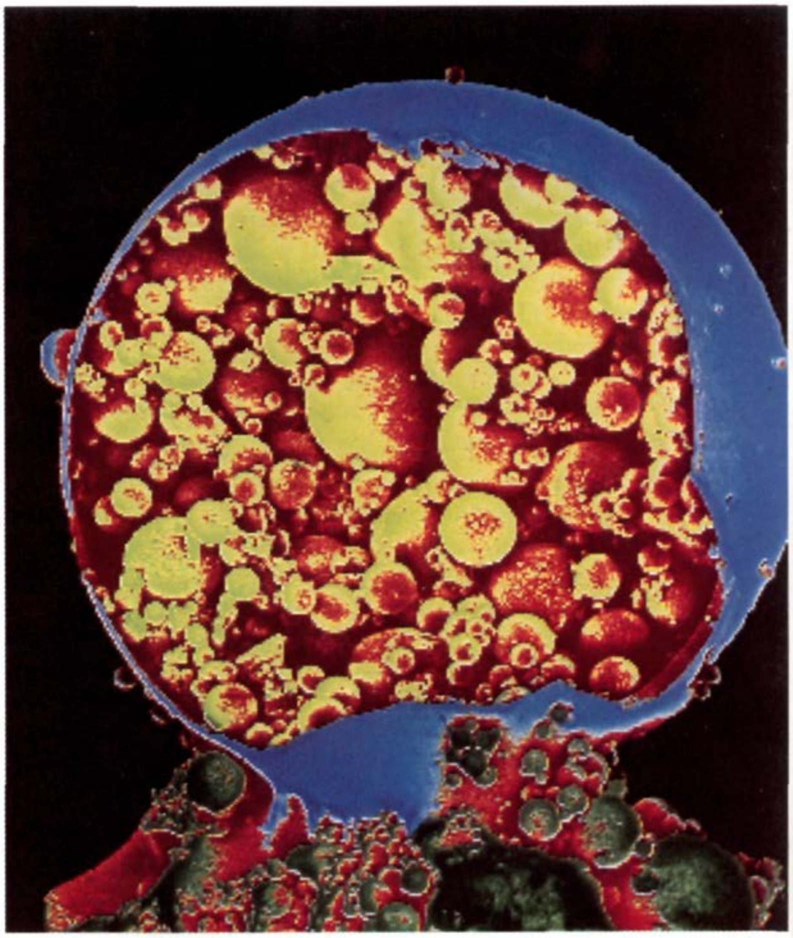


Figure 12.1 False-color electron micrograph image of fly ash, a common aerosol. Fly ash comes from power plants and other sources of comhustion. Individual spheres range in size from about 0.001 to 0.01 mm in diameter.

Because both the water vapor and aerosol contents of the air vary widely, the relative amounts of the remaining gases in the air are generally reported on a dry (meaning free of water vapor) and aerosol-free basis. Once these two variable constituents are removed, the relative proportions of the remaining gases in the air turn out to be essentially constant. As shown in Figure 12.2, three gases—nitrogen, oxygen, and argon—make up 99.96 percent of dry air by volume. Even though the relative amounts of the remaining gases are very small, these minor gases are profoundly important for life on the Earth because they act both as a warming blanket and as a shield from deadly ultraviolet radiation.

Carbon dioxide, methane, ozone, and nitrous oxide are the minor gases that, together with water vapor, create the Earth's life-maintaining blanket. These five gases are commonly called the *greenhouse gases* because, like a glass-covered greenhouse, they

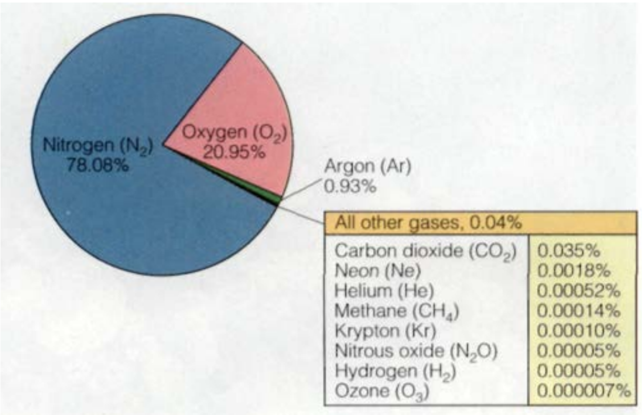


Figure 12.2 Composition of dry, aerosol-free air in volume percent. Three gases—nitrogen, oxygen, and argon—make up 99.96 percent of the air.

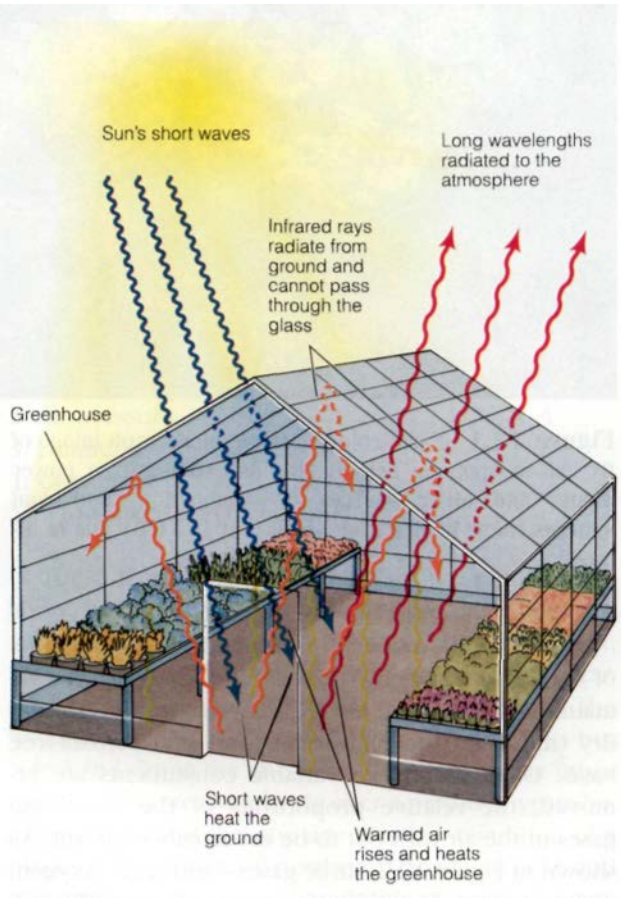


Figure 12.3 The way a greenhouse works. Short-wavelength radiation from the Sun passes through the glass roof and heats the ground. Some of the heat from the ground then warms the air in the greenhouse; the rest is reradiated back as infrared radiation, which is then trapped by the glass roof, producing additional heating inside. The warmed air emits long-wavelength radiation, which passes through the glass and escapes into the atmosphere. When a balance is reached, the incoming radiation equals the escaping radiation.

create a warm environment. A greenhouse is kept warm because air heated by the Sun's radiation is prevented from escaping (Fig. 12.3). The atmosphere works in a similar way. Like the glass of a greenhouse, the atmosphere lets the Sun's short-wavelength radiation pass through and warm the Earth's surface. However, the greenhouse gases then absorb the infrared radiation given off by the Earth's surface. This absorbed radiation heats the atmosphere and keeps the air in contact with the Earth's surface in a comfortable temperature range. (If you move ahead to Chapter 18, you will find that Figure 18.8 illustrates the radiation absorption effect.)

The part of the solar radiation spectrum that is dangerous to humans and to many other living creatures is the ultraviolet (see Fig. C2.2). Most of this deadly radiation is prevented from getting to the Earth's surface as a result of absorption by three forms of oxygen gas: O, O_2, and O_3 (ozone) (Fig. 12.4). The most important of the three shielding gases is ozone, which, even though it occurs only in minute amounts high in the atmosphere, is able to absorb the most lethal of the ultraviolet rays.

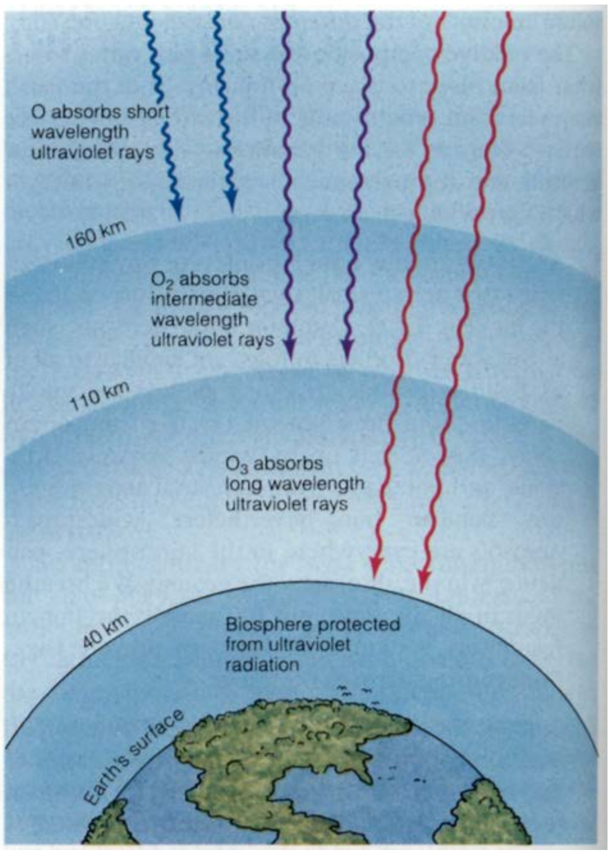


Figure 12.4 Life-protecting layers of 0, 0_2, and 0_3 in the atmosphere absorb lethal ultraviolet radiation.

At very great altitudes, 80 km (50 mi) and higher, the composition of dry, aerosol-free air changes a little from what it is at the Earth's surface; it is depleted in the heavier gases, such as neon and argon, and enriched in the lighter gases, such as helium. For most purposes of discussion of the Earth system, however, we don't have to be worried about such high-altitude changes in the composition because weather and climate effects occur in the lower atmosphere, where the relative compositions do not vary.

Temperature

Temperature is the most important variable used to define the state of the atmosphere; it is also the most familiar. The human body is sensitive to changes in temperature as small as 1°C (1.8°F), and as a result everyone is aware that temperature varies from hour to hour, from place to place, from day to night, and from season to season.

Temperature Versus Heat

Before we discuss temperature variations in the atmosphere, it is important to establish the difference between heat and temperature. The definition of heat and *heat energy* (the two terms mean exactly the same thing) is the total kinetic energy (energy of motion) of all the atoms in a substance. Not all the atoms in a given sample move with the same speed, however, and so there is a range of kinetic energies among them. **Temperature** is a measure of the *average* kinetic energy of all the atoms in a body. Note the difference: heat is the *total* energy, and temperature is a measure of the *average* energy. Even though two bodies of the same substance, such as a cup of water and a pail of water, have the same temperature, say 25°C (77°F), there are so many more water molecules in the pail than in the cup that the pail has far more heat energy. Furthermore, if the temperature of the cup of water is raised from 25°C (77°F) to boiling (100°C or 212°F), which means the average speed of the atoms in the water molecules rises, the heat energy in the lower-temperature but much larger pail would probably still exceed the heat energy in the small, higher-temperature cup. The reason is that the heat energy of the sum of a large number of slow-moving atoms may well exceed the heat energy of a small number of fast-moving atoms.

The atmosphere gets its heat energy from the Sun. As we discussed in Chapter 2, the flux of energy coming in from the Sun is 1370 W/m^2. This is the energy flux that would be measured by a satellite orbiting the Earth outside the atmosphere. The **insolation,** which is the energy that reaches the surface, is considerably less than 1370 W/m^2. (See "A Closer Look: Sunlight and the Atmosphere.")

There are two principal reasons why the insolation is less than 1370 W/m^2. The first reason is the atmosphere, which absorbs some of the incoming radiation. The second reason is the shape of the Earth. Even if the Earth were devoid of an atmosphere, there is only one place on it, as shown in Figure 2.5, that would receive 1370 W/m^2, and that is the spot where the Sun is directly overhead. At all other places, because of the Earth's curvature, 1370 W would be spread over an area larger than a square meter.

Temperature Profile of the Atmosphere

The way the temperature of the atmosphere changes with altitude is shown in Figure 12.5. Note that there are four thermal layers separated by boundaries called *pauses.* The layers are as follows.

1. The **troposphere** extends to variable altitudes of 10 to 16 km (6 to 10 mi) and is named from the

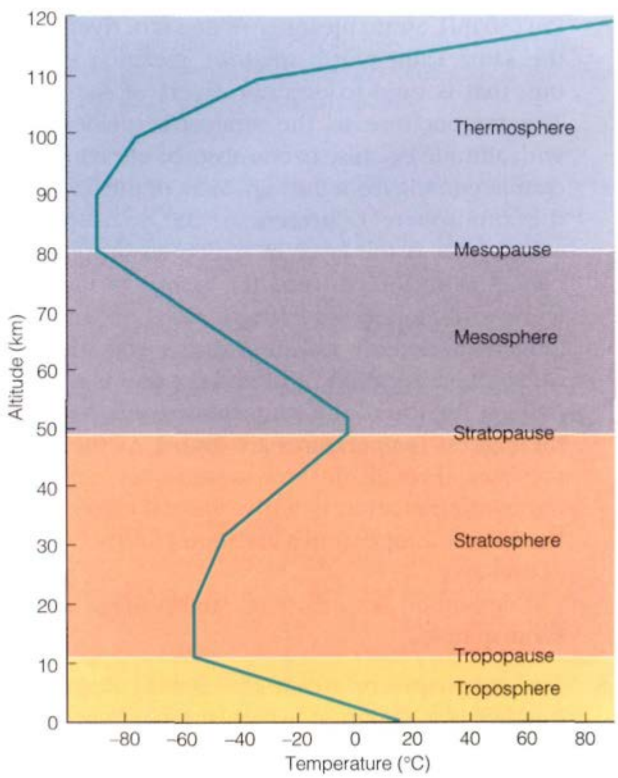


Figure 12.5 The variation of temperature with altitude in the atmosphere. The atmosphere is divided into four temperature zones. The outermost zone, the thermosphere, continues to an altitude of about 700 km.

Greek, *tropos,* meaning to change or to turn. Temperature in the troposphere decreases with altitude because air at the bottom is continually warmed by the ground and the ocean. The troposphere is so named because it is endlessly convecting, with warm ground-level air rising upward and colder air from above sinking downward to take its place. Most of our weather arises in the troposphere and is a consequence of these thermal motions.

The top of the troposphere, the **tropopause**, is 16 km (10 mi) high at the equator but only 10 km (6 mi) or less at the poles. The tropopause height does not change smoothly from the equator to the poles. Rather, it declines very gently as one moves from the equator to a latitude of about 40°, then it decreases sharply to about 10 km, and finally, it continues near that height to the poles (Fig. 12.6). As will be discussed in Chapter 13, the incline in the tropopause has important consequences for weather because it gives rise to the phenomenon known as *the jet stream.*

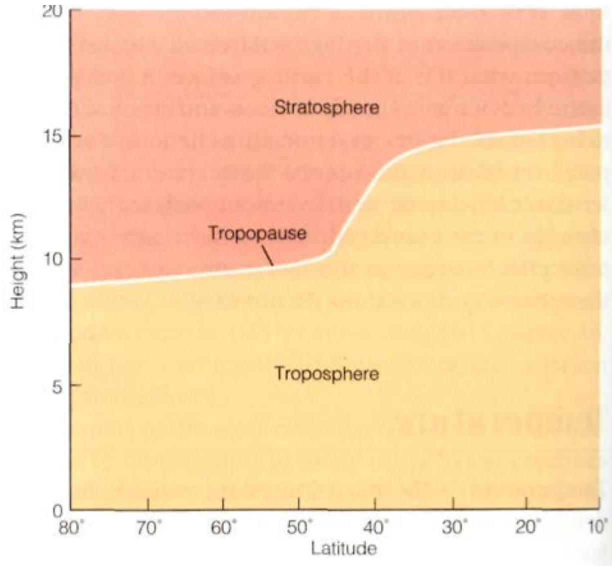


Figure 12.6 Schematic section through the tropopause showing the altitude variation with latitude. As discussed in Chapter 13, the sharp change in altitude in middle latitudes plays a major role in the polar Jet stream.

2. The **stratosphere** lies above the tropopause and is a region in which the temperature increases with altitude, reaching a maximum at about 50 km (30 mi). Strato means layer and is derived from the same Latin word, *stratum,* meaning spread out, that is used to describe layers of sediment. The temperature in the stratosphere increases with altitude because ozone absorbs ultraviolet radiation coming from the Sun. Most of the ozone in the atmosphere is present in the stratosphere, and so that is the layer in which absorption occurs. Absorption converts the energy of ultraviolet rays into longer-wavelength radiation, and this longer-wavelength radiation heats the air. Because the absorption of ultraviolet rays is a maximum at the top of the stratosphere, this is where the highest temperatures are found. As the Sun's rays pass through the stratosphere, less and less ultraviolet radiation is left to be absorbed; thus, the lowest temperatures are found at the bottom of the layer.

   The upper boundary of stratosphere is the **stratopause.**

3. The **mesosphere** (from the Greek *mesos,* for middle) is a region in which temperature again decreases with increasing altitude, reaching a minimum of about - 100°C (- 148°F) at about 85 km (53 mi). The mesosphere is terminated by the **mesopause.**

4. The **thermosphere,** which reaches out to about 500 km (310 mi), is a region of increasing temperature. The temperature increase in the thermosphere arises partly from the absorption of solar radiation by gases in the atmosphere and partly from bombardments of gas molecules by protons and electrons given off by the Sun. During periods of strong sunspot activity, when the flux of protons and electrons is at a maximum, the bombardment is so great that the temperature at the top of the thermosphere may reach as high as 1500°C (2732°F).

Despite the high temperatures reached in the thermosphere, very few molecules of gas are present and so there is very little heat. Strange as it may seem, we would feel very cold if we were exposed to a 1500°C atmosphere as thin as that in the thermosphere.

One of the most spectacular sights on the Earth, an aurora, occurs in the thermosphere (see Fig. 2.12C). When radiation from the Sun is absorbed by molecules of gas in the thermosphere, some of the molecules are broken apart to form electrically charged ions. The region of ionized gases, from 100 to 400 km (60 to 250 mi) in altitude, is called the *ionosphere.* Auroras occur when electrons streaming in from the Sun combine with the ionized gases, form neutral atoms, and give off light rays in the process.

A Closer Look

Sunlight and the Atmosphere

Many things happen to sunlight when it passes through the atmosphere. Recall from Chapter 2 and specifically from Figure 2.11 that the spectrum of the Sun's radiation measured on the surface of the Earth differs from the spectrum measured in space.

When sunlight enters the atmosphere, four things can happen: it can pass through unchanged and be absorbed by the land or the sea; it can be reflected, unchanged, back into space; it can be scattered by particles in the air; or it can be absorbed by gases in the atmosphere. The last two of these four effects—scattering and absorption—are the reasons why the sunlight spectrum at sea level differs from the spectrum in space (Fig. C12.1).

Scattering

Scattering is the dispersal of radiation in all directions, as Figure C12.2 shows. (It is often called spherical scattering to emphasize that radiation moving in a straight line is scattered equally in all directions.) Radiation comes in from the Sun in a straight line; if some of the radiation is scattered, an observer will obviously notice a reduction in the amount of radiation reaching the surface of the Earth. We are all familiar with this effect because clouds, which are simple masses of suspended particles, scatter sunlight in all directions. When a cloud passes overhead, the intensity of sunlight drops.

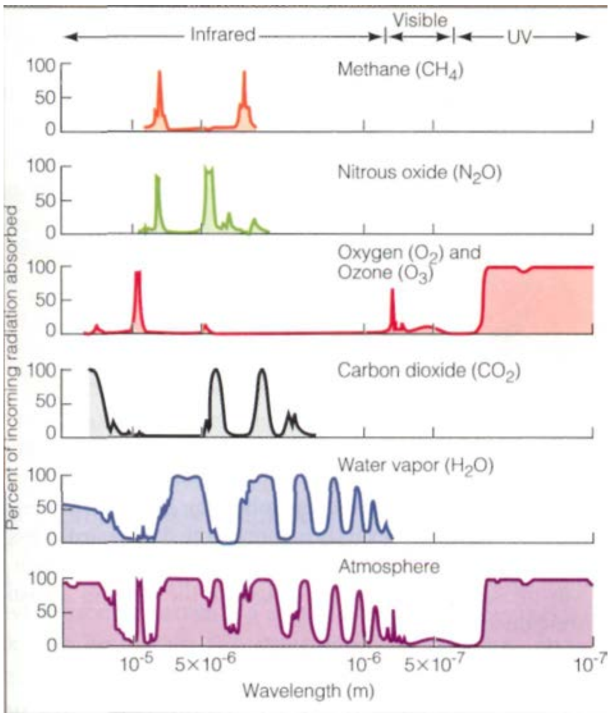


Figure C12.1 Absorption of incoming solar radiation by gases in the atmosphere. The percentage of a given wavelength range absorbed is indicated by the height of the peak. The bottom panel is the sum of all the panels above. Except for its longest wavelengths, ultraviolet radiation is almost fully absorbed.

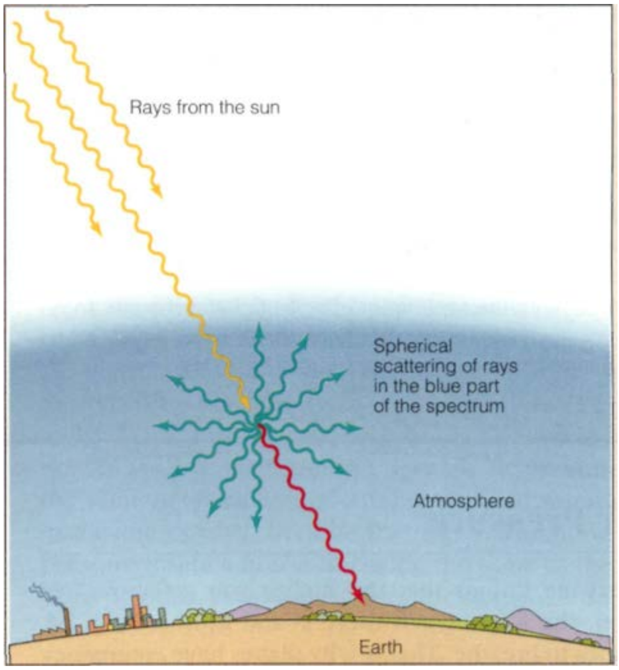


Figure C12.2 Light coming from the Sun is scattered. Air molecules are so small they scatter shorter blue wavelengths more easily than longer red and yellow wavelengths. The scattered blue wavelengths make the sky appear blue in all directions. If there were no scattering by air molecules, the sky would appear pitch black and stars would be visible all day.

Aerosols and gas molecules also cause scattering, but there is an important difference: aerosols scatter all wavelengths of visible light; gas molecules do not. In general, if the diameters of the scattering particles are less than one-tenth the diameter of the wavelength (g) of the incoming radiation, as they are in gas molecules, scattering is governed by the Rayleigh Law. The Rayleigh scattering relationship, which was discovered in 1881 by an English physicist, Lord Rayleigh (1842-1919), states that the amount of scattering is proportional to $1/(g)^4$ The smaller g is, the larger $1/(g)^4$ will be, and the greater the scattering will be for the short-wavelength end of the spectrum.

In the visible portion of the solar spectrum, the predominant scattering is at the blue end because that is the short-wavelength end (see Chapter 2). The sky appears blue to us because white, unscattered light comes straight through, whereas blue radiation in sunlight is scattered in all directions. What we see when we look at the sky is this scattered blue radiation. Similarly, because the blue end of the spectrum is reduced in intensity by scattering, the Sun appears more yellow to an Earth-bound observer than it does to an observer in space.

Absorption

Absorption can be of two kinds. Certain specific wavelengths of solar radiation make atoms or molecules vibrate with the same frequency as the wavelength. In effect, such an atom or molecule absorbs the radiation and then re-emits it at the same wavelength. However, the radiation is re-emitted equally in all directions, and so the observer of the incoming radiation sees a diminution of that wavelength. Most of the absorption by H_2O, CO_2, N_2O and CH_4 shown in Figure C12.1 is of this kind.

In the second type of absorption, molecules absorb the radiation and break apart into atoms or smaller molecules as a result. This is the process by which ozone is formed in the stratosphere. Very-short-wavelength radiation, in the ultraviolet range (g less than 3 X 10^{-7} m, 3 X 10^{-7} yd), is almost entirely absorbed by oxygen and ozone. When the molecules re-form, the trapped energy is released at a different and nonvisible wavelength.

Air Pressure

Everyone knows that the higher you go above sea level, the less oxygen there is and the harder it becomes to breathe. That is why planes have emergency oxygen masks in case cabin pressure should fail and also why mountain climbers often carry tanks of oxygen.

The supply of oxygen diminishes not because of a change in the composition of the atmosphere but rather because of a reduction in the air pressure and therefore a reduction in the air density. In as much as the size of a human lung is fixed, we obtain less oxygen when we breathe less dense air. Air at sea level and air at 9000 m (5.6 mi), near the summit of Mount Everest, have the same relative amount of oxygen—20.9 percent by volume in each case. However, a lungful of air at the top of Mount Everest has only 38 percent of the pressure that a lungful of air has at sea level and therefore only 38 percent of the amount of oxygen.

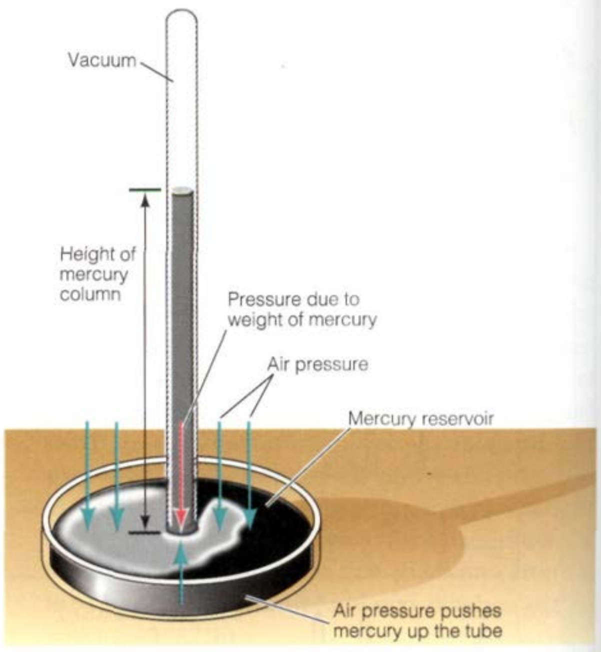


Figure 12.7 Sketch of a simple mercury barometer. Air pressure on the surface of the open bowl holds up the column of mercury in the glass tube. The downward pressure exerted by the air exactly balances the downward pressure exerted by the column of mercury on the bowl. When the air pressure changes, the height of the column adjusts in response.

Measuring Air Pressure

Air pressure is measured with a device called a **barometer.** Barometers are of two kinds, mercury and aneroid. The mercury barometer was invented in 1644 by the Italian physicist Evangelista Torricelli (1608-1647) who performed the following experiment. He sealed a 1-m-long (1-yd long) glass tube at one end and then filled it with mercury. Then, with his finger over the open end in order to prevent the mercury from running out, he inverted the tube and put the open end into a bowl of mercury (Fig. 12.7). When he removed his finger, some of the mercury flowed into the bowl, but most stayed in the tube. Torricelli reasoned that the air pressing on mercury in the bowl must be holding up the column of mercury in the glass tube.

Scientists were quick to exploit Torricelli's great discovery. Day-to-day measurements soon showed that the height of the mercury column fluctuated slightly and therefore that the air pressure must vary from time to time.

Meteorologists still measure air pressure. When television weather forecasters talk about "highs" and "lows," they are referring to air pressure that is higher or lower than average. Similarly, when a weather-

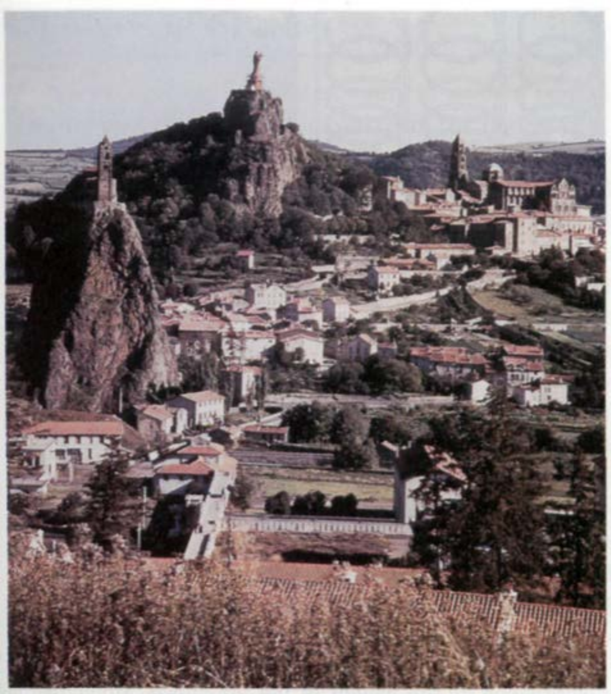


Figure 12.8 Puy-en-Velay, one of many ancient volcanic rocks in France. It was on such a puy that Blaise Pascal arranged for rock climbers to carry out the experiment proving that air pressure decreases with altitude.

watcher says "the glass" is falling or rising, the reference is to the mercury in a glass barometer that is rising (air pressure is increasing) or falling (air pressure decreasing).

The aneroid barometer (from the Greek *a* and *news,* meaning no liquid) employs a sealed metal bellows that expands and contracts as the air pressure changes.

Air Pressure Variation with Altitude

Avery important experiment was carried out in 1658 by a young French scientist, Blaise Pascal (1623-1662), who arranged for rock climbers to ascend a prominent volcanic rock in France, Puy-de-Dome (Fig. 12.8), and measure the air pressure at several places during the ascent. Despite the inconvenience of carrying a meter-long glass tube and a flask of mercury up the steep slope of the Puy, the climbers successfully performed the task. From their measurements Pascal demonstrated that air pressure decreases with altitude.

Today, meteorologists use balloons filled with helium to carry pressure-recording instruments aloft. Such balloons, which are called *radiosondes* (Fig. 12.9), can reach an altitude of 30 km (19 mi), at which point they burst and the recording instruments are parachuted back to the ground. To make measurements above 30 km, meteorologists resort to rockets and, most recently, to orbiting weather satellites.

At any given altitude, the air pressure is caused by the weight of air above. The average air pressure at sea level is about 105 Pa,[1] equivalent to the pressure produced by a column of water about 10 m (11 yd) high. Why doesn't this pressure crush us and the houses we live in? It doesn't because air pressure is the same in all directions—up, down, and sideways, inside and outside. Because the outward-pointing air pressure inside a house is exactly the same as the inward-pointing pressure outside, the net pressure on the house is zero.

As shown in Figure 12.10, the air pressure decreases smoothly with altitude. The air pressure curve is not a straight line, however, because gases are

Figure 12.9 A radiosonde, a helium-filled balloon used to carry measuring instruments into the upper atmosphere. This large, modern balloon has just been filled with helium and is about to be released. The size of the balloon can be judged from the scientists on the ground.

[1] because the earliest pressure measurements were made with mercury barometers, air pressures are still sometimes reported as the height of a mercury column. A model, or average pressure, called the standard atmosphere at sea level, is 760 mm or 29.9 inches of mercury. Another commonly employed unit is the bar, which is a pressure of 1 kg/cm^2. The standard atmosphere is 1013.25 millibars. In this book SI units are used throughout. The SI pressure unit is the pascal (Pa); the standard atmosphere is 101,325 Pa or 101.325 kilopascals (kPa).

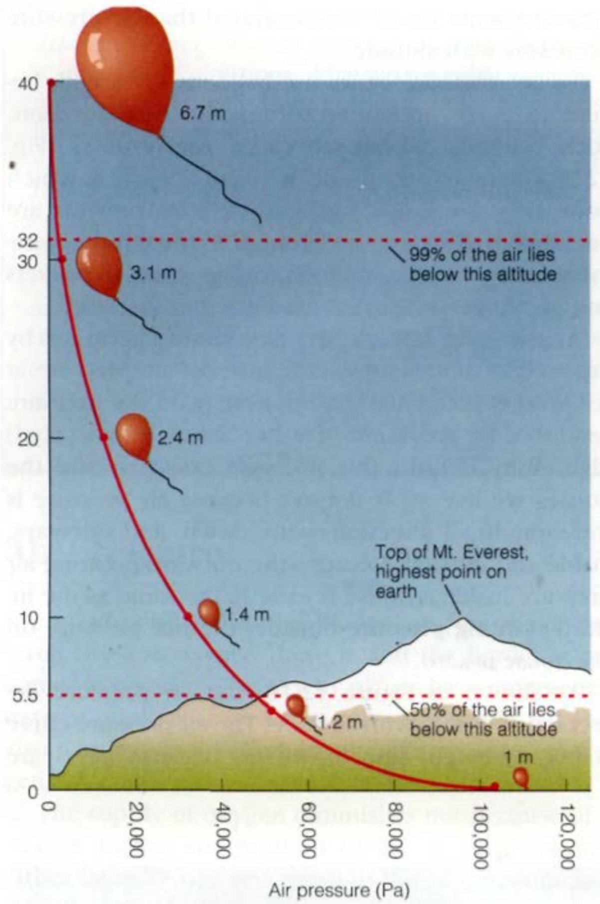


Figure 12.10 Air pressure decreases smoothly with altitude. If a helium balloon 1 m in diameter is released at sea level, it expands as it floats upward because of the pressure decrease. If the balloon did not burst, it would be 7 m in diameter at a height of 40 km.

highly compressible. This means that the air near the ground is compressed by the weight of air above (Fig. 12.11). If air were not compressible (if it were more like water, for example), the pressure-versus-altitude curve would be a straight line.

As a result of the compressibility of air, half of the mass of the atmosphere lies below an altitude of 5.5 km (3.4 mi) and 99 percent lies below 32 km (20 mi). At a height of 32 km, the air is so thin that it is like a laboratory vacuum. The 1 percent of the atmosphere that lies above 32 km continues out to an altitude of about 500 km (310 mi), with the air getting thinner and thinner until it simply merges into the vacuum of space. The few gas atoms present at the outermost fringe of the atmosphere are mostly atoms of hydrogen and helium that have reached the Earth from the Sun.

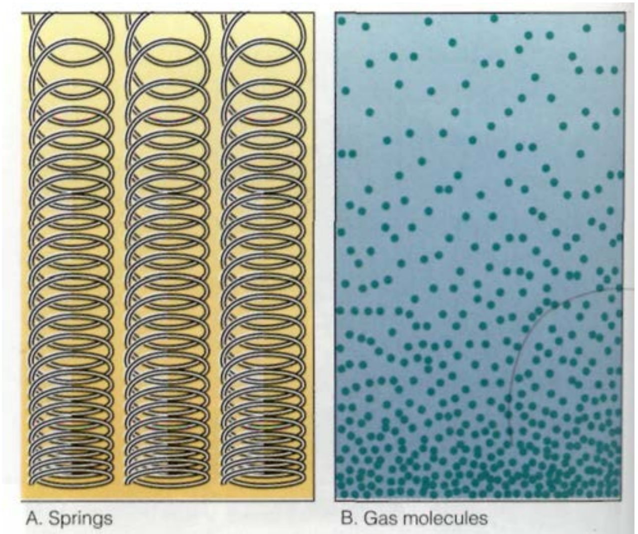


Figure 12.11 Air pressure decreases with altitude because air is compressible and behaves like a pile of springs. A. The springs near the base are compressed by the weight of the springs above. B. Air, like the springs, is compressed by the weight of the air above. Molecules of the gases nearest the ground are squeezed closer together than molecules higher up. Compression is the explanation for the shape of the curve in Figure 12.9.

MOISTURE IN THE ATMOSPHERE

The compound H_2O is so familiar that it is sometimes easy to forget what an unusual substance it is. In truth, H_2O is probably the most remarkable compound around, and in no small measure the Earth system works the way it does because H_2O has the properties it does.

One of the unusual properties of H_2O is its ability to exist in three physical states at the Earth's surface—as a solid (ice), a liquid (water), and a gas (water vapor). Of course, under some conditions of temperature and pressure, all compounds can form solids, liquids, and gases. Among naturally occurring compounds, however, only H_2O has the ability to do so under the conditions that exist at the Earth's surface. To establish why this property of H_2O is so important, first we have to explore what happens when H_2O changes from one state to another.

Changes of State

Whenever a change of state occurs in matter, energy is either absorbed by or released from the matter (Fig. 12.12). In going from a more ordered state (a solid) to a less ordered one (a liquid) or to a fully disordered one (a gas), energy is absorbed. The reverse process occurs and heat is released when the change is from a less ordered to a more ordered state. The amount of heat *released* or absorbed per gram during a change of state is known as the **latent heat** (from the Latin *latens*, meaning hidden, hence hidden heat). For example, the latent heat of condensation (less ordered water vapor condensing to more ordered liquid water) is 2260 joules per gram (J/g),whereas the latent heat of freezing (again less ordered to more ordered) is 330 J/g. The latent heat of evaporation (more ordered liquid water vaporizing to less ordered water vapor) is 2260 J/g, whereas the latent heat of melting (more ordered solid ice melting to less ordered fluid water) is 330 J/g.

One familiar phenomenon involving a change of state is evaporation. The 2260 J needed to evaporate a gram of water has to come from somewhere. The reason you feel cool after you wet yourself down on a hot day is that some of the heat needed for evaporation is absorbed from your skin and as a result your body temperature drops. Before the invention of ice chests and refrigerators, the best way to keep food cool in hot weather was to place it in a "cool safe" in which the evaporation of water kept the temperature low.

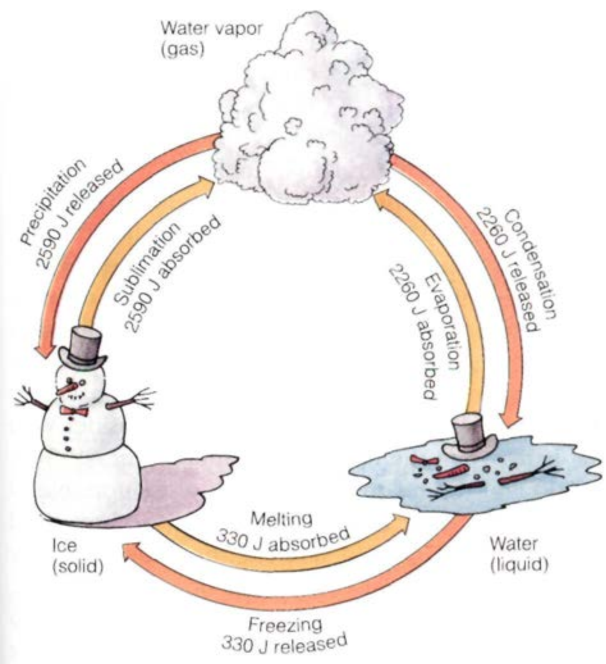


Figure **12.12** Amount of heat added to or released from a gram of H_2O during a change of state.

The six changes of state shown in Figure 12.12 (the six arrows) all play a role in weather, but evaporation and condensation far outweigh the other changes in importance. Evaporation and condensation play vitally important roles in the weather (1) because they give rise to clouds, fogs, and rain, and (2) because they are the means by which huge amounts of heat are moved from equatorial regions toward the poles. To explore how these phenomena occur, it is necessary to discuss humidity, which, as mentioned previously, is the amount of water vapor in the atmosphere.

Relative Humidity

Water vapor gets into the air by evaporation. Evaporation is a process by which fast-moving liquid molecules manage to escape from the liquid and pass into the vapor above. Of course, because molecules in a vapor move randomly in all directions, some of the gas molecules in the vapor will also move back into the liquid. When the number of molecules that evaporate (going from liquid to gas) equals the number condensing (going from gas to liquid), the vapor is said to be *saturated.*

It is common practice to report vapor properties in terms of the vapor pressure. One of the important properties of gases is that pressures in a mixture of gases are additive; this property is known as *Dalton's law of partial pressures*, and it means that the total pressure of a mixture of gases is the sum of the partial pressures exerted by all the individual gases present. Partial pressure, in turn, is a measure of the volume percent of a gas in a mixture. For example, because the content of oxygen in dry air is 20.9 percent by volume, the fraction of the pressure of standard air (101.325 kPa) attributable to oxygen is 101.325 X 0.209 = 21.2 kPa. The additivity of gas pressures is the reason that the water vapor content of air is just as often reported as a pressure and as a percentage. The saturation vapor pressure of water at various temperatures is shown in Figure 12.13-

The *saturation vapor pressure,* which is also known as the *water vapor capacity* of air at any given temperature, cannot be exceeded. If the vapor pressure reaches the capacity, condensation starts and the saturation vapor pressure is maintained. (**Condensation** is the formation of a more ordered liquid from a less ordered gas.) The vapor pressure can, however,

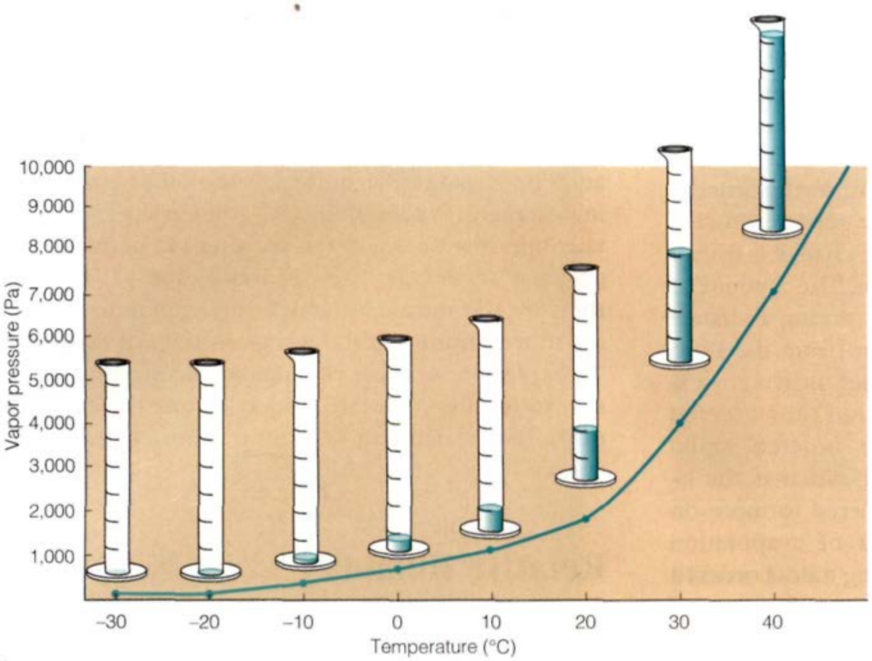


Figure 12.13 Saturation pressure of water vapor as a function of temperature. The amount of H_2O present is shown by the size of the drop of water that would form if all the vapor in a kilogram of saturated air at sea level were condensed.

be lower than the saturation value. For example, if saturated air is removed from contact with water and is then heated, the vapor pressure will fall below saturation level and the air will then be undersaturated.

Meteorologists prefer to use the term **relative humidity** rather than undersaturation when they are discussing the amount of water vapor in undersaturated air. The **relative humidity** is the ratio of the vapor pressure in a sample of air to the saturation vapor pressure at the same temperature, expressed as a percentage. For example, saturated air at 20°C (68°F) has a water vapor pressure of 2.338 kPa. Air at 20°C with a water vapor pressure of 1.403 kPa will therefore have a relative humidity of 1.403/2.338 X 100% or 60 percent.

Note that the relative humidity does not refer to a specific amount of water vapor in the air; rather, it refers to the ratio of what is present at a given temperature to the maximum possible amount that the air could hold at the same temperature. The fact that relative humidity is a ratio sometimes confuses people and leads to misconceptions. One misconception is that if air feels damp and humid, it must contain more H_2O than air that feels dry. The confusion arises because temperature exerts such a strong control on the water vapor capacity of air. For example, desert air at 30°C (86°F) and relative humidity of 25 percent feels very dry, even though it contains 6.62 g of H_2O per kg of air, whereas air at 10°C (50°F) and relative humidity of 80 percent, which feels damp and humid, contains only 5.60 g (0.2 oz) of H_2O per kg of air.

Relative humidity can be changed in two ways—by addition of water vapor or by change of temperature. When the relative humidity is below 100 percent and air is in contact with water, evaporation will raise the relative humidity. This is why air in contact with the ocean usually has a high relative humidity and why so much of the water vapor that enters the atmosphere does so over the oceans (Fig. 12.14).

Temperature changes can also change the relative humidity whether or not H_2O is added. If the amount of water vapor in the air is kept constant and the temperature drops, the relative humidity will rise. The temperature at which the relative humidity reaches 100 percent and condensation starts is called the **dew point.** By contrast, if the temperature rises, the relative humidity drops. People who live in centrally heated houses are very familiar with the problem of decreases in relative humidity. In winter, as air is drawn in from outside and heated by a furnace, the relative humidity drops. For example, consider what happens when the outside air temperature is —10°C (14°F) and the relative humidity is a comfortable 60 percent; if the air is now drawn into the house and heated to 25°C (77°F) without the addition of any water vapor, the relative humidity drops to 6 percent, a level at which many people feel discomfort. To counteract the effects of low humidity, a humidifier must be used to add water vapor to the heated atmosphere.

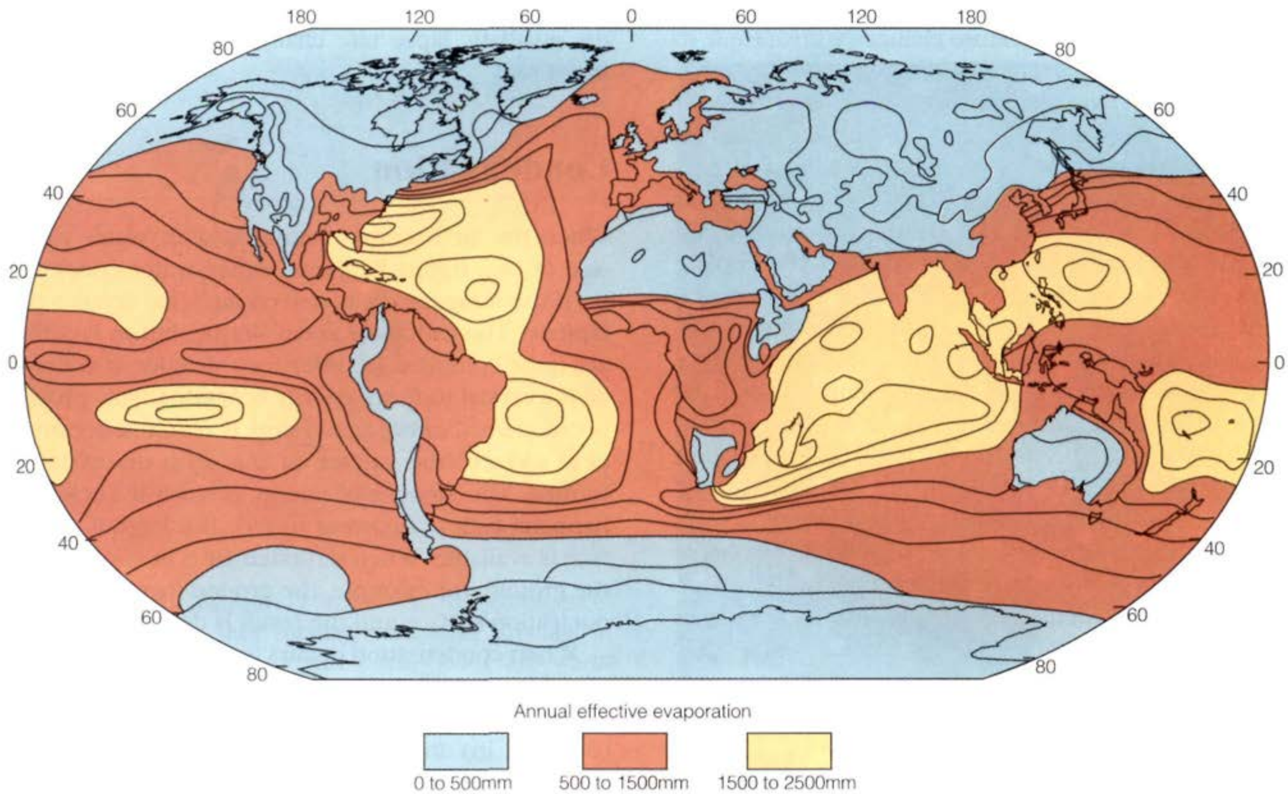


Figure 12.14 The annual addition of water vapor to the atmosphere as a function of geography. The amount evaporated per year is measured in millimeters of water. Note that areas of highest evaporation are over the oceans in equatorial and midlatitudes. Evaporation is low in the deserts because deserts have so little water to evaporate.

CONDENSATION AND THE FORMATION OF CLOUDS

If you have ever pumped up a bicycle tire, you will have noticed that the pump becomes hot when the air was compressed. Similarly, if the valve is opened and the compressed air in a tire is allowed to escape, the air is noticeably cool as it expands. These two effects—compressional warming and expansional cooling—are examples of what are called **adiabatic processes**, from the Greek *adiabatos,* meaning no passage. Adiabatic processes are so named because they are processes that occur without the addition or subtraction of heat from an external source. When air is compressed, the mechanical energy of pumping is converted to heat and the temperature rises as a result. When compressed air expands, the energy required comes from the heat energy of the gas, and consequently a temperature drop follows.

Warm air is less dense than cold air and therefore rises, creating a convection cell in the process. Because it is warmest at ground level, air in the troposphere is continually rising or, after cooling, falling. However, because air pressure decreases with increasing altitude, the rising air expands, and because there is no heat source in the troposphere, the rising air expands adiabatically and so its temperature falls. In the case of sinking air, the reverse happens—the air is compressed and so its temperature rises.

Adiabatic Lapse Rate

When a parcel of unsaturated air rises and expands adiabatically, the temperature drops at a constant rate of 1°C/100 m. Conversely, if cool, unsaturated air sinks toward the Earth's surface, it is compressed and the temperature rises at a rate of 1°C/100 m (Fig. 12.15). The way temperature changes with altitude in rising or falling unsaturated air is called the **dry adiabatic lapse rate.** Eventually, when air rises far enough, it will cool sufficiently to become saturated and condensation will start. The latent heat of condensation that is released as the vapor condenses works against the adiabatic cooling process; in other words, the release of latent heat slows the cooling rate. The greater the amount of latent heat released (the greater the condensation), the less the temperature increase with altitude. The way temperature drops in a rising mass of saturated air is called the **moist adiabatic lapse rate** and it ranges from 0.4°C/100 m to 0.9°C/100 m, with an average of 0.6°C/100 m (Fig. 12.15).

Note that the moist adiabatic lapse rate is always less than the dry rate because of the addition of latent heat to the rising air above the level of condensation. Note, too, that when clouds start to form as a parcel of air reaches saturation and condensation begins, the dry adiabatic lapse rate changes immediately to the moist rate.

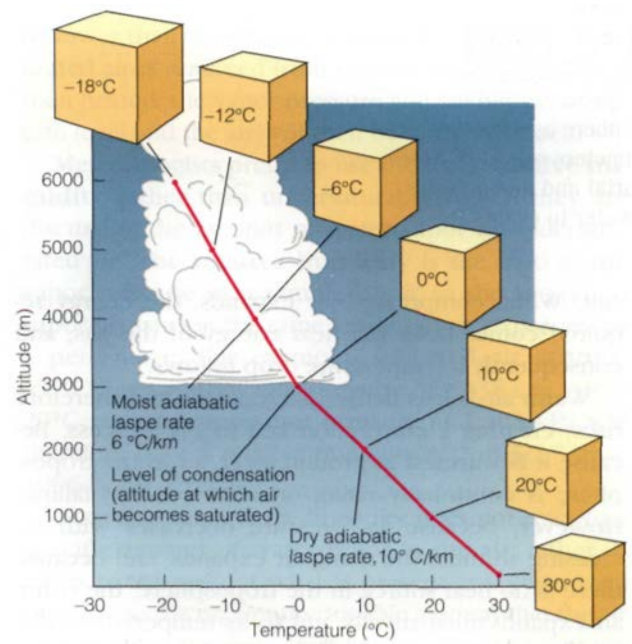


Figure 12.15 As an unsaturated mass of air rises, it expands and cools at the dry adiabatic lapse rate (10°C/km). When the air temperature falls to the point where the air is saturated, condensation commences and latent heat is released. A further increase in altitude causes more condensation and the release of more latent heat; the air temperature now decreases at the moist adiabatic lapse rate (6°C/km). Note that the speed with which the air rises does not necessarily change; what changes is the temperature drop with altitude. Shown beside the curve are the volumes of a mass of rising air that starts as a cube 1 km on an edge.

Condensation

When the air becomes saturated with water vapor, one of two things happens: either water condenses or, if the temperature is low enough, ice crystals precipitate. The processes seem simple, but in fact they are quite complex. In order for a droplet of water or an ice crystal to form, energy is needed. This process is called *nucleation,* and energy is required because a new surface (the surface of the drop or crystal) is formed. The amount of energy is small if nucleation happens on a preexisting surface but large if no surface is available. When saturated air is in contact with the ground, for example, the ground itself serves as a nucleation surface and the result is dew or frost.

When condensation occurs in a rising parcel of air, aerosols provide the nucleation surfaces. Most aerosols are tiny—no more than 1 micrometer (0.00003 in) in diameter. When condensation on aerosols commences, it happens very rapidly and water droplets reach a diameter of 20 to 25 micrometers (0.0008 to 0.001 in) in about a minute. Thereafter, droplet growth rate slows because the remaining water vapor must be spread over billions of droplets as it condenses.

Cloud droplets are so small that air turbulence within the cloud keeps them suspended. A density of about 1000 droplets/cm^3 (or 1 drop/mm^3) is sufficient to keep the drops apart. When the density of droplets increases above this value, they start to coalesce, and eventually a few drops become too big to remain suspended. As the drops fall, they hit and coalesce with more and more droplets until finally a raindrop has formed (Fig. 12.16). A single raindrop contains about 1 million cloud droplets.

When the cloud temperature is below 0°C (32°F), the process of precipitation is more complex than simple condensation. Complexity arises because water droplets can be supercooled below 0°C without freezing to ice. The person who first recognized the role that supercooled water plays in cloud precipitation was a Swedish scientist, Tor Bergeron (1891-1977), and the process is now named the **Bergeron process** in honor of his discovery. What Bergeron discovered is that ice crystals grow at the expense of supercooled water drops.

Clouds with temperatures between 0°C and -9°C (32°F and 16°F) contain only supercooled water droplets. When the temperature is between —10°C

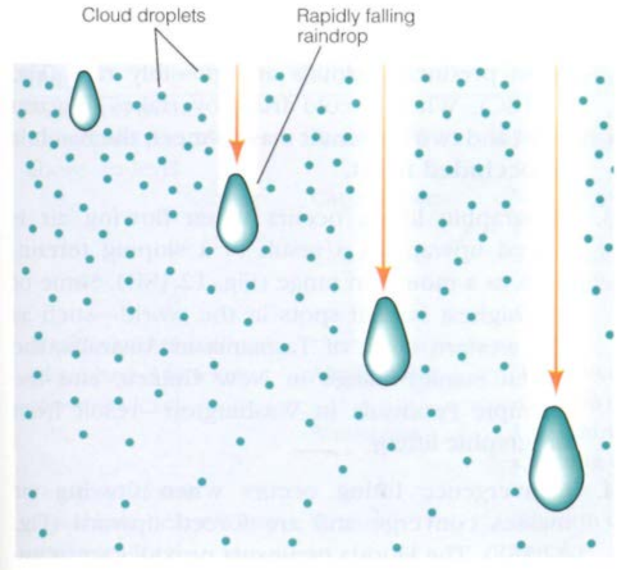


Figure 12.16 Growth of raindrops by coalescence. When droplets fall, they combine with other droplets in their path. Coalescence occurs when clouds are warm enough to keep ice from forming.

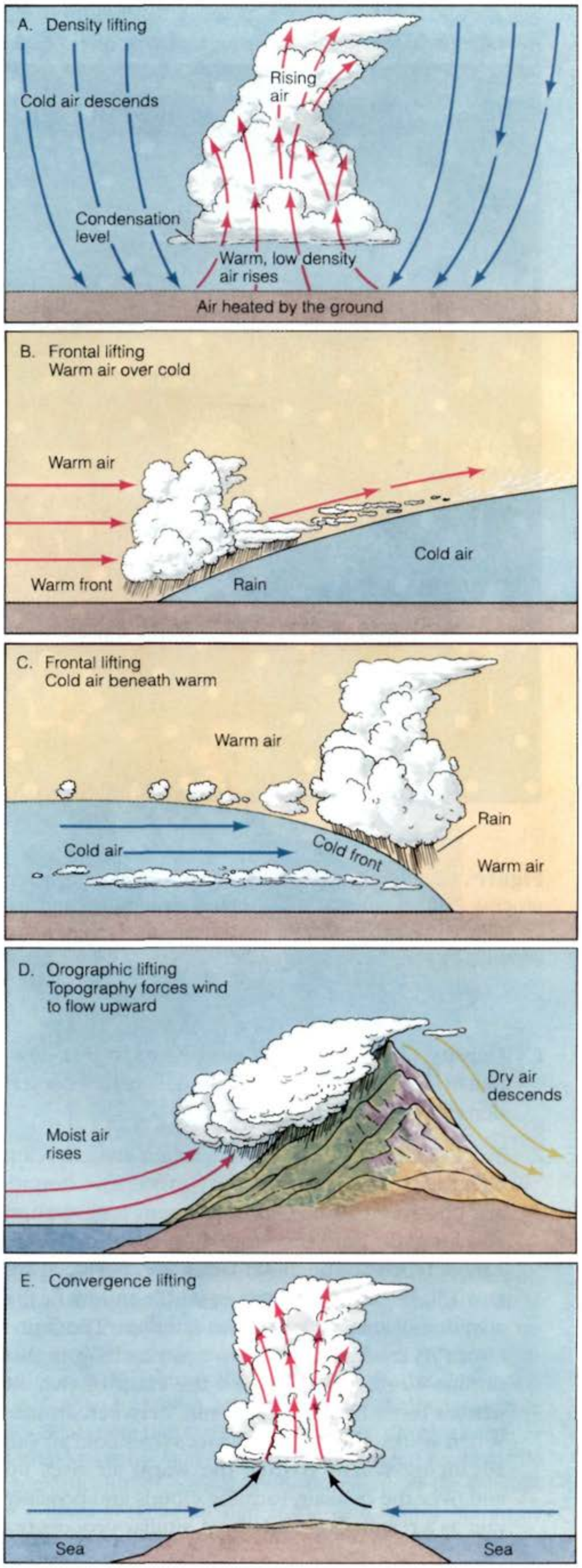
and -20°C (14°F and —4°F), ice crystals also nucleate so that the cloud becomes a mixture of supercooled water drops and ice crystals. Below -20°C, water drops disappear and clouds contain only ice crystals. As air cools and a cloud grows, therefore, the mixture of supercooled water droplets and ice crystals becomes increasingly dominated by ice crystals. Bergeron discovered that, in a mixture of supercooled water droplets and ice crystals, the water droplets slowly evaporate and release water vapor that is then deposited on the ice crystals, making them grow larger. Eventually the ice crystals become so large that they start to fall (Fig. 12.17). If the temperature all the way to the ground is everywhere below 0°C, the result is snowflakes. If the temperature near the ground is above 0°C (32°F), the ice crystals melt and raindrops hit the ground. (In other words, most rain starts out as ice crystals.) If raindrops fall through a layer of air near the ground where the temperature is below 0°C, the drops freeze and sleet is the result.

Clouds

Clouds form when air rises and becomes saturated in response to adiabatic cooling. There are four principal reasons for the upward movement of air. Although it is possible to separate the reasons for discussion, most individual circumstances involve more than one lifting force. The four lifting forces are:

Figure 12.17 Growth of ice crystals by the Bergeron process. Supercooled water droplets evaporate, and ice crystals grow by incorporating the newly formed water vapor.

1. Density lifting, which occurs when warm, low-density air rises convectively and displaces cooler, dense air (Fig. 12.18A).
2. Frontal lifting, which occurs when two flowing air masses of different density meet. The boundaries between air masses of different temperature and humidity, and therefore different density, are called **fronts.** The boundaries are between 10 and 150 km (6 and 93 mi) in width and mark the advance of one air mass into another. The name "front" is used because Norwegian meteorologists during World War I likened the clash of two air masses to battle lines, or fronts, between armies. When warm, humid air advances over cold air (an advancing **warm front),** the warm air rises up and over the cold air, forming clouds and possibly rain as a result (Fig. 12.18B). A similar process oc-



curs when denser, cold air flows in and displaces warm air by pushing it upward (a **cold front**), again producing clouds and possibly rain (Fig. 12.18C). When a cold front overtakes a warm front and two cooler air masses meet, the result is an **occluded front.**

3. Orographic lifting occurs when flowing air is forced upward as a result of a sloping terrain, such as a mountain range (Fig. 12.18D). Some of the highest rainfall spots in the world—such as the western coast of Tasmania in Australia, the Owen Stanley Range in New Guinea, and the Olympic Peninsula in Washington—result from orographic lifting.

4. Convergence lifting occurs when flowing air masses converge and are forced upward (Fig. 12.18E). The Florida peninsula provides an example; air flows landward off the ocean from both east and west; the two flowing air masses collide and force some of the air to rise. Clouds therefore form, and the result is the familiar frequent afternoon thunderstorms.

Clouds are visible aggregations of minute water droplets, tiny ice crystals, or both. They are such prominent and beautiful features of the sky that poets from time immemorial have written about them, painters beyond number have painted them, and meteorologists spend a great deal of time studying them.

Because clouds form by condensation of water vapor, all common clouds are phenomena of the troposphere. They are classified on the basis of shape, appearance, and height into three families: cumulus, stratus, and cirrus (Table 12.1, Figs. 12.19 and 12.20).

Cumulus clouds are puffy, globular, individual clouds that form when hot, humid air rises convectively, reaches a level of condensation where cloud formation starts, but continues to rise. These are the flat-based cauliflower-shaped clouds that children like to draw—the flat base marking the level of condensation. When cumulus clouds coalesce to form a puffy layer, the term *stratocumulus* is applied.

When large cumulus clouds rise to the top of the troposphere, they expand horizontally and form *at-*

Figure 12.18 Lilting forces. A. Density lifting causes convection cell as warm, low-density air rises and cold, higher-density air sinks. B. Frontal lifting. A warm front occurs when flowing warm air overrides cold air and is forced upward. C. Frontal lifting. A cold front occurs when a wedge of forward-moving cold air slides under warm air and forces it upward. D. Orographic lifting occurs when flowing air is forced upward by mountains or other sloping ground. E. Convergence lifting occurs when masses of air collide and are forced upward.

Table 12.1
Classification of Clouds in the Troposphere by Altitude

Height	Name	Shape and Appearance
High-level clouds		
Cloud base 6 to 15 km above sea level	Cirrus	Feathery streaks
	Cirrocumulus	Small ripples and delicate puffs
	Cirrostratus	Translucent to transport sheet, like a veil across the sky
Middle-level clouds		
Cloud base 2 to 6 km above sea level	Altocumulus	White to dark gray puffs and elongate ripples
	Altostratus	Uniform white to gray sheet covering the sky
Low-level clouds		
Cloud base below 2 km above sea level	Stratus	Uniform dull gray cover over the sky
	Nimbostratus	Uniform gray cover, rain generally falling
	Stratocumulus	Patches of soft gray; in places, patches may coalesce to a layer
Clouds with great vertical development		
Cloud base below 3 km above sea level	Cumulus	Puffy cauliflower shape with flat base
	Cumulonimbus	Large, puffy; white, gray, and black; great vertical extent, often with anvil-shaped head

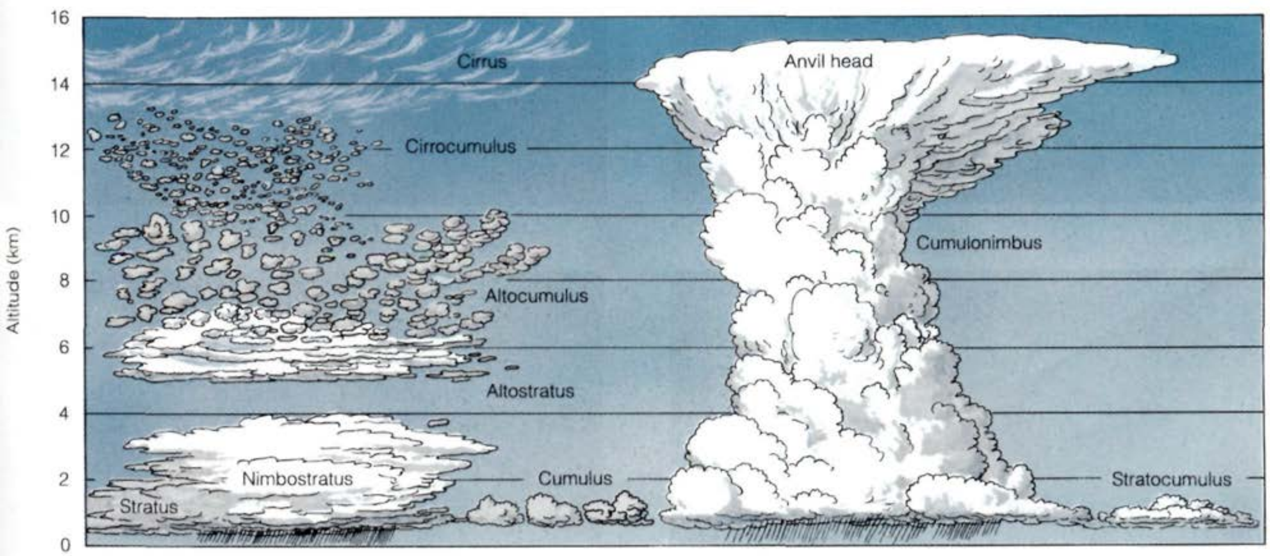


Figure 12.19 The altitudes of clouds. An anvil head is the flattened top of a cumulonimbus cloud that spreads across the top of the troposphere.

mulonimbus clouds. These are the familiar thunderstorm clouds or "thunderheads" of summer. A cumulonimbus cloud contains a great deal of energy and turbulence. Some of the energy causes thunder and lightning within a cloud, between adjacent clouds, and between clouds and the ground.

Cumulus clouds that form at altitudes between 2 and 6 km (1 and 4 mi) are given the modifying name *altocumulus* clouds; those that form between 6 and 15 km (4 and 10 mi) are called *cirrocumulus* clouds. Frequently, altocumulus and stratocumulus clouds are arranged in regular rows or clumps separated by clear sky. Convection cells up to several hundred meters across give rise to the patterns.

Stratus clouds are sheets of cloud cover that form at altitudes from 2 km to about 15 km and cover the entire sky. Stratus clouds form when air rises as a result of frontal lifting, reaches its level of condensation, and then spreads laterally but not vertically. If the cloud blanket is several kilometers thick, the day is dark and dreary and the cloud is called *nimbostratus.*

Depending principally on the altitude, stratus clouds, analogous to cumulus clouds, are given modifying names; between 6 and 8 km (4 and 5 mi), *altostratus;* between 8 and 12 km (5 and 8 mi), *cirrostratus.*

Cirrus clouds are the highest clouds in the troposphere. Looking like fine, wispy filaments, or feathers, cirrus clouds form only above 6 km (4 mi) in altitude and are composed entirely of ice crystals.

Two rare kinds of cloud are known to form in the stratosphere. *Nacreous clouds* are beautiful translucent sheets (meaning light gets through, like a frosted window pane) of minute ice crystals that form at altitudes between 20 and 30 km (12 and 19 mi). Even less common are *noctilucent clouds,* which are so thin they look like gossamer veils. They are composed entirely of minute ice crystals and form at altitudes as high as 90 km (56 mi). No clouds have ever been reported above 90 km, and none are likely to be reported because water vapor is too scarce to form clouds at such altitudes.

A.

B.

C.

Figure 12.20 Principal types of clouds. A. Cumulus B. Cumulonimbus. Note the plume spreading sideways. C. Altocumulus D. Cirrostratus E. Stratocumulus F. Stratus G. Altostratus H. Nimbostratus I. Cirrus

D.

E.

F.

G.

H.

I.

Guest Essay

Atmosphere on Other Planets

James C. G. Walker is professor of atmospheric, oceanic and space Sciences, professor of geological sciences, and Arthur F. Thurnau Professor at the University of Michigan, where he also serves as director of the program in Environmental Studies. He has conducted research in ionospheric and space physics, the evolution of planetary atmospheres, atmospheric chemistry, and climate change. He has written several books and numerous scientific articles dealing with these subjects.

Although atmospheres on other planets are individually interesting, their scientific importance lies in their status as a class of objects against which to test our knowledge and understanding of the science of planetary atmospheres. It would be hard to discover the fundamental laws and processes of biology if we had only one organism to study. Like biologists with different organisms, atmospheric scientists seek a deeper understanding of the natural laws of atmospheres by comparing the atmospheres of different planets. The practical goal is to improve our knowledge of the origin, evolution, and probable fate of our own atmosphere.

Mercury and Earth's moon have no atmospheres. Because they are small bodies, their gravitational fields are too small to bind gases against their tendency to expand into the vacuum of space. At the other extreme are the giant planets, Jupiter, Saturn, Uranus, and Neptune. They are so massive that their gravitational fields have retained much of the gas that was present in their regions of space when the solar system formed. Like the Sun, they are composed mainly of hydrogen and helium, the most abundant elements in the universe. Data from spacecraft missions to these planets have raised interesting questions about the chemical reactions in hydrogen-rich atmospheres and about the circulation of thick, deep atmospheres on rapidly rotating planets. Some of the satellites of the giant planets are less exotic, but the greatest rewards for Earth system scientists come from study of the Earthlike planets, Mars and Venus.

In the early 1960's, when I was a graduate student, Earth's was the only atmosphere for which we had much data. It was believed that the atmospheres of Mars and Venus were composed principally of nitrogen, as Earth's atmosphere is. This idea seems foolish in retrospect but must have made sense at the time. In fact, both atmospheres are composed mostly of carbon dioxide, with ni-

A HABITABLE PLANET

In the opening essay of this chapter we learned that oxygen must be present within well-defined limits in an atmosphere if a planet is to be habitable. The reason that no other planet or moon in the solar system is habitable is the absence of oxygen (see the "Guest Essay" by Professor Walker).

In addition to the presence of *oxygen* in the atmosphere, two other essential criteria make a planet habitable: water vapor must be present, and the temperature must be neither too high nor too low.

Humans cannot live for long where the air is completely dry because water vapor is needed for our lungs to work properly. A habitable planet must therefore have a hydrosphere so that there is some way of getting water vapor into the atmosphere.

Where temperature is concerned, the limits of a habitable planet are made clear by the way we humans live. The parts of the body that must be protected from temperature fluctuation are the core organs—the brain, heart, lung, liver and digestive system. Although the temperature of the skin can vary widely with no harmful consequences, the temperature of the core organs, which is 37°C (98.6°F), cannot vary safely by more than ± 2°C (2°F). To maintain a stable core-organ temperature, the body has two cooling mechanisms—perspiration and dilation of blood vessels in the skin, both of which cool because they get rid of heat—and two heating mechanisms-shivering and contraction of blood vessels in the skin. In addition to these natural mechanisms, humans can put on clothes and take other measures to avoid exposure to extreme temperatures.

The most effective way to handle the body's temperature requirements, of course, is to live where the annual mean temperature is comfortable. Over 90 percent of the global population lives in places where the annual mean temperature is between 6°C (43°F) and 27°C (81°F). A habitable planet must therefore have regions that fall in this temperature range. Much of the Earth enjoys annual mean temperatures in the equitable 6°C to 27°C range largely as a result of the atmosphere, which acts as a warming blanket.

trogen concentrations of just a few percent. Earth is the odd planet because nearly all of Earth's carbon dioxide has been removed from the atmosphere and deposited as carbonate minerals in sedimentary rocks. If this carbon were in the atmosphere, the ratios of carbon to nitrogen would be similar in the atmospheres of all three planets. Why has Earth's carbon dioxide been almost completely extracted from the atmosphere? Probably because Earth has abundant water, which has made possible the weathering reactions that extract carbon dioxide from the atmosphere. The atmospheres of Mars and Venus are both very dry. As far as we know, Earth, Mars, and Venus were assembled out of more or less the same material with more or less the same complements of water and other volatile compounds. Why, then, are these atmospheres so dry? Probably because Mars is too cold, Venus is too hot, and Earth is just right. Planetary scientists call this the Goldilocks problem.

Mars is further from the Sun than Earth and has a thin atmosphere with a surface pressure 160 times smaller than ours. Over most of the surface, temperatures never rise as high as 0°C, so most of Mars's water is presumed to be preserved in permanant deposits of subsurface ice; much of Mars's carbon may be similarly locked in solid form as dry ice and as carbonate minerals. The thin atmosphere appears to be a consequence of the low temperature. Venus is closer to the Sun than Earth and has a massive atmosphere with a surface pressure 70 times larger than ours. Because of the greenhouse effect of this massive atmosphere, the surface is a searing 480°C, well above the boiling point of water. Perhaps Venus has always been too hot for water to condense. Instead, water vapor in the atmosphere may have been broken apart by ultraviolet radiation from the Sun into its constituent elements, oxygen and hydrogen. The light hydrogen could have escaped into space, while the heavier oxygen could have reacted with rocks to become incorporated into the solid part of the planet.

From the study of planetary atmospheres, therefore, we learn that a habitable planet like ours is an improbable object. Small differences in planetary origins lead to widely divergent evolutionary paths. A planet a little too close to the Sun becomes hot and dry like Venus; a planet a little too far grows cold and dry like Mars. Too large a planet captures a massive atmosphere like those of the giant planets; too small a planet ends up with no atmosphere at all, like Mercury and the Moon. The requirements for habitability are stringent indeed.

Our explanations of why planetary atmospheres are so different are plausible but not necessarily correct. It is much easier to explain differences once they have been discovered than it is to predict differences in advance, and the best test of a scientific theory is its ability to predict what has not yet been observed. Planetary atmospheric science has achieved very few predictive successes so far. We need more planets.

Summary

1. Weather is the state of the atmosphere at a given time and place; the variables that define weather are temperature, air pressure, humidity, cloudiness, and wind speed and direction.
2. Climate is the average weather condition at a given place. In order to determine the climate, the weather must be averaged over a period of years.
3. Two energy sources drive the atmosphere: the Sun's heat and the Earth's rotation.
4. The relative composition of the atmosphere is constant on a dry and aerosol-free basis. Both the humidity and the amount of aerosols vary from place to place and time to time, but the gases of the atmosphere have fixed relative proportions.
5. The three major gases, nitrogen, oxygen, and argon, account for 99.96 percent of dry, aerosol-free air.
6. The minor gases, water vapor, carbon dioxide, methane, nitrous oxide and ozone are important for their roles in trapping the Sun's heat and shielding the Earth from ultraviolet rays.
7. The temperature profile of the atmosphere is highly structured. In the troposphere, to an altitude between 10 and 16 km, temperature decreases with altitude. In the stratosphere, 16 to 50 km, the temperature increases with altitude; temperature then decreases with altitude in the mesosphere (50 to 85 km) and finally increases again with altitude in the thermosphere (85 to 500 km).
8. The atmosphere extends out to an altitude of about 500 km, at which point it blends into the vacuum of space.
9. Air pressure decreases with altitude, but because air is highly compressible, the decrease is not linear. Fifty percent of the atmosphere lies below an altitude of 5.5 km and 99 percent below 32 km.
10. The troposphere is where most of the Earth's weather is generated, where clouds form, and where rain and snow develop.
11. When H_2O changes from one state to another, heat is absorbed or released. The most important

changes of state as far as the weather is concerned are condensation, precipitation, and evaporation. Condensation and precipitation release latent heat. Evaporation absorbs heat.

12. The amount of water vapor in the atmosphere cannot exceed the saturation capacity. Air that has reached the saturation capacity has 100 percent relative humidity.
13. In air that contains less water vapor than the saturation value, the water vapor content is measured by the relative humidity (the ratio of the vapor in a given sample of air to the saturation vapor pressure at the same temperature, expressed as a percentage).
14. Rising air cools adiabatically, which means without losing or gaining heat energy. When adiabatically cooled air becomes saturated with water vapor, condensation commences and clouds start to form.
15. Rising or falling air masses cool or warm respectively as a result of adiabatic expansion or compression. If the air is unsaturated, the temperature changes at the adiabatic lapse rate; if unsaturated air becomes saturated, so that latent heat is added, the temperature changes at the moist adiabatic lapse rate.
16. Aerosols serve as the nuclei on which water droplets and ice crystals nucleate.
17. In clouds that contain a mixture of supercooled water droplets and ice crystals, the water droplets evaporate and the ice crystals grow larger. This is known as the Bergeron process.
18. Droplets of water coalesce in warm clouds, and when the coalesced drops are large enough, they fall as rain. Most clouds are so cold that ice particles eventually predominate over water droplets. As ice particles fall into warmer air below the clouds, they melt and reach the ground as rain.
19. Clouds form when air rises adiabatically, becomes saturated, and condensation commences as a result of four kinds of lifting forces: density lifting, frontal lifting, orographic lifting, and convergence.
20. Clouds are classified by shape, appearance, and height. There are three cloud families, based on shape: cumulus, stratus, and cirrus clouds. Modifying prefixes are used to designate the altitude of the clouds; for example, cirrocumulus and cirrostratus are high-altitude cumulus and stratus clouds, respectively.
21. Two more kinds of clouds form in the stratosphere—nacreous and noctilucent clouds.

Important Terms to Remember

adiabatic lapse rate
(dry and moist) (p. 326)
adiabatic process (p. 325)
aerosol (p. 315)
air (p. 315)
barometer (p. 320)
Bergeron process (p. 326)
climate (p. 314)
cloud (p. 328)
cold front (p. 328)
condensation (p. 323)
dew point (p. 324)
front (p. 327)
heat (= heat energy) (p. 317)
humidity (p. 315)
insolation (p. 317)
latent heat (p. 323)
mesopause (p. 318)
mesosphere (p. 318)
occluded front (p. 328)
relative humidity (p. 324)
stratopause (p. 318)
stratosphere (p. 318)
stratus clouds (p. 330)
temperature (p. 317)
thermosphere (p. 318)
tropopause (p. 318)
troposphere (p. 317)
warm front (p. 327)
weather (p. 314)

Questions for Review

1. What are the five variables used to define weather and climate?
2. How does weather differ from climate and how is climate measured from weather variables?
3- What are the sources of energy that drive activities in the atmosphere?
4. Mars and Venus have atmospheres, but are their atmospheres air? Why or why not?
5. The main ingredient of air is nitrogen. What is the second most common ingredient? What percentage of the air is made up by this second most abundant ingredient?

6. What is the difference between humidity and relative humidity?
7. Why is it possible for dry air in a desert to contain more water vapor than moist air in the Arctic?
8. Which minor gases in the atmosphere are known as the greenhouse gases, and why?
9. Ultraviolet radiation from the Sun is lethal to many forms of life, including humans. Explain how it is possible for humans to live on the Earth's surface without being harmed by these lethal rays.
10. Explain why, since the composition of the air doesn't change, there is less oxygen to breathe at the top of Mount Everest than at sea level.
11. The way air pressure decreases with increasing altitude is nonlinear. Explain why this is so.
12. How does heat differ from temperature?
13. The energy flux from the Sun that reaches the outer edge of the atmosphere is 1370 W/m^2; the energy flux that reaches the Earth's surface (the insolation) is considerably less than 1370 W/m^2. Cite three effects that cause the reduction.
14. Name the four temperature regions of the atmosphere in order of altitude, starting with the lowest region. Give the approximate altitudes where one temperature zone passes into another.
15. Why does temperature decrease with altitude in the troposphere but increase with altitude in the stratosphere?
16. What is latent heat? Give two examples of a change of state in which latent heat is released. Give two examples of a change in which latent heat is absorbed.
17. When air is rapidly compressed, as in a bicycle pump, it becomes heated. Explain 'why this is so. What is the name given to processes such as the heating or cooling of a gas as a result of compression or expansion?
18. When a mass of air rises, the rate at which temperature changes with altitude above cloud level (the level of condensation) is different from that rate below cloud level. Explain why this is so.
19. What is the Bergeron process, and what role does it play in the formation of raindrops?
20. Describe four ways by which a mass of air can be lifted. Why are these lifting processes important for the formation of clouds?
21. What is the difference between a cold front and a warm front? Rain is commonly associated with both kinds of front. Why should that be so?
22. Name the three major families of clouds and describe their general differences.

Questions for A Closer Look

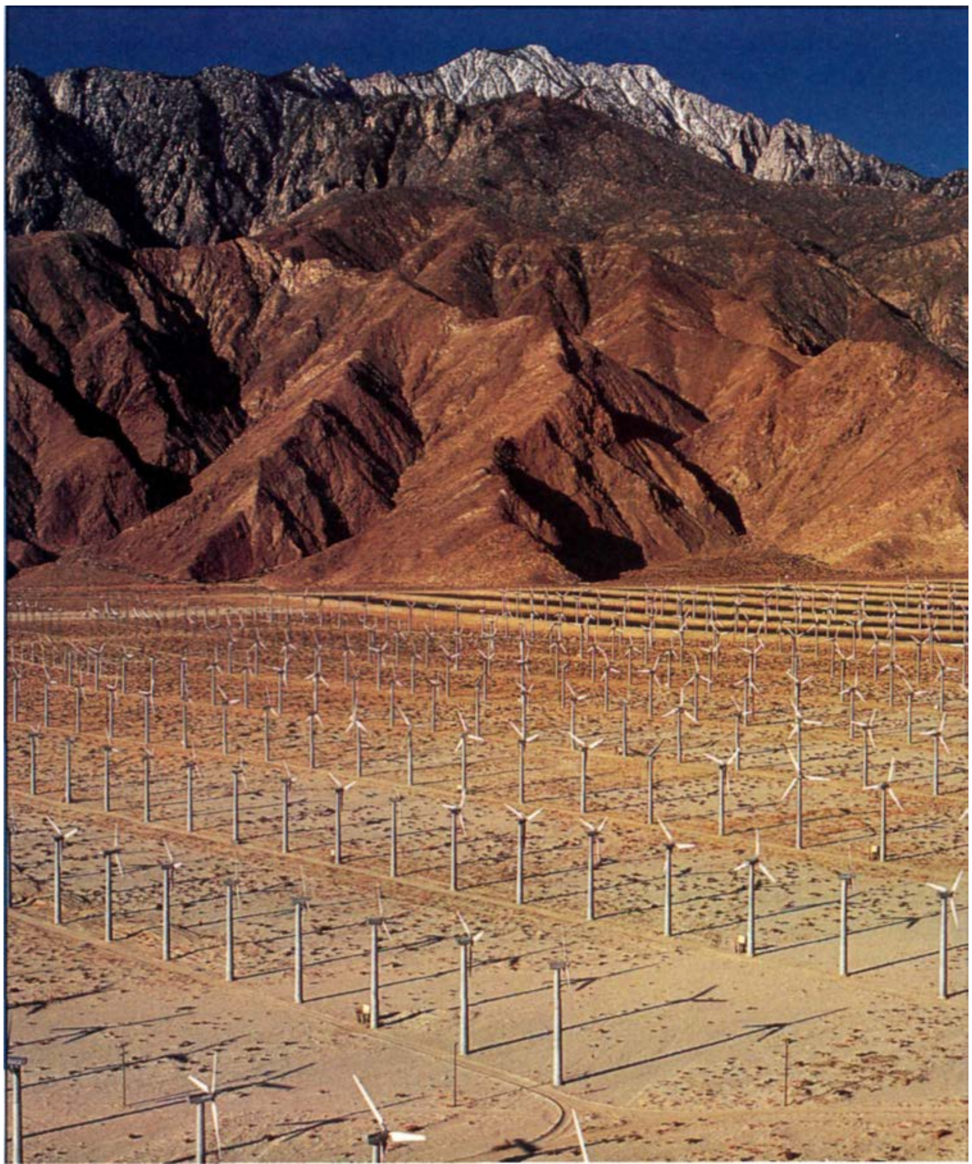
1. Why does scattering reduce the amount of incoming solar radition that reaches the Earth's surface?
2. Why is the sky blue?
3. Which of the gases in the atmosphere removes the greatest number of wavelengths from the solar radition?

Questions for Discussion

1. When Venus is viewed through a telescope, all that can be seen is a thick cover of clouds. We know that Venus does not have a hydrosphere. What explanations can you offer for the formation of Venusian clouds?
2. Would you expect the Earth's cloud cover 25,000 years ago, during the most recent ice age, to have been more extensive, less extensive, or about the same as it is today? Did precipitation during the ice age have to be different from what it is today, or could precipitation have been the same and temperature the only thing that changed?
3. When atom bombs are exploded, a huge cloud rises into the atmosphere. At some height in the troposphere, the cloud starts to spread sideways and takes on a mushroom shape. The shape is similar to that of a great thunderhead. Why do thunderheads and atomic clouds spread sideways?
4. Air pressure is usually the same in all directions. However, during the passage of a tornado or hurricane, all the windows in a house may break if they are closed. Why might this be so?

CHAPTER 13

Winds, Weather, and Deserts



A wind farm at Corgonia Pass, near Palm Springs, California. Each windmill drives a generator.

Harvesting the Wind

A hard-working human can generate just enough power in eight hours to keep a 75-watt light bulb burning. Since 75 watts is only one-tenth of a horsepower, it is hardly surprising that our ancestors sought other sources of power. They solved the problem by harnessing draft animals and by "harvesting" the wind. Long before written history, animals were being harnessed to plows, and by 8000 years ago wind-filled sails were pushing boats down rivers and around the shores of the eastern Mediterranean Sea. Many millennia were to pass, however, before the next step in the use of wind power took place. When the step came, it was a giant one: the invention of windmills provided a way for wind energy to drive machines.

The earliest written accounts on windmills, dating to about the middle of the seventh century, record their use for grinding grain in Persia (modern-day Iran). Knowledge of windmills appears to have reached Europe about A.D. 1100, and from that time onward there is a reliable record of their use. By the nineteenth century, Dutch windmills had reached such a peak of technological excellence that individual mills could generate 30,000 watts of power.

The availability of inexpensive gasoline and the spread of electricity into rural communities sent windmills into decline from the 1930s onward. However, when the price of petroleum started to rise in the 1970s, attention returned to wind power. This time the attention was directed principally to the use of windmills to generate electricity. The first windmill designed specifically to produce electricity is believed to have been built in Denmark in about 1895. In the United States, a 1,250,000-watt wind-driven turbine was installed at Grandpa's Knob in Vermont in the 1940s; this system eventually failed when one of its two propeller blades broke. Now, in the 1990s, the approach is to place hundreds of smaller turbines in wind farms located where the wind blows for long periods at high speeds. Large banks of wind-driven turbines have been installed at Altamount Pass in California, in New Hampshire, in Denmark, and elsewhere.

What is the likely future for wind power? Two problems are obvious: first, winds blow only intermittently, and so wind-driven generators need to be used in conjunction with other power sources. The second problem is location. Electricity is needed in cities, and land near cities is expensive. An additional consideration is that no one wants to live close to a wind farm because wind-driven turbines are noisy. Despite all the problems, many experts believe wind power has a great future. Many locations for wind farms do exist (offshore, for example), the efficiency is growing rapidly, and all developed countries have electrical networks that can hook wind farms to other generators of electricity. Some optimistic experts foresee a future in which the wind will supply as much as 10 percent of the world's electricity.

WHY AIR MOVES

Wind is horizontal air movement arising from differences in air pressure. Nature always moves to eliminate a pressure difference, and wind is the result when air flows from a place of high pressure to one of low pressure. Because air pressure is related to density (high pressure means the air is more compressed and therefore more dense, low pressure means less compression and lower density), horizontal movement is always associated to some degree with vertical movement.

Wind Speed

When discussing wind speed, it is necessary to separate momentary gusts from steady flow. For that reason, wind speed measurements are averaged over a specific period, commonly five minutes.

A wind speed of 20 km (12 mi/h)[1] is a pleasant breeze that rustles and moves all the leaves in a tree and produces small, white-capped waves on a lake. At 45 km/h (28 mi/h), all the branches in a tree start to sway and spray forms on open water. When wind speeds reach 65 km/h (40 mi/h), twigs and small branches break off trees, waves are high, and foam forms on wave crests. At 90 km/h (56 mi/h), trees are uprooted and the wind will knock you down; at 180 km/h (112 mi/h), it can pick you up. Fortunately, such high-speed winds are very rare. When they do occur, they tend to be associated with tornadoes or hurricanes, so that the damage is localized.

The greatest wind speed ever recorded on the surface of the Earth is 372 km/h (231 mi/h) on Mount Washington, New Hampshire, in April 1934. Speeds of 325 km/h (202 mi/h) have been recorded in hurricanes, and winds up to 335 km/h (208 mi/h) have been reported during severe storms in Greenland. Such high-speed winds can do remarkable things (Fig. 13-1). High-speed tornado winds, for example, have been reported to kill chickens and geese and also to pluck off all their feathers.

Most places around the world have wind speeds that average between 10 and 30 km/h (6 and 19 mi/h). The windiest place on the Earth is at Cape Dennison, in Antarctica, where the average speed is 70 km/h (43 mi/h). Mount Washington, winner of the crown for the highest recorded speed, averages only 55 km/h (34 mi/h) year round.

[1] Because wind was the power source for sailors for so many centuries, wind speeds are sometimes still reported in nautical units. We do not follow the practice in this book, but if the weather in your newspaper or on television use nautical units, here is how to convert them: The unit of distance at sea is the nautical mile, and the unit of speed is the knot; a knot is 1 nautical mile/hour. A nautical mile Is equal to 1.1508 land miles. To convert knots to miles/h, multiply by 1.1508; to convert knots to km/h, multiply by 1.852. Thus, 30 knots is 34.5 milcs/h or 55.6 km/h.

Figure 13.1 All that remains of a town in Florida after Hurricane Andrew passed through in 1992. All the damage seen here was caused by high speed winds.

The Windchill Factor

In places where temperatures drop below freezing, it has become customary for weather forecasters to report a **windchill factor.** This variable measures the heat loss from exposed skin as a result of the combined effects of low temperature and wind speed. Here's how it's derived.

Immediately adjacent to the human body (and also to any other solid surface) is a thin layer of still air called the *boundary layer* (Fig. 13-2). This layer is still because friction prevents movement. Heat escaping from the body must pass through the boundary layer by conduction. Because air is a poor conductor, the boundary layer serves as an effective insulator. The key to the windchill factor is that as wind speed increases, the thickness of the boundary layer decreases, thereby reducing its effectiveness as an insulator and increasing the rate at which heat is lost from the body.

Note that the air temperature does not drop as a result of the wind. What happens is that the higher the wind speed, the faster heat is lost from the skin and therefore the faster the skin temperature approaches the temperature of the air. If the skin reaches freezing temperature, frostbite ensues.

The windchill factor should correctly be called the windchill equivalent temperature because, for a given air temperature and given wind speed, the windchill factor is the air temperature at which exposed parts of the body would lose heat at the same rate if there were no wind. For example, if the air temperature is -3°C (about 27°F) and the wind speed is a sprightly 32 km/h (20 mi/h), then the windchill equivalent temperature is - 18°C (equal to 0°F). Windchill is not the only factor that affects our comfort in cold weather, but it is certainly one of the most important as far as safety is concerned.

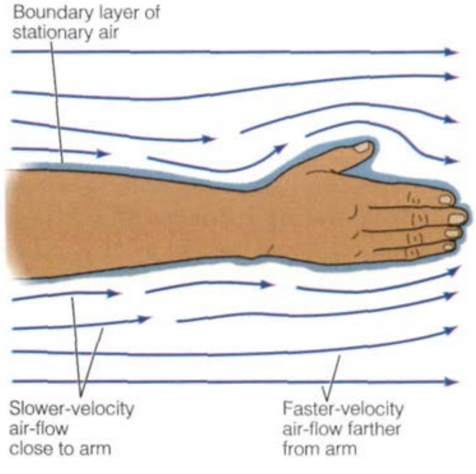


Figure 13.2 Adjacent to any solid body, such as a human arm, there is a thin layer of air held stationary by friction. Away from the body, wind speed, indicated by the length of the arrows, increases as the effects of friction become weaker and weaker.

Factors Affecting Wind Speed and Direction

If the Earth did not rotate, wind would blow in a straight line; if the Earth had a frictionless surface, the wind would flow longer and harder than it does. Neither of these ifs applies, of course, and wind is therefore controlled by the following factors.

1. The **air pressure gradient,** which is the air pressure drop/unit distance.
2. The Coriolis effect, which, as discussed in Chapter 8, is the deviation from a straight line in the path of a moving body owing to the Earth's rotation.
3. **Friction,** which is the resistance to movement when two bodies are in contact.

Air Pressure Gradient

A pressure gradient is determined from the isobars on a weather map. **Isobars** are lines on a map connecting places of equal air pressure[2] (Fig. 13.3), and the spacing between isobars determines the air-pressure gradient. When isobars are close together, the gradient is steep; when they are far apart the gradient is low. When isobars are close together, air flows rapidly down the pressure gradient and a high-speed wind is the result (Fig. 13.4).

As was mentioned earlier in this chapter, air pressure differences develop both horizontally and vertically in the atmosphere. In order to assess the effects of horizontal pressure gradients, vertical pressure differences must be avoided, and doing that requires measurements at constant altitude. Mean sea level is the elevation generally chosen. Since air pressure decreases with altitude, a vertical pressure gradient is everywhere present. In order to draw sea-level isobars on maps, air pressure measurements made at elevations higher than sea level must be corrected to the sea-level value. Once the corrections are made, isobars are drawn, generally at a contour interval of 0.4 kPa.

[2]Although the SI pressure unit is the pascal, weather maps usually report air pressure in an older unit, the millibar (Mb). To convert Mb to kilopascals (kPa), divide the Mb by 10.

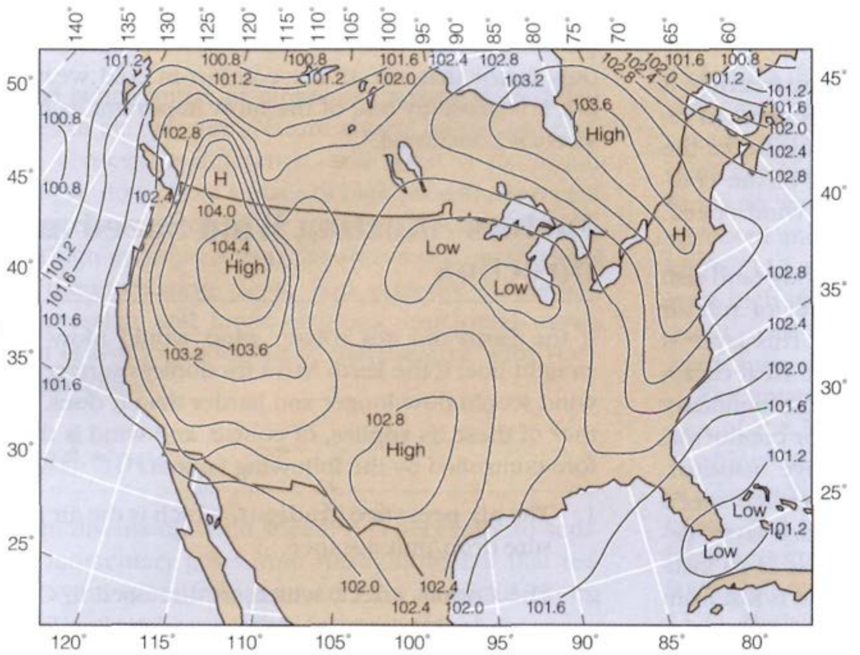


Figure 13.3 A typical air pressure map of the United States and part of Canada for November 26, 1993. The isobars are in kPa.

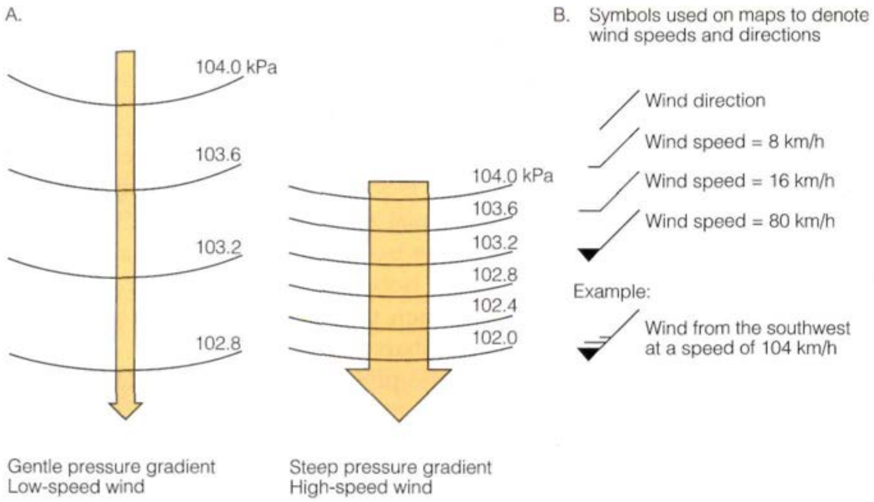


Figure 13.4 Winds and pressure gradients. A. Closely spaced isobars indicate a rapid pressure drop over a short distance and thus a steep pressure gradient; high-speed winds are the result. Widely spaced isobars indicate a low-pressure gradient and low-speed winds. B. Symbols used on weather maps to indicate wind direction and speed. The orientation of the stem indicates wind direction, and the barbs indicate speed. If more than one barb is on the wind stem, add the barbs together to get the wind speed.

Coriolis Effect

The Coriolis effect, as discussed in Chapter 8 in the section on ocean currents, influences all moving bodies. Because wind is moving air, the directions of all winds are subject to the Coriolis effect—that is, a deflection toward the right in the northern hemisphere and to the left in the southern hemisphere (see Fig. 8.8).

The speed of a moving object influences the magnitude of the Coriolis effect because a fast-moving object covers a greater distance in a given time than a slow-moving object. The longer the trajectory, the greater the change in angular velocity and therefore the greater the Coriolis deflection. Where air flow is concerned, the Coriolis effect is of greatest importance in large-scale wind systems such as the tradewinds so loved by mariners, but it is of only minor importance in small-scale, local wind systems such as thunderstorms.

Friction

When wind blows across the ground, through trees, or over solid objects of any kind, friction slows its speed. Remember that the magnitude of the Coriolis effect is proportional to the speed of a moving body. A *reduction* in speed resulting from friction will therefore reduce the Coriolis deflection, in effect causing northern hemisphere winds to turn a little to the *left* and southern hemisphere winds a little to the *right.*

Friction is important for small-scale air motions and for all winds within 1 km (0.6 mi) of the Earth's surface. It is less important for winds 1 km or higher above the Earth's surface, such as the winds of the jet stream.

Combinations of Factors

Winds are always subject to more than one factor. Consider the least complicated example, a high-altitude wind that is not in contact with the ground and therefore not affected by friction. Such a wind starts to flow because an air pressure gradient exists, and the direction of flow is down the gradient perpendicular to the isobars. Once flow starts, the Coriolis effect becomes important, and so the flowing air is deflected. Deflection means that the wind direction is no longer perpendicular to the isobars; instead, the wind must cross them at an oblique angle (Fig. 13.5). Eventually, when the pressure-gradient flow and the Coriolis deflection are in balance, the wind flows parallel to the isobars, with the low-pressure air to the left and the high-pressure air to the right.

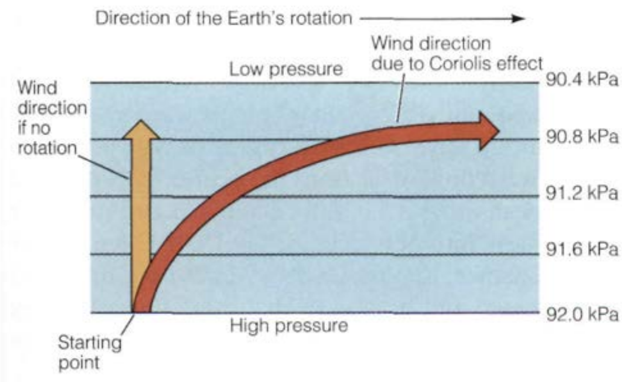


Figure 13.5 A geostrophic wind. A high-altitude wind is deflected by the Coriolis effect until a balance is reached between the direction of flow due to the pressure gradient and the direction due to the Coriolis deflection, at which point flow is parallel to the isobars.

Geostrophic Winds

Winds that result from a balance between pressure-gradient flow and the Coriolis deflection are called **geostrophic winds.** (Recall from Chapter 8, especially Figs. 8.11A and 13, that geostrophic flow also occurs in the ocean.) The daily weather maps published by the National Weather Service reveal that geostrophic winds are almost always blowing in the upper part of the troposphere. Figure 13.6 is a map of North America on which the height above sea level of the 50-kPa air pressure contours are shown. The contour heights are about 5.5 km (3.4 mi), well above any frictional effects; note that wind directions are parallel (or nearly so) to the isobars.

Within 1 km (0.6 mi) of the Earth's surface, friction complicates air flow by upsetting the balance be-

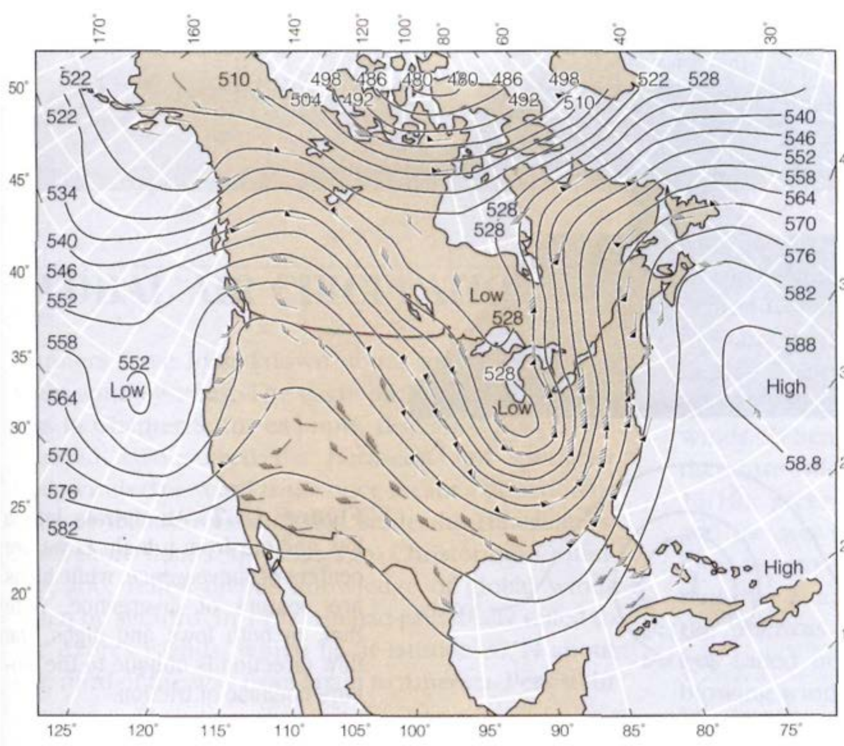


Figure 13.6 Map of North America showing upper-atmosphere winds at 7:00 a.m., November 28, 1993. The lines represent the height in meters above ground of the 50 kPa pressure surface. Note that winds are nearly all parallel to the isobars and therefore are geostrophic. Map compiled by National Weather Service.

tween pressure-gradient flow and Coriolis deflection. Friction slows the wind and thereby reduces the Coriolis deflection. Now a balance must be reached between pressure-gradient flow, Coriolis deflection, and frictional slowing. As a result, winds near the surface flow at oblique angles to the isobars. The angle size is a function of the roughness of the terrain. If the surface is very rough and the friction effect large, the angle between the air flow and the isobars can be as great as 50°. If the surface is smooth, such as the surface of the sea, the angle will be closer to 10 or 20°.

Convergent and Divergent Flow

As air near the ground flows inward from all directions toward a low-pressure center, frictional drag causes the flow direction to be across the isobars at an oblique angle. As a consequence, winds around a low-pressure center develop an *inward* spiral motion (Fig. 13-7). By the same process, air flow spirals *outward* from a high-pressure area. In the northern hemisphere, the inward-flowing low-pressure spirals rotate counterclockwise, and the high-pressure spirals rotate clockwise. In the southern hemisphere, the reverse is true.

Spiral flow was first explained by a Swedish scientist, Valfrid Ekman (1867-1954), and for this reason the spirals are sometimes called *Ekman spirals.* Ekman actually explained the spirals from his study of oceanography, as mentioned in Chapter 8, but the phenomenon, which arises from three counterbalancing effects—the pressure-gradient flow, the Coriolis effect, and friction—is the same in the atmosphere as it is in the ocean.

The spiral pattern of air flow in the lower atmosphere can be seen almost daily on the weather map and is dramatically seen in the satellite images of cloud patterns shown on television (Fig. 13.8). Air spiraling inward around a low-pressure center, which is designated **L** for **Low** on the weather map, is called a **cyclone.** Air spiraling outward, away from a high-pressure center designated **H** for **High** on the map, is called an **anticyclone.**

The inward spiral flow in a cyclone causes **convergence**, which leads to an upward flow of air at the center of the low. This upward flow leads to cloud cover and rain (Fig. 13.9A). The outward spiral flow in an anticyclone causes **divergence**, which leads to an outward flow of air from the center. This outward flow means that a high must draw high-altitude air downward into the center (Fig. 13.9B). Remember from Chapter 12 that cold air drawn downward is compressed and heated adiabatically, thus dropping the relative humidity and leading to clear, cloudless skies.

Lows tend to be associated with cloudy, unsettled weather, and highs with clear, dry weather. This is why weather forecasters always emphasize the location and movement of high- and low-pressure zones.

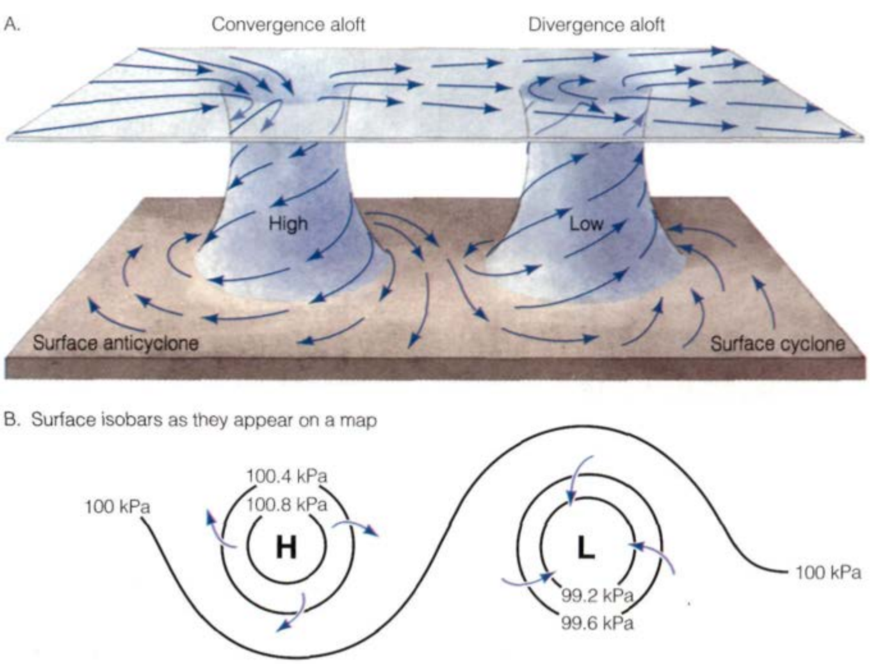


Figure 13.7 Air spirals into a low and out from a high. Lows are centers of convergence, while highs are centers of divergence. Note that, in both lows and highs, the flow direction is oblique to the isobars because of friction.

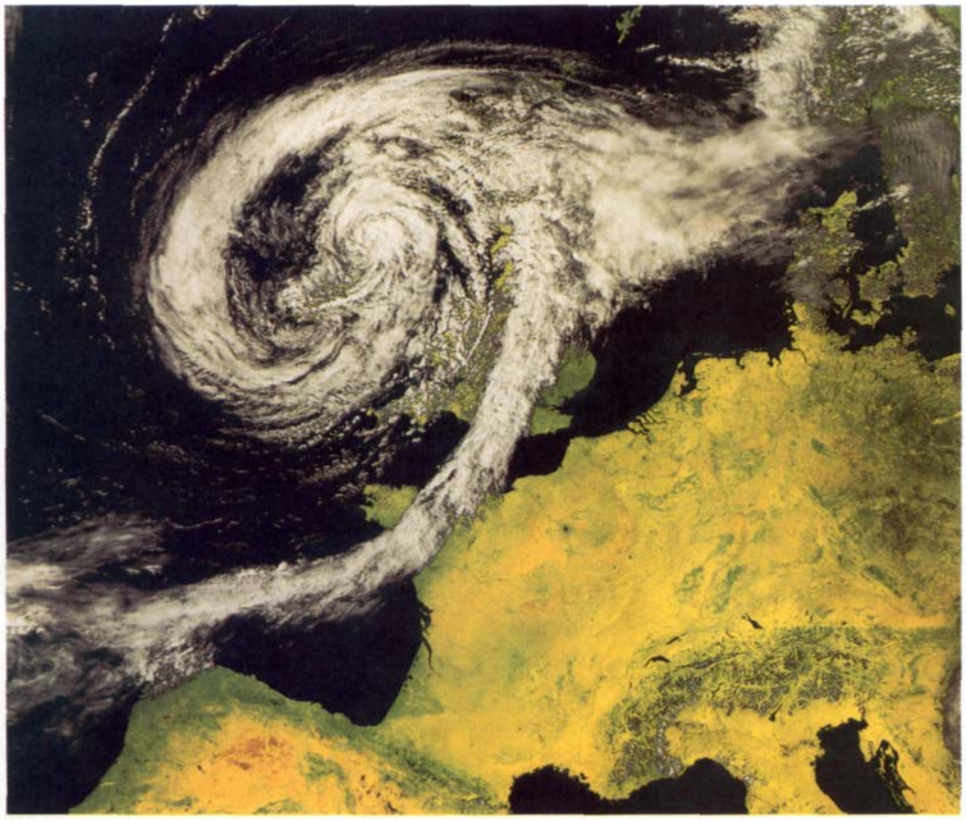


Figure 13.8 A low pressure center (cyclone) centered over Ireland and moving eastward over Europe. The counterclockwise winds of a northern hemisphere low are clearly shown by the spiral cloud pattern.

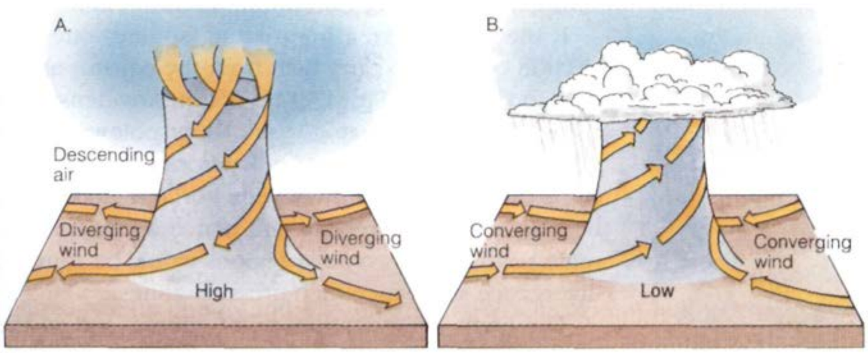


Figure 13.9 A. Convergence in a cyclone causes a rising updraft of air and with it clouds and probably precipitation. B. Divergence in an anticyclone draws in high-altitude air, creating a downdraft; clear skies and fair weather.

GLOBAL AIR CIRCULATION

Mariners have long known about and used global-scale wind systems. The essay on Polynesian navigators in Chapter 8, for example, discussed how those intrepid sailors used the northeast and southeast tradewinds (the word *trade* once meant a direction or course) to discover and then settle the Hawaiian islands more than 1300 years ago. Christopher Columbus, too, relied on his knowledge of global winds when he set forth in 1492. He had previously sailed to the Azores islands, which lie at latitude 37°N about one-third of the way from Spain to America. Persistent westerly winds slowed the trip to the Azores and prevented Columbus as well as earlier sailors from getting any farther west. From reports of Portuguese sailors, however, Columbus knew that, if he sailed south down the coast of Africa, he would find easterly winds. When he left Spain on August 3, 1492, he therefore sailed south as far as the Canary islands (Fig. 13.10), picked up the easterly blowing tradewinds, and the rest is history. With the tradewinds behind him, he crossed the Atlantic Ocean and blazed the trail that would be used by the Europeans to invade the Americas. On his return voyage to Europe, Columbus sailed northward and picked up the westerly blowing winds.

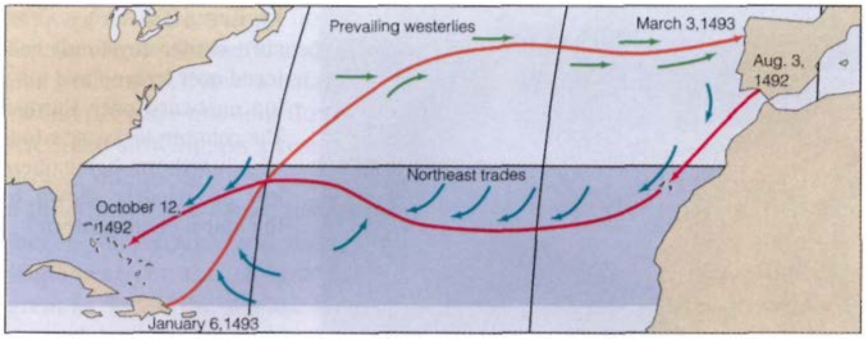


Figure 13.10 The winds used by Columbus on his first voyage to America. Outward bound after visiting the Canary islands from August f 2 to September 8, 1492, he sailed west with the northeast trades behind him. On his return voyage, Columbus sailed north to pick up the prevailing westerlies that had prevented previous European mariners from sailing any further west than the Azores. Columbus stayed in the Azores from February 15 to 24, 1493.

Hadley Cell Circulation

The person who first offered an explanation for the persistent easterly and westerly winds reported by mariners was George Hadley (1685-1768), an English mathematician. Hadley pointed out in 1735 that the underlying cause of global winds is that more of the Sun's heat reaches the surface at the equator than at the poles. The reason for the disparity of heat reaching the surface, as explained in Chapter 2, is that the Earth is round. The solar heat imbalance, Hadley pointed out, means that warm equatorial air must flow toward the pole and cold air must flow toward the equator, creating a huge convection cell.

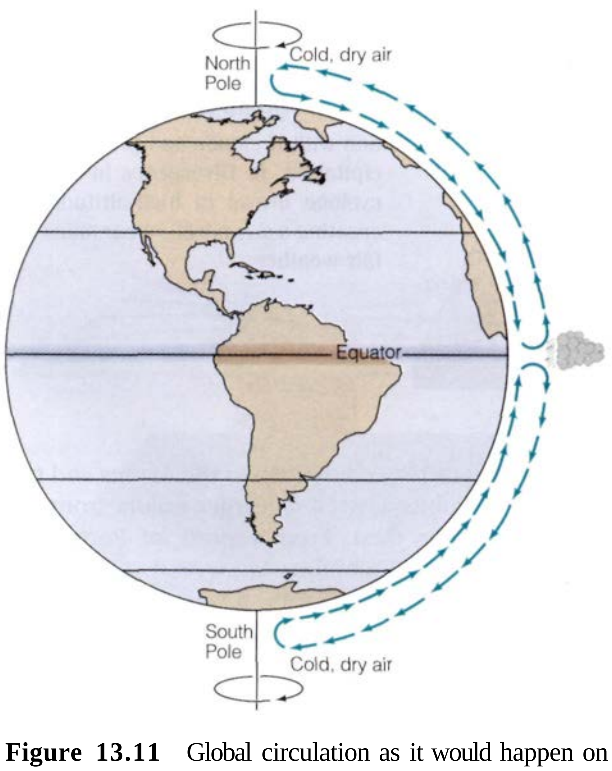


Figure 13.11 Global circulation as it would happen on a nonrotating Earth. Huge convection cells would transfer heat from equatorial regions, where the solar energy/unit area is greatest, to the poles, where the solar input is least. The equatorial region would be a zone of low pressure, while the poles would be high-pressure zones.

If the Earth were a nonrotating sphere, one convection cell would carry heat from the equator all the way to the poles (Fig. 13-11)- Warm, low-density air rising above the equator would flow poleward, and cool polar air would flow back across the surface toward the equator. Thus, the equatorial region would be a zone of convergence and hence a low-pressure region, while the two polar regions would be zones of divergence and hence high-pressure zones.

The Earth is not stationary, of course, and so the poleward air flow and the equatorward return flow are deflected as a result of the Coriolis effect. Convection does operate, but the flow is not as simple as the case described for a nonrotating Earth. On a rotating Earth, as on a nonrotating one, warm air rises in the tropics and creates a low-pressure zone of convergence called the **intertropical convergence zone**[3] (Fig. 13.12). By the time the poleward-flowing air, high in the troposphere, reaches latitudes of 30N or 30S, it has been deflected by the Coriolis effect and is a westerly geostrophic wind. Remember that a westerly wind flows to the east. Obviously, a wind that

[3]Sea-level air pressure readings in equatorial regions are generally below the standard sea-level pressure of 101.3kPa. They fall in the range 100.0 to 101.1 kPa, whereas pressures over the poles can be as high as 103.0 kPa.

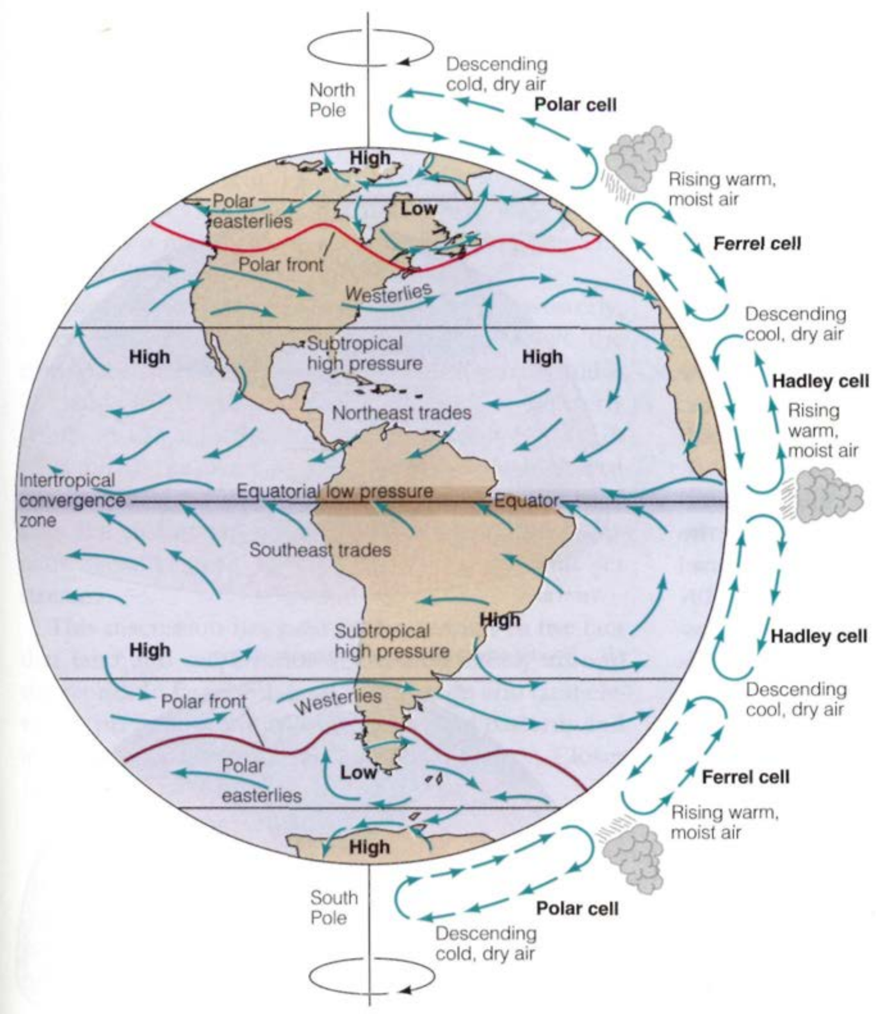


Figure 13.12 The Earth's global wind system. Moist air, heated in the warm equatorial zone, rises convectively and forms clouds that produce abundant rain. Cool, dry air descending at latitudes 20-30°N and S produces a belt of subtropical high pressure in which lie many of the world's great deserts.

flows due east cannot reach the poles, and so air tends to pile up at 30N and 30S, creating two belts of high-pressure air around the world centered approximately on those latitudes. Air in these high-pressure belts sinks back toward the surface, creating a zone of divergence. Some of the divergent air flows toward the poles, but most flows back toward the equator, creating convection cells on both sides of the equator that dominate the winds in tropical and equatorial regions; the cells, which are labeled in Figure 13.12, are called **Hadley cells** in honor of the man who explained their existence. The exact position of the intertropical convergence zone and of the two high-pressure belts varies with the seasons because the place where the Sun is directly overhead, and therefore where the Earth is receiving the greatest amount of heat, moves with the seasons.

In the Hadley cells, the high-level winds are westerlies, and the low-level winds bringing the return air toward the tropics are almost easterlies. The "almost" is necessary because friction comes into play. In the northern hemisphere, the lower-level winds are northeasterly winds, called the *northeast trades,* and in the southern hemisphere, they are the *southeast trades* (Fig. 13.12).

The Polar Front, Rossby Waves and Jet Streams

In each hemisphere, poleward of the Hadley cells, a second, middle-latitude circulation occurs. The mid-latitude cells are called *Ferret cells* after the American meteorologist, William Ferrel (1817-1891). In these middlelatitude Ferrel cells, the surface winds are westerlies because they are created, in part, by poleward flows of air from the high-pressure divergence regions at 30N and 30S. These westerlies were the winds that prevented mariners before Columbus from sailing any farther west than the Azores islands.

A third region of high-latitude circulation, called *polar cells,* lies over the polar regions. In each polar cell, cold, dry, upper air descends near the pole, creating a high-pressure area of divergence. Then air from this area moves equatorward in a surface wind system called the polar easterlies. As this air moves slowly equatorward, it encounters the middle-latitude belt of surface westerlies in the Ferrel cells. The two wind systems meet along a zone called the **polar front** and create a low-pressure zone of convergence. The polar front is a region of unstable air along which severe atmospheric disturbances occur.

The high-level winds in the polar cells are westerly. Indeed, because some of the high-level air in the Hadley cells spills over into the midlatitude Ferrel cells, the prevailing high-level winds poleward of 30N and 30S are all westerlies. Flow in the upper atmosphere is not uniform, however. Rather, the winds flow in great undulating streams called *Rossby waves* (named for Swedish scientist Carl Gustav, 1899-1951, who discovered them); these undulations resemble the meanders of streams and rivers.

Recall from Chapter 12 that the top of the troposphere (the tropopause) is much lower at high latitudes than at low latitudes (16 km, or 10 mi, at the equator and 10 km, or 6 mi, at the poles). The region where the height of the tropopause changes most rapidly is over the polar front (Fig. 13.13). A large body of cold, polar air fills the troposphere poleward of the polar front, while warmer, subtropical air fills the troposphere on the equatorial side. The tropopause is an isobar. This means that, in the stratosphere, there is a very steep pressure gradient over the polar front. Steep pressure gradients mean high-speed winds, and because friction is not involved, the winds are geostrophic. Upper-atmos-

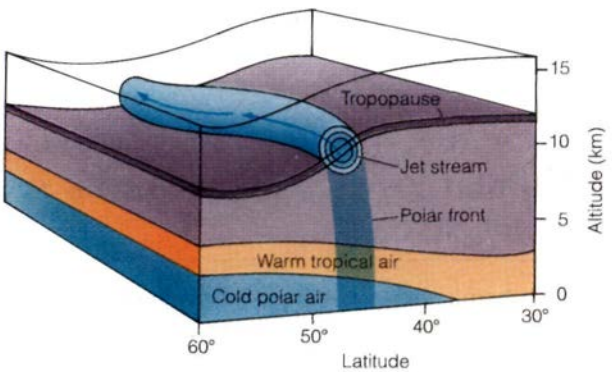


Figure 13.13 The jet stream is a high-speed westerly geostrophic wind that occurs at the top of the troposphere over the polar front where a steep pressure gradient exists between cold polar air and warm subtropical air.

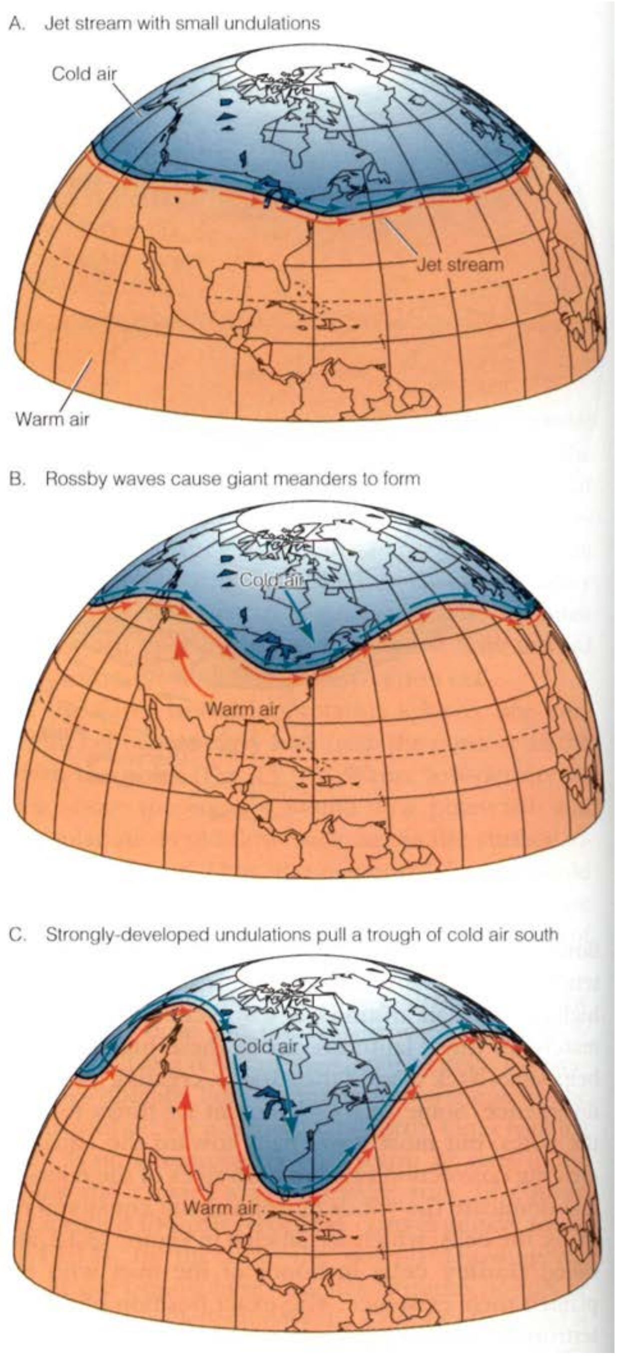


Figure 13.14 Rossby waves in the jet stream pull masses of cold air south as meanders form. A. The axis of the jet stream starts out flowing to the east in a nearly straight line. B. and C. Undulations grow into gigantic meanders that pull masses of cold polar air down over the United States.

phere westerlies associated with this steep pressure gradient, called the *polar front jet stream*, can develop exceptionally high speeds; as high as 460 km/h (286 mi/h) have been reported by high-flying planes.

Rossby waves distort the polar-front jet stream into great undulations (Fig. 13.14). As the jet stream undulates, it pushes and pulls the polar front with it and thus plays a major role in weather patterns between 45° and 60° north and south latitudes.

A second jet stream, also a geostrophic westerly, called the *subtropical jet stream*, forms above the tropopause over the Hadley cell between latitudes 20° and 30° north and south. Speeds of westerly winds in the subtropical jet stream reach 380 km/h (236 mi/h), but because the troposphere is higher at low latitudes, the subtropical jet is at a higher altitude than the polar-front jet and does not play the dominant weather role exerted by the polar-front jet stream.

This discussion has paid little attention to the fact that land and sea are not distributed evenly around the world. In fact, both land distribution and land elevation play important roles in local wind patterns and in what are called monsoon systems. (See "A Closer Look: Monsoons.")

GLOBAL PRECIPITATION AND THE DISTRIBUTION OF DESERTS

There are three global belts of high rainfall and four of low rainfall. The high-rainfall belts are the three regions of global convergence—the tropics and the two polar-front regions. The four belts of low rainfall are the regions of divergence—the two belts of subtropical highs (see Fig. 13.12) centered on latitudes 30N and 30S, and the two polar regions. The effect of the global air-circulation system is most clearly demonstrated by the distribution of deserts.

Deserts

Desert lands of various kinds total about 25 percent of the land area of the world outside the polar regions. In addition, a smaller, though still large, percentage of semidesert land exists in which the annual rainfall ranges between 250 and 500 mm (10 and 20 in). Together these desert and semidesert regions form a distinctive pattern on the world map (Fig. 13.15). The regions are not randomly scattered across the globe but instead are related to the global atmospheric circulation and to local features of the Earth's geography. In all, five types of desert are recognized (Table 13.1)-

When we compare Figure 13.15 with Figure 13.12, a relationship is immediately apparent. The most extensive deserts, the Sahara, Kalahari, Great Australian, and Rub-al-Khali, are associated with the two circumglobal belts of dry air descending on the downward-flowing limbs of the Hadley cells, centered between latitudes 20° and 30°. These and other subtropical deserts comprise one of the five recognized types of desert. They are associated with anticyclonic regions of high pressure.

A second type of desert is found in continental interiors, far from sources of moisture, where hot summers and cold winters prevail (that is, a continental-type climate). The Gobi and Takla Makan deserts of central Asia fall into this category. These deserts form because wind that travels a very long distance over land, especially land that rises up to high plateaus, eventually contains so little water vapor that hardly any is available for precipitation.

A third kind of desert is found where a mountain range creates a barrier to the flow of moist air, causing orographic lifting and heavy rains on the windward side along with a zone of low precipitation called a *rainshadow* on the downwind side (see Fig. 12.19D). The Cascade Range and Sierra Nevada of the western United States form such barriers and are responsible for desert regions lying immediately east of these mountains.

Coastal deserts constitute a fourth category. They occur locally along western margins of certain continents. The flows of surface ocean currents can cause the local upwelling of cold bottom waters. The cold, upwelling seawater cools maritime air flowing onshore, thereby decreasing its ability to hold moisture. As the air encounters the land, the small amount of moisture it holds condenses, giving rise to coastal fogs. In spite of the fog, the air contains too little moisture to generate much precipitation, and so the coastal region remains a desert. In fact, coastal deserts of this type in Peru and southwestern Africa are among the driest places on the Earth.

The four kinds of desert mentioned thus far are all hot deserts where rainfall is low and summer temperatures are high. In the fifth category are vast deserts of the polar regions where precipitation is also ex-

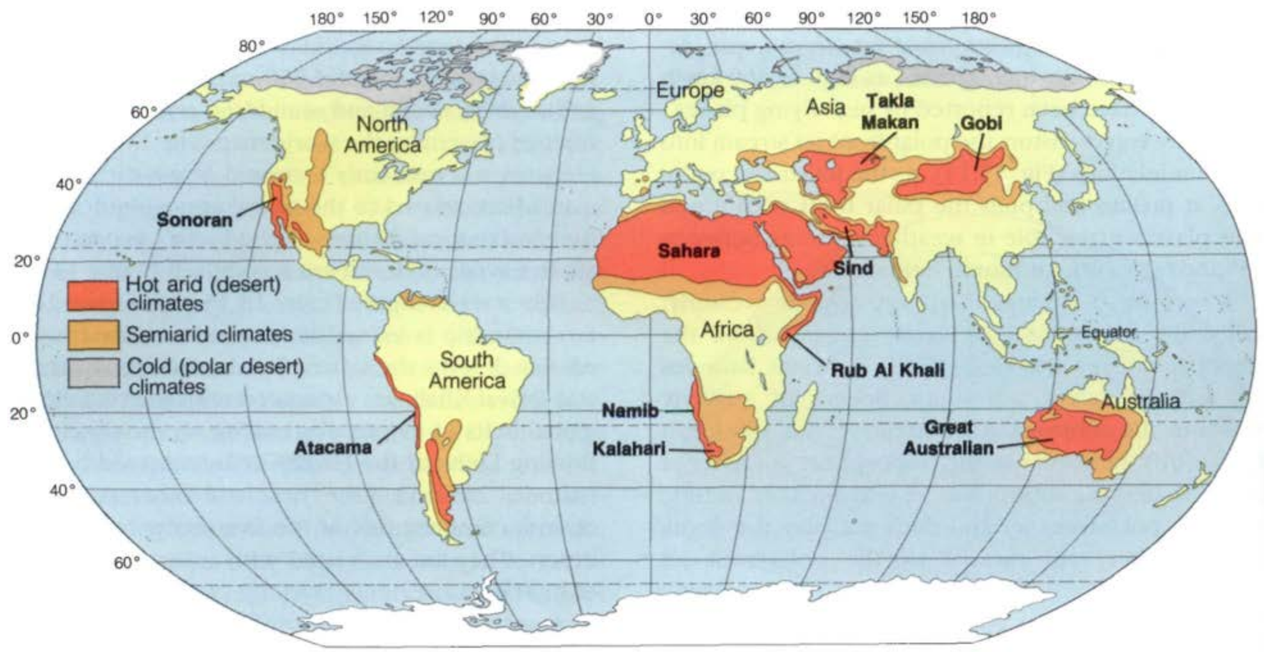


Figure 13.15 Arid and semi-arid climates of the world and the major deserts associated with them. The very dry areas of the polar regions are polar deserts.

Table 13.1
Main Types of Deserts and Their Origins

Desert Type	Origin	Examples
Subtropical	Centered in belts of descending dry air at 20-30° north and south latitude	Sahara, Sind, Kalahari, Great Australian
Continental	In contental interiors, far from moisture sources	Gobi, Takla Makan
Rainshadow	On the sheltered side of mountain barriers that trap moist air flowing from oceans	Deserts on the sheltered sides of Sierra Nevada, Cascades, and Andes
Coastal	Continental margins where cold, upwelling marine water cools maritime air flowing onshore	Coastal Peru and southwestern Africa
Polar	In regions where cold, dry air descends, creating very little precipitation	Northern Greenland, ice-free areas of Antarctica

tremely low owing to the sinking of cold, dry air. Remember that polar regions are high-pressure areas where cold, high-altitude air descends from the upper troposphere. However, cold deserts differ from hot deserts in one important respect: the surface of a polar desert, unlike those of warmer latitudes, is often underlain by abundant H_2O, nearly all in the form of ice. This ice accumulates, even though precipitation is exceedingly low, because the precipitation is always in the form of snow and the snow doesn't melt. Even in midsummer, with the Sun above the horizon *24* hours a day, the air temperature may remain below freezing. Polar deserts are found in Greenland, arctic Canada, and Antarctica. Such deserts are considered

A Closer Look

Monsoons

Monsoonal circulation is characteristic of regions where local conditions bring about a seasonal reversal of the direction of surface winds. The places on the Earth where this phenomenon is most distinct are Asia and Africa, although a weak monsoon develops over eastern and central North America, too.

The Asian monsoon is the most distinct. Because the equator lies just south of the tip of India, the normal surface-wind pattern is a northeast trade blowing offshore from India into the Indian Ocean. For half a year during the winter months, the expected northeasterly wind pattern is observed because a high-pressure anticyclone sits over the high, cold plateau of central Asia while the low-pressure intertropical convergence zone lies south of the equator, where the Sun is overhead (Fig. C13.1). The winter months are therefore a time of cool, dry, cloudless days and northeast winds. For more than 2000 years, Arab sailors have used these northeasterly winds to sail home from India.

During the summer months, the Asian wind pattern is reversed. With the Sun overhead in the northern hemisphere, the intertropical convergence zone shifts north of the equator, and so the landmass of Asia heats up and is covered by low-pressure cyclones. The winds now blow southwesterly, from the Indian Ocean on to the land. Summer months are therefore a time of hot, humid weather and torrential rains. The summer monsoons start in southern India and Sri Lanka in late May, progress to central India by mid-June, and reach China by late July. During the summer months, the Arabs sailed from Arabia to India. The sailors referred to the change in wind direction as *mausim,* Arabian for change, and from this comes our word *monsoon.*

As seen in Figure C13.1, the monsoon system that affects India also occurs in North and West Africa. There are local differences, of course, but the main controlling factor in both cases is the seasonal movement of the intertropical convergence zone.

A weak monsoon system occurs in North America. During the summer months, there is a tendency for surface winds to bring warm, moisture-laden air from the Gulf of Mexico into the central and eastern United States. Humid weather and summer rains are the result. In the winter months, the winds reverse, and there is a tendency for cold air to move southward from Canada into the Gulf.

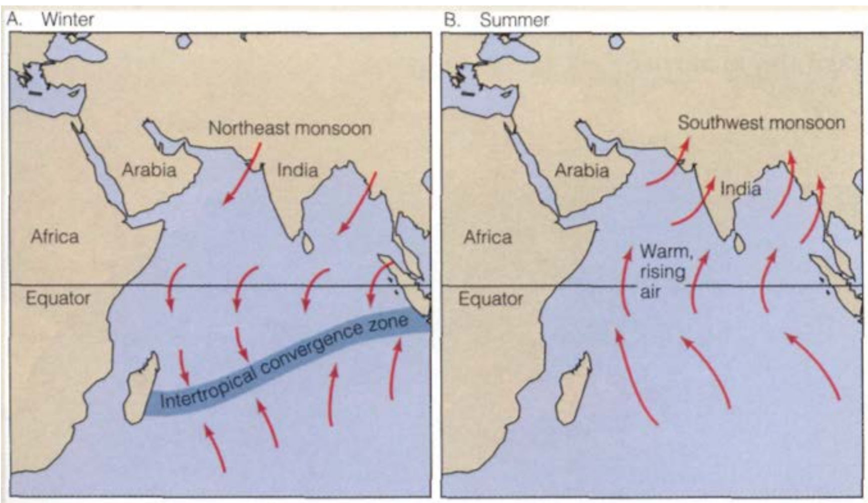


Figure C13.1 The reversing winds of the Indian monsoon. A. During the winter months when the Sun is overhead in the southern hemisphere, winds flow offshore from the northeast toward the intertropical convergence zone. Note how the winds curve toward the east as they cross the equator. B. During the summer months, the land heats up and winds flow from the southwest across Asia. When the Sun is overhead on land, the intertropical convergence zone is not a distinct band of low pressure.

to be the closest earthly analogs to the surface of Mars, where temperatures also remain below freezing and the rarefied atmosphere is extremely dry.

Dust Storms

One of the most striking features of deserts is dust storms.

As discussed both in Chapter 11 and earlier in this chapter in the section, the Windchill Factor, at the surface of an object across which wind is flowing there is a boundary layer of still air less than 1mm thick. Above the boundary layer is a thin zone in which air flow is laminar, and above the laminar air flows turbulently (Fig. 13.16). Large grains that protrude above the quiet, laminar-flow zone can be rolled and bounced along or may be swept aloft by rising turbulent winds. The larger grains that roll and bounce along mobilize

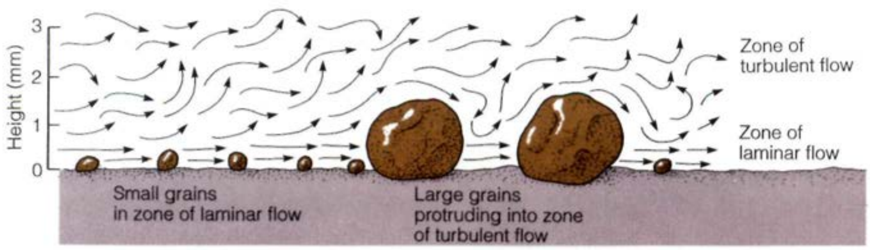


Figure 13.16 Particles of fine sand and silt at the ground lie within the boundary layer where wind speed is extremely low. As a result, it is difficult for the wind to dislodge and erode these small grains. Larger grains protrude into a zone of faster moving, turbulent air. The turbulence, which exerts a greater push on the top of the grains than does the still boundary-layer air at their base, starts the grains moving.

fine sediment, which is then carried upward by the turbulent air. In this manner dust storms start.

In a dust storm, the visibility at eye level is reduced to 1000 m (0.6 mi) or less. Such storms are most frequent in the vast arid and semi-arid regions of central Australia, western China, former Soviet Central Asia, the Middle East, and North Africa (Fig. 13.17). In the United States, blowing dust is especially common in the southern Great Plains and in the desert regions of California and Arizona.

The frequency of dust storms is commonly related to cycles of drought, with a marked rise in atmospheric dust concentration coinciding with severe drought. The frequency also has risen with increasing agricultural activity, especially in semi-arid lands. An example of how human activities can contribute to an increase in dustiness is seen in records from the western desert of Egypt in the 1930s and 1940s: the number of dust storms rose from three or four per year before the Second World War to more than 40 between 1939 and 1941, when tank action and artillery bombardment were at a peak, and then declined to four per year after military activity ceased.

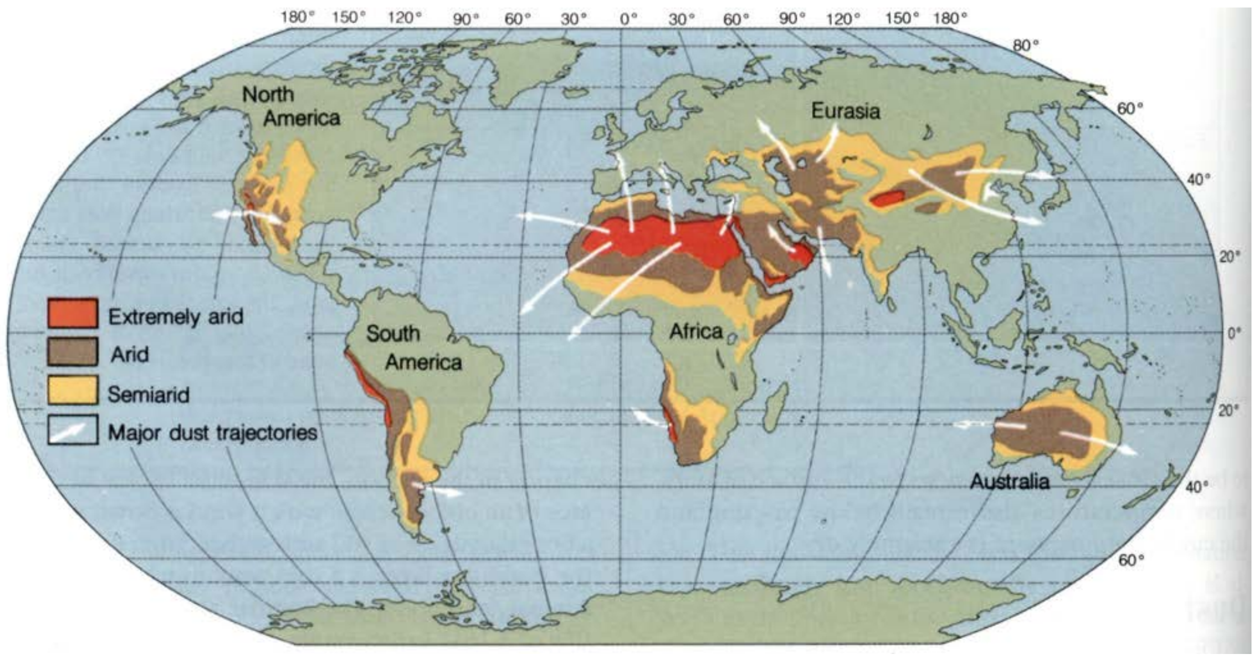


Figure 13.17 Major dust storms are most frequent in arid and semi-arid regions that are concentrated in the areas of subtropical high-pressure belts north and south of the equatorial zone. Arrows show the most common trajectories of dust transported during major storms.

LOCAL WIND SYSTEMS

In many localities, local winds are more important than global winds. Local winds, which may flow for tens or hundreds of kilometers rather than the thousands of kilometers involved in global winds, are the result of the local terrain.

Sea and Land Breezes

The least complicated example of a local wind system is the coupled land breeze and sea breeze that is familiar to anyone who lives on or near a coast. The origin of these breezes is illustrated in Figure 13.18. During the day the land heats up more rapidly than the sea, and the heated land causes the air in contact with it to heat up and expand. A pressure gradient develops, and the lower air layer flows toward the land, creating a *sea breeze.* Higher in the atmosphere, an upper-level reverse flow sets in; the coupled flows—rising air over the land and sinking air over the sea—form a convection cell.

During the night, heat is radiated more rapidly from the land than from the sea, and consequently the situation reverses. The sea is now warmer than the land, and air moves from the land to the sea, creating a *land breeze.*

Mountain and Valley Winds

Mountain winds and valley winds have a daily alternation of air flow in the same way that land and sea breezes do. During the day, the mountain slopes are heated by the Sun, and so air flows from the valley upward over the slopes. At night, the mountain slopes cool quickly, and so the flow reverses with air flowing from the mountainsides down into the valleys. Just as in the case of the land and sea breezes, moun-

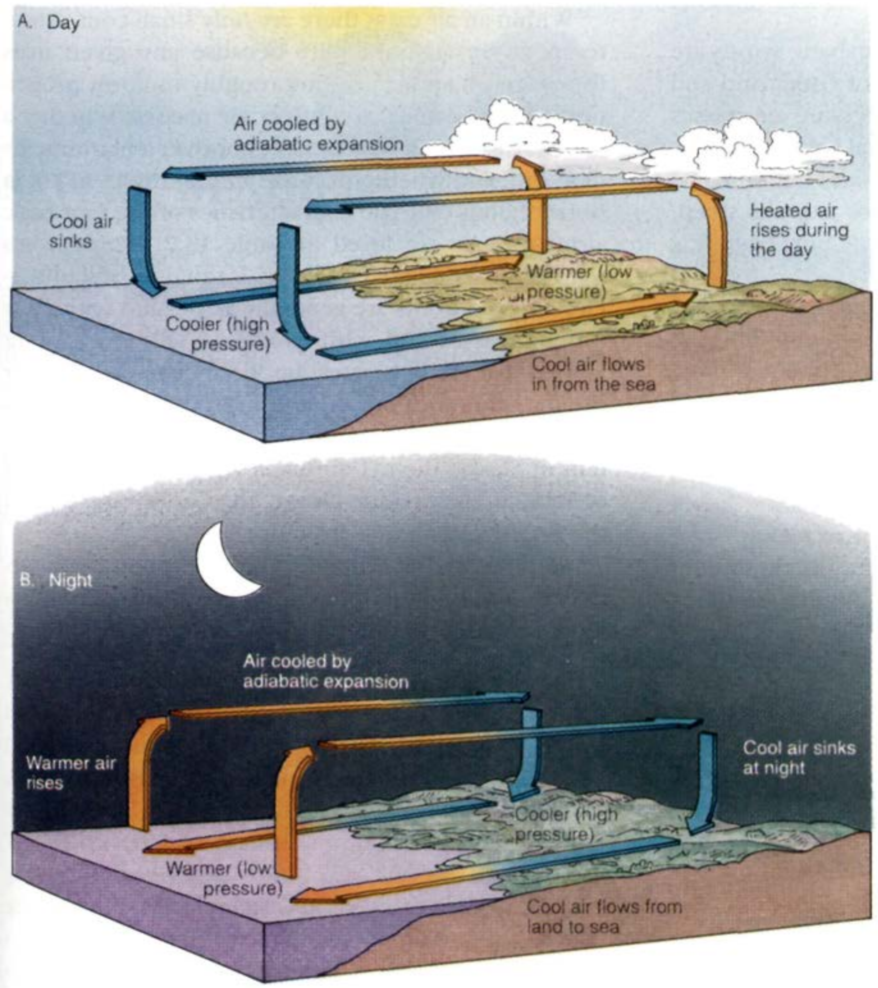


Figure 13.18 Land and sea breezes. A. During the day, the land heats up more rapidly than does the sea. Air rises over the land, creating a low-pressure area. Cooler air flows in to this area from the sea, creating a sea breeze. B. During the night, the land cools more rapidly than the sea, and the reverse flow, a land breeze, occurs.

tain and valley winds respond to localized pressure gradients set up by heating and cooling of the lower air layer.

Katabatic Winds

The flow of cold, dense air under the influence of gravity is called a *katabatic wind.* Such winds occur in places where a mass of cold air accumulates over a high plateau or in a high valley in the interior of a mountain range. As the cold air accumulates, some eventually spills over a low pass or divide and flows down valleys onto the adjacent lowlands as a high-speed, cold wind.

Katabatic winds occur in most mountainous regions around the world and commonly have local names. The *mistral* is a notable example: it is a cold, dry wind that flows down the Rhone Valley in France, past Marseilles, and out onto the Mediterranean Sea. Another notable example is the *bora,* a northeasterly that rushes down from the cold highlands of Yugoslavia to the Adriatic Sea near Trieste. Wind gusts in Trieste during a bora can reach speeds of 150 km/h (93 mi/h).

The most striking examples of katabatic winds are those that occur around the edges of Greenland and Antarctica, where the frigid, high-pressure air masses that accumulate above the continental ice sheets pour down the sloping margins of the ice and out onto the adjacent ocean waters. When the ice slope is steep, the katabatic wind speed can be terrifyingly high. It is because of a katabatic wind that Cape Dennison in Antarctica has a higher annual average wind speed than any other place on Earth.

Chinooks

Related to katabatic winds is another class of down-slope land winds called by various local names—*Chinook* along the eastern slopes of the Rocky Mountains, *John* in Germany, and *Santa Ana* in southern California. For simplicity, we speak now only of chinooks, but of course all we say is true of these winds regardless of the name being used. Chinooks are warm, dry winds. Because warm, dry air has a low density and so does not sink naturally, a chinook must be forced downward by large-scale wind and air pressure patterns. This forcing happens when strong regional winds, commonly associated with anticyclones, rise and compress higher level air masses as they pass over a mountain range and then are forced to flow down on the downwind side by the pressure of higher level air. The result is that the downward-flowing air is adiabatically heated and therefore dry-in short, a chinook.

AIR-MASS TYPES

People who dwell in the middle latitudes know that weather patterns generally last several days. The reason is that weather is controlled by huge air masses up to 2000 km (1240 mi) across and several kilometers high. Such an air mass requires several days to cross a continent.

Within an air mass there are only small contrasts of temperature and humidity because any given mass forms over a surface having roughly uniform properties. Four variables may affect air masses: whether a mass forms over a continent (c) or over a maritime region (m), and whether it forms in the tropics (T) or in polar regions (P). The characteristics of the four basic airmass types are listed in Table 13.2. Warm fronts tend to be associated with mT air masses and anticyclones. Cold fronts are generally associated with cP or mP air masses and with cyclones. The kinds of air masses that are of greatest importance to the weather of North America are cP and mT. The cP air masses originate in Canada, in the Arctic, and to a lesser extent in Alaska; the mT air masses originate in the Gulf of Mexico, the Atlantic Ocean, the Caribbean Sea, and the Pacific Ocean (Fig. 13.19).

As discussed in Chapter 12, the boundaries between air masses of different temperature and humidity, and therefore different density, are called fronts.

Table 13.2
Characteristics of Air Masses

Origin of Air	Temperature	Humidity
Continental polar (cP)	Cold	Low
Maritime polar (mP)	Cool	High
Continental tropical (cT)	Hot	High
Maritime tropical (mT)	Warm	High

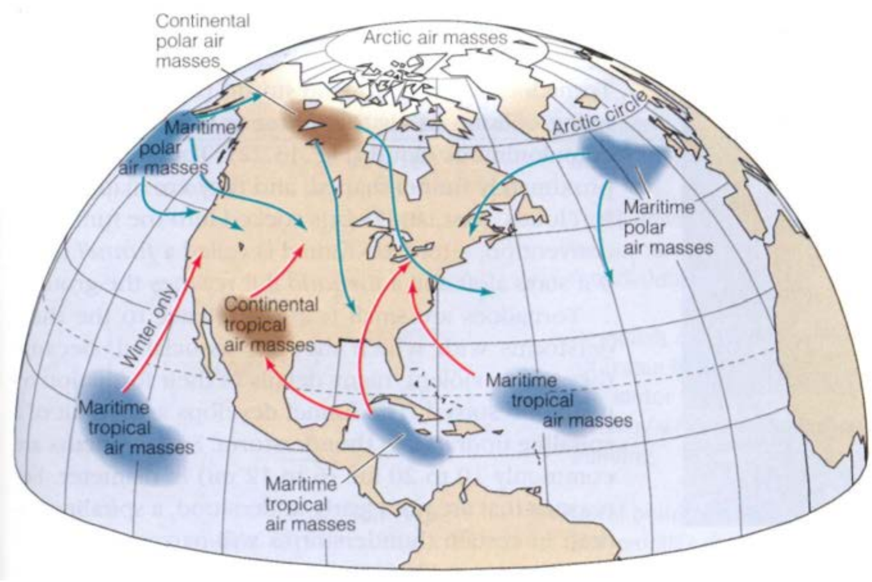


Figure 13.19 Sources of the air masses that control the weather of North America.

The boundaries are between 10 and 150 km (6 and 93 mi) in width and mark the active advance of one air mass into another.

SEVERE WEATHER

"Severe weather" and "storm" mean the same thing—a violent disturbance of the atmosphere attended by strong winds and commonly rain, snow, hail, sleet, thunder, and lightning. Severe weather can have many causes, but most occurs along cold fronts. Three kinds of severe weather—thunderstorms, tornadoes, and hurricanes—cause so much damage, and even loss of life, that they deserve special attention in our coverage.

Thunderstorms

Thunderstorms develop when an updraft of warm, humid air releases a lot of latent heat very quickly and becomes unstable. Most thunderstorms in North America form along cold fronts and are associated with mT air masses formed over the Gulf of Mexico. The released heat causes stronger updrafts, which pull in more warm, moist air, and so the process grows and the updraft intensifies. Cumulonimbus clouds form, and heavy rainfall, commonly hail, thunder, and lightning, are the result.

The towering masses of cumulonimbus clouds associated with thunderstorms can reach as high as 18 km (11 mi) (Fig. 13.20), and winds in a thunderhead can exceed 100 km/h (60 mi/h). Updrafts in a thunderhead can be so strong that large hailstones can form by coalescence of tiny ice particles and be held aloft until a sudden downdraft deposits them on the ground (Fig. 13.21).

Lightning and *thunder* accompany each other and are due to electrical charges. The electrical charges

Figure 13.20 A thunderstorm over Tucson, Arizona. Note the dark cumulonimbus clouds, the dense rain and the lightning in the clouds.

Figure 13.21 Hail falling from a thunderstorm, Santa Cruz County, Arizona.

form during the growth of a cumulonimbus cloud. The turbulent movement of precipitation inside the cloud causes particles in the upper part to become positively charged and particles in the lower part to become negatively charged. Exactly how the charge builds up is not clearly understood, but the buildup can reach hundreds of millions of volts. The charges can be released by a lightning strike either to the ground or to another cloud. As the lightning strike passes, it heats the surrounding air so rapidly that the air expands explosively and we hear the effect as thunder.

Tornadoes

Tornadoes are violent wind storms produced by a spiraling column of air that extends downward from a cumulonimbus cloud (Fig. 13.22). Tornadoes are approximately funnel-shaped, and they are made visible by clouds, dust, and debris sucked into the funnel. By convention, a tornado funnel is called a *funnel cloud* if it stays aloft and a *tornado* if it reaches the ground.

Tornadoes are small features relative to the thunderstorms with which they are associated. Because they are so violent, many details of their formation remain unresolved. The funnel develops as a result of a spiraling updraft in a thunderstorm. Such updrafts are commonly 10 to 20 km (6 to 12 mi) in diameter. For reasons that are not clearly understood, a spiraling updraft in certain thunderstorms will narrow and spiral down to a tornado funnel from 0.1 to 1.5 km (0.06 to 0.9 mi) in diameter.

Fortunately, most tornadoes are not especially strong and do not cause much damage. At the other end of the strength scale, there are some tornadoes that completely destroy any object in their path. The strength of a tornado can be estimated from the damage it causes by referring to the F-scale, named for Professor T. Theodore Fujita of the University of Chicago, who devised it (Table 13-3).

An average of 780 tornadoes strike each year in the United States. They can occur in all states and during any month of the year. However, there is a distinct intense period of tornado activity from April to August, with a peak in May. Because the severe thunderstorms that are the parents of tornadoes form along cold

Figure 13.22 A tornado crossing the plains of North Dakota.

Table 13.3
The F-Scale for Tornado Intensity

F-Scale	Category	Estimated Wind Speed, km/h	Damage
0	weak	65-118	Minor. Twigs broken.
1	weak	119-181	Trees down, mobile homes moved off foundation.
2	strong	182-253	Demolish mobile home; roof off frame houses.
3	strong	254-332	Lift motor vehicles. Destroy well-constructed building.
4	violent	333-419	Level buildings, toss automobiles around.
5	violent	420-513	Lift and toss around houses.

fronts and because the most violent cold fronts are those associated with cP air from the Canadian Arctic and mT air from the Gulf of Mexico, most tornadoes occur in the midcontinent states because that is where the two air masses are most likely to meet. (See the "Guest Essay" at the end of the chapter for more discussion of tracking tornadoes.)

Hurricanes

Hurricanes are violent, oceanic cyclones that, by definition, have maximum wind speeds in excess of 119 km/h (74 mi/h). (See Fig. 13.23.) Hurricanes are particularly devastating to island settlements and coastal regions. Once a hurricane leaves the ocean and moves onshore, wind speeds diminish and the hurricane quickly dies down. For this reason, most hurricane damage occurs within 250 km (155 mi) of the coast.

Besides wind damage, two other hurricane effects can be devastating. A *storm surge* is a local, exceptional flood of ocean water. The center of a hurricane is a region of very low air pressure—below 92 kPa in the greatest hurricanes—and a drop in air pressure raises local sea level. In the eye of a great hurricane, sea level may be 8 or 9 m (9 or 10 yd) above normal, and when hurricane-force winds drive such high seas onshore, extensive flooding results. The second effect associated with hurricanes is rain. Torrential rains and consequential flooding are associated with most hurricanes—rainfalls of 25 cm (10 in) are not uncommon—and even after wind speeds have dropped below hurricane force, violent rainstorms can continue.

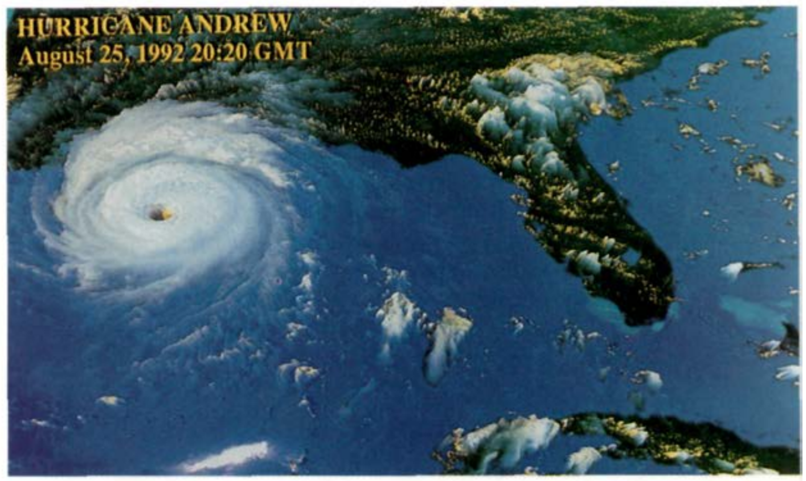
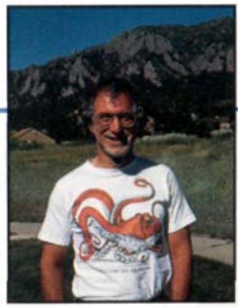


Figure 13.23 Hurricane Andrew, one of the largest and strongest hurricanes in modern times, spawned over the Atlantic Ocean and slammed into Florida in August 1992. The hurricane, here photographed from above, packed winds in excess of 200km/h.

Hurricanes start as cyclones over warm ocean water. Experience has shown that they require a sea-surface temperature of at least 26.5°C (80°F). This temperature is required because the energy source that sustains a hurricane is warm water—the water evaporates and subsequently condenses in the hurricane, releasing latent heat. Hurricanes die over land or over bodies of cold water because they no longer have an energy source to sustain them. In as much as hurricanes develop from cyclones, they can be spawned only in latitudes where the Coriolis effect is strong enough for cyclonic circulation to develop—

Guest Essay

Tracking Tornadoes Through the Southern Plains of the United States



Howard B. Bluestein is from the Boston area. He holds B.S. and M.S. degrees in electrical engineering, and M.S. and Ph.D. degrees in meteorology from the Massachusettes Institute of Technology. He is currently professor of meteorology at the University of Oklahoma, where he teaches both undergraduate courses, and does research on severe storms, tornadoes, mesoscale and synoptic meteorology, and tropical cyclones.

Tornadoes have been described as one of the last frontiers of meteorology. Relatively little is known about them because they are so difficult to study. Since they are only 100 m in diameter and usually last for less than half an hour, tornadoes affect very small areas of the Earth for very short periods of time and are extremely difficult to predict. The impact of their damage, however, can be enormous.

Prior to the 1970s, what we knew about tornadoes came mainly from serendipitous observations by nonmeteorologists. In the early 1970s, my predecessors at the University of Oklahoma and the National Severe Storms Laboratory in Norman, Oklahoma, began to increase our knowledge of tornado structure and behavior by setting out to intercept ("chase") tornadoes and tornado-producing thunderstorms, significantly increasing the number of observations.

Tornado interception works as follows. A forecast is made of where tornadoes are expected to occur, and meteorologists drive to within a 300-km radius of home base to the general area where the parent storms might form. If storms do form, the meteorologists decide which storm has the most tornado-producing potential and position themselves at a safe distance from the portion of the storm where tornadoes typically occur. At this distance (3 to 6 km), cloud features can still be seen clearly. In supercell storms (long-lived, rotating solitary storms), meteorologists look for tornadoes near the wall cloud, a rotating, lowered cloud base that is located near the rear (with respect to storm motion) of the storm. In nonsupercell storms, meteorologists look near the cloud base of rapidly building cloud towers in growing thunderstorms. We guess which storm has the most tornado-producing potential by combining theory with observational experience: it is both a science and an art to "pick" the tornadic storm from the zoo of storms out in the field.

I arrived in Norman in 1976, when meterologists were beginning to evaluate the usefulness of Doppler radar for issuing severe weather warnings to the public. Conventional radar can assess only the intensity of the precipitation in a storm by measuring the backscattered radiation from raindrops, ice crystals, and hail. Doppler radar, however, can also reveal features of the storm's

that is, higher than about latitude 5°. Hurricanes cannot form along the equator. The necessary conditions for hurricane breeding are found in only a few places around the world (Fig. 13.24). Note, however, that, although the phenomenon is everywhere the same, hurricanes are called typhoons in the western Pacific; in northern Australia, cyclone, the general term for convergent, spiraling airflow, is also used for hurricanes.

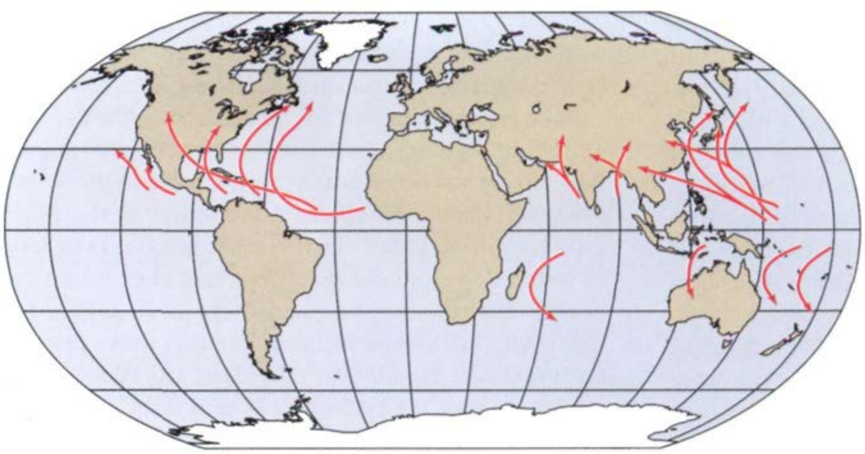


Figure 13.24 Hurricanes form in those places in the world where the right conditions of ocean-water temperature and the Coriolis effect occur. Arrows show the usual directions followed by hurricanes once they form.

wind field by measuring the shift in frequency of the backscattered radiation. At Norman we would observe a storm visually and correlate the wind "signature" seen by the Doppler radar with what we saw. After many observations, it became possible to devise the technique currently being used by the National Weather Service to issue tornado warnings up to 20 or 30 minutes prior to the touchdown of a tornado within a supercell storm.

Our visual observations also proved useful for training storm spotters. These individuals watch approaching storms and provide the National Weather Service with information about whether a tornado is actually observed and whether a rotating wall cloud, a precursor to tornado development, is present.

As we became more proficient at intercepting tornadic storms, it became apparent to me that we should attempt to obtain more quantitative measurements. The wind, pressure, temperature, and moisture distributions within tornadoes were not known. From 1980 to 1983, we attempted to make these measurements, using TOTO (the Totable Tornado Observatory, not Dorothy's pet), an instrumented device designed at the Wave Propagation Laboratory in Boulder, Colorado. Our goal was to place TOTO directly in the path of a tornado, and then retrieve it and analyze the recorded data. Although we obtained data under wall clouds and near tornados, we found that it was too difficult to place TOTO directly in the path of a tornado: usually the tornado dissipated or changed direction before hitting TOTO, or there were no roads leading to the path of the tornado. We were also concerned that a strong tornado might damage TOTO.

From 1984 to 1989, we used a commercially available portable radiosonde (instrumented weather balloon) to learn about the vertical distribution of temperature and moisture inside and outside of severe storms. This information was especially useful for scientists who simulate the life history of severe storms on a computer and need to know atmospheric conditions that precede the development of the storm.

In 1987 we first began to use a portable 3-cm wavelength Doppler radar designed at the Los Alamos National Laboratory. Using the radar, we could position ourselves at a safe distance from a tornado and make wind measurements, without having to get directly in its path.

Unfortunately, 1987 and 1988 were among the "worst" storm seasons we had ever seen in the Southern Plains: there were very few tornadoes to make measurements in! It really was not until 1990 and 1991 that we successfully collected data. On April 26, 1991, we made measurements of wind speeds as high as 120 to 125 m/s in a large tornado in north-central Oklahoma. Prior to this, wind speeds this high were only inferred from damage assessments or from photogrammetric analysis of debris and cloud tags in tornado movies.

Tracking tornadoes and making scientific measurements in them is an exhilarating, challenging, and often frustrating endeavor. On the average, we intercept tornadoes on only one out of nine chases. We drive great distances and often see nothing, or we miss a tornado that occurred just minutes earlier, or we intercept a tornado and have an instrument fail. However, when we do have a successful chase, we have the opportunity to witness one of nature's most awesome displays of power and to unravel its mysteries.

Summary

1. Wind results from the horizontal movement of air in response to differences in air pressure.
2. The windchill factor results from wind reducing the insulating effect of the boundary layer of stationary air adjacent to the skin.
3. Wind speeds and wind directions are controlled by air pressure gradients, the Coriolis effect, and friction.
4. Air pressure gradients can be determined from a weather map by measuring the distance between isobars, which are lines connecting places of equal air pressure at sea level. When isobars are close together, the pressure gradient is steep and winds are strong.
5. The Coriolis effect, which arises as a result of the Earth's rotation, deflects wind toward the right in the northern hemisphere and to the left in the southern hemisphere.
6. The magnitude of the Coriolis effect is a function of latitude and wind speed. The effect is zero at the equator and a maximum at the poles. Wind speed contributes to the Coriolis deflection because, at high speed, a body moves a long distance in a short time. The longer the trajectory, the greater the deflection.
7. Friction between air and the ground slows winds and therefore reduces the Coriolis effect.
8. High-altitude geostrophic winds are Coriolis deflected and eventually flow parallel to isobars, with the low pressure on the left and the high pressure on the right.
9. Friction causes spiral air flow directed in toward

a low-pressure area. Such lows are called cyclones and rotate clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. The opposite flow occurs as air spirals out from a high-pressure area called an anticyclone. The spiral direction of an anticyclone is counterclockwise in the northern hemisphere and clockwise in the southern hemisphere.

10. Inward air flow in a cyclone produces a low air pressure zone of convergence; outward air flow in an anticyclone produces high air pressure and is a zone of divergence.
11. The global air-circulation pattern arises from a combination of two factors—the flow of air from the equator toward the poles in response to a thermal imbalance and the Coriolis effect.
12. At the top of the troposphere, along steep pressure gradients formed above the polar fronts, are westerly geostrophic winds called the polar-front jet streams. Similar subtropical jet streams occur above the descending limbs of the Hadley cells.
13. Around the equator is a region of low pressure caused by rising currents of warm, humid air. Centered on latitude 30°N and S are two belts of high pressure owing to descending, low-humidity air. The world's major deserts are located in these belts. The cells of rising moist air and descending dry air are called Hadley cells.
14. Deserts form in four ways: as a result of the global air circulation, in continental interiors far from sources of moisture, in rainshadows, and along coasts adjacent to cold-upwelling seawater.
15. Local wind systems, such as land and sea breezes, mountain and valley winds, katabatic winds, and chinooks, arise from local terrain effects. Such winds are often of much greater importance locally than global winds.
16. Thunderstorms form along cold fronts, as a result of updrafts of warm, humid air, and are maintained by the latent heat of condensation from the humid air.
17. Tornadoes, which are violent, upward-spiraling columns of air associated with cumulonimbus clouds, form as a result of spiral updrafts in certain thunderstorms. Many aspects of their formation remain uncertain.
18. Hurricanes are violent oceanic cyclones in which maximum wind speeds exceed 119 km/h. Because hurricanes are oceanic phenomena, they cause their greatest damage to island and coastal regions.

Important Terms to Remember

air-pressure gradient (p. 339)
anticyclone (p. 342)
convergence (p. 342)
cyclone (p. 342)
divergence (p. 342)
friction (p. 339)
geostrophic wind (p. 345)
Hadley cell (p. 345)
high(H) (p. 342)
intertropical convergence zone (p. 344)
isobar (p. 339)
Iowa) (p. 342)
polar front (p. 346)
wind (p. 338)
Windchill factor (p. 339)

Questions for Review

1. Explain why air density and air pressure are related.
2. Why do weather reporters give the windchill factor during cold weather but not during warm weather? Could windchill cause you harm if the air temperature was 15°C?
3. Name the three factors that control the speed and direction in which wind flows and briefly explain how each factor works.
4. On a weather map on which the isobar contour interval is 0.4 kPa, one region has the contours 20 km apart while another has them 200 km apart. In which region would you experience the stronger winds?
5. Explain why the Coriolis deflection of wind direction is always to the right in the northern hemisphere and always to the left in the southern hemisphere.
6. What are geostrophic winds and how do they arise? Name a well-known geostrophic wind.
7. How do cyclones and anticyclones form? What is the relationship between the highs and lows

marked on a weather map and cyclones and anticyclones?

8. What kind of weather tends to be associated with cyclones?
9. What is a Hadley cell and how does it form? Describe the relationship between the intertropical convergence zone and the Hadley cells.
10. How do the tradewinds arise?
11. Use a drawing to illustrate the global surface wind systems.
12. What is the polar front? Is it a zone of convergence or divergence? How does the polar-front jet stream form?
13. Why are the world's major desert regions centered between latitudes 20° and 30°?
14. List three ways deserts can form other than as a result of the global air circulation.
15. What is the boundary layer, and what role does it play in the formation of dust storms?
16. How and why do land and sea breezes occur?
17. What are katabatic winds? Give an example of a well-known katabatic wind.
18. List the four major categories of air masses. Which kinds of air masses are of greatest importance in weather development in North America?
19- What is a cold front and how does it differ from a warm front? Describe the kind of weather you might expect as a warm front advances into the area in which you live.
20. Briefly describe how a thunderstorm forms. Where does its energy come from?
21. What is the relationship between thunderstorms and tornadoes? Why are tornadoes most frequent in the central part of the United States?

Questions for A Closer Look

1. What is monsoonal circulation and where on the Earth does it occur?
2. Describe the Asian monsoon and explain how it was used by Arab sailors to sail back and forth between India and Arabia.
3. How and where does the North American monsoon occur?

Questions for Discussion

1. Discuss in general terms the criteria needed for successful wind farms. Are the criteria most likely to be met by local or by global wind conditions?
2. About 250 million years ago, the continents of today were grouped together in a supercontinent called Pangaea. The site of New York City at that time was in the center of the supercontinent, thousands of kilometers from the sea. The latitude of the future New York was about 15°N. What was the weather like?
3. What distributions of continents and oceans would effectively stop the formation of hurricanes? What distribution would make their formation rate even more frequent than it is today? Do some research and see if there have been times in the past when your predicted positions have occurred.

CHAPTER 14

The Earth's Climate System

The mummified body of a prehistoric man, exposed by retreat of a glacier high in the Tyrolean Alps, provides important clues about changing climate during the past 5000 years.

The Tyrolean Iceman

In the late summer of 1991, a remarkable discovery was made by a pair of German trekkers high in the Tyrolean Alps. The mummified body of a prehistoric man was seen protruding from slowly melting ice near the margin of Similaun Glacier at 3200 m (10,500 ft) altitude. With the corpse were a fur robe, woven grass cape, leather shoes, flint dagger, copper ax, wooden bow, and 14 arrows. Radiocarbon dating of the man's skin and bone indicated that he died about 5300 years ago. His antiquity, together with the associated artifacts and his bodily characteristics, showed him to be a member of the Late Neolithic and Bronze Age population of south-central Europe. The man was judged to be between 25 and 35 years old and about 1.6 m (5.2 ft) tall. He apparently had remained frozen more than five millennia until progressive thinning of the glacier since the middle nineteenth century eventually led to his exhumation.

The frozen corpse and the implements and clothing found with it proved to be a treasure trove for archaeologists. Prior to this discovery, scientists could only speculate about many aspects of Neolithic life on the basis of limited artifacts found at scattered sites. Now they had an actual person, complete with the tools of everyday life.

This discovery also was important to scientists studying climatic change, for it lent support to the interpretation of climate history in central Europe that had been pieced together from various lines of evidence. In the high Alpine valleys, receding glaciers have disgorged the remains of trees and bog vegetation that once flourished at sites subsequently covered by ice for many millennia. Dated by radiocarbon, the fossil trees and other plant remains tell us of an interval of mild climate during the early to middle Holocene Epoch (ca. 9500-5500 years ago). This was a time, following the last ice age, when glaciers retreated high up in the mountains and plants invaded the upper slopes of alpine valleys. At the time the iceman lived, the climate was becoming cooler and glaciers were growing larger. The next 5000 years would see a succession of cool intervals, the latest of which is commonly referred to as the Little Ice Age. Apparently, none of the intervening milder climatic periods achieved temperatures equal to those of the first half of the Holocene. Thus, when the Tyrolean man died near the margin of Similaun Glacier, his remains became entombed in an alpine deep-freeze that kept him perfectly preserved for more than 5 millennia until the recent warming trend exposed him to view.

THE CLIMATE SYSTEM

In Chapter 12 we learned that climate is a measure of the average weather conditions of any place on the Earth and commonly is expressed in terms of mean temperature and mean precipitation. However, other parameters, including humidity, windiness, and cloudiness, also are important in characterizing climate, even though measurements of them are not routinely recorded in all places.

As shown in Figure 14.1, the Earth's climate system is complex and consists of several subsystems—atmosphere, ocean, cryosphere, lithosphere, and biosphere. The subsystems interact so closely that a change in one of these subsystems can lead to changes in one or more of the others. All but the lithosphere are driven by solar energy. Some of the incoming radiative energy is reflected back into space by clouds, atmospheric pollutants, ice, snow, and other reflective surfaces. The remainder is absorbed by the air, ocean, and land. Of these three energy reservoirs, the atmosphere responds most rapidly to outside influences, commonly within a month or less. The ocean surface responds more slowly (generally over months or years), whereas changes involving the deep ocean may take centuries. Although the land may respond rapidly or slowly to changes in the other components of the climate system, it has special significance on long time scales affecting the distribution of continents and ocean basins and the location and height of mountain ranges. The distribution and topography of the land directly influence the location and extent of glaciers and sea ice, as well as the character and extent of vegetation. Vegetation is impor-

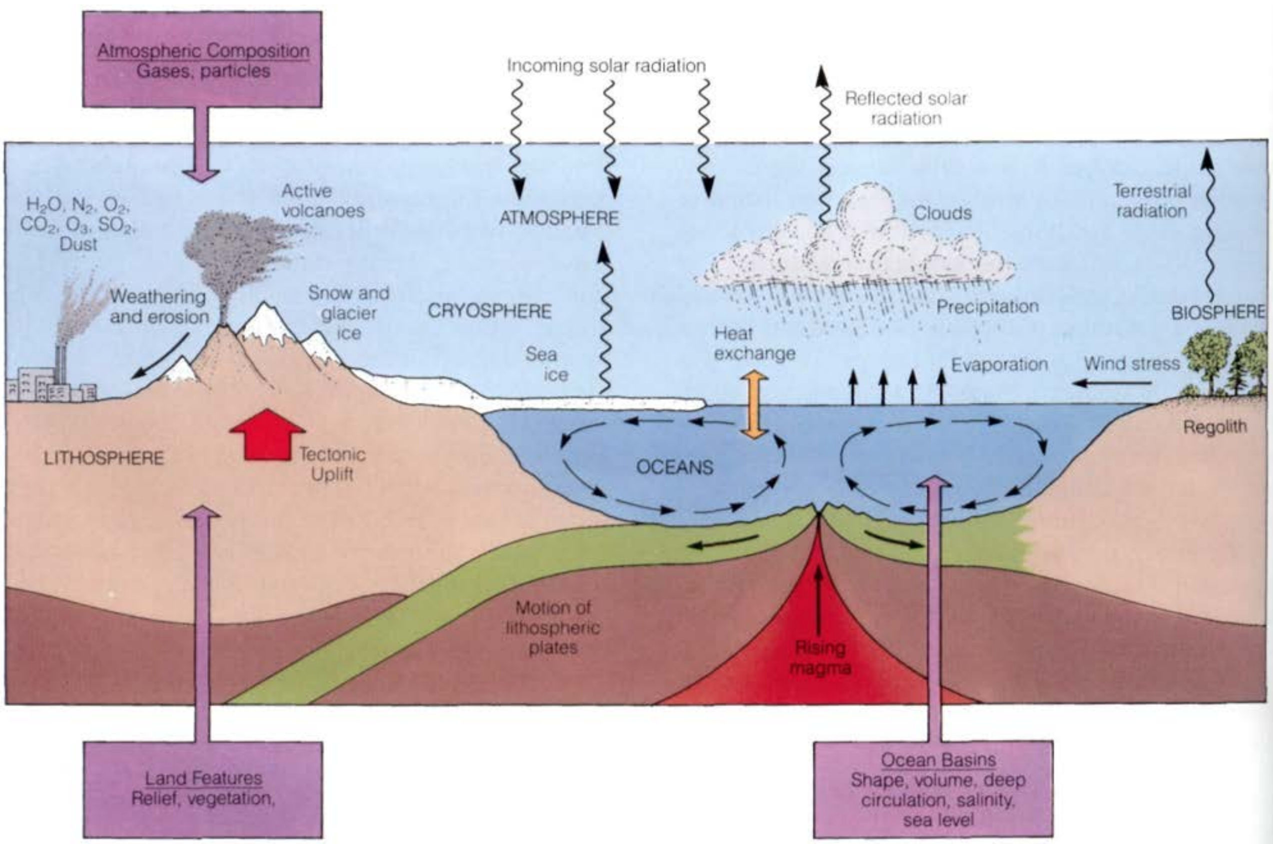


Figure 14.1 A diagrammatic representation of the Earth's climate system showing its five interacting components: lithosphere, atmosphere, oceans, cryosphere, and biosphere.

tant in the climate equation because it helps determine the reflectivity of the land surface. It also influences the composition of the air by absorbing carbon dioxide, and it affects humidity and therefore local cloud cover. Its absence can increase wind erosion, which in turn can influence climate by affecting the dustiness of the atmosphere.

Because of its ability to absorb and retain heat, the ocean serves as a great reservoir of heat energy that helps moderate climate. The ocean's effect is clearly illustrated by the contrast between a coastal region having a maritime climate (little contrast between seasons) and an area farther inland having a continental climate (strong seasonal contrast). The world ocean also is extremely important in controlling atmospheric composition, for the ocean contains a large volume of dissolved carbon dioxide. If the balance between oceanic and atmospheric carbon dioxide reservoirs were to change by even a small amount, the radiation balance of the atmosphere would be affected, thereby bringing about a change in world climates (Chapter 18).

Understanding how the Earth's climate system works is a challenging task, and we are far from having all the answers. Important insights have been gained through the study of past climates, evidence of which is preserved in the geologic record. Such evidence offers important clues that can help tell us what causes climate to change, and how the different physical and biological systems of the Earth respond to changes of climate operating on different time scales.

In this chapter we will investigate some of the evidence demonstrating that climates have changed during Earth history, and we will see how this evidence provides clues about why climates change. In Chapter 18, we will look more closely at how human activities are changing the atmosphere in ways that might lead to a significant change of climate during our lifetime.

EVIDENCE THAT CLIMATES CHANGE

Last winter may have been colder than the winter before, and last summer may have been wetter than the previous summer, but such observations do not mean that the climate is changing. The identification of a climatic change must be based on a shift in average conditions over a span of years. Several years of abnormal weather may not mean that a change is occurring, but trends that persist for a decade or more may signal a shift to a new climatic regime.

Our experience tells us that weather changes from year to year, but because climate is based on average conditions over many years, we may not find it easy to tell if the climate is changing. A grandparent may recall that winters seemed colder half a century ago, but do such recollections actually point to a change of climate? Fortunately, weather records are kept throughout the world, and in some places they have been maintained for a century or more, long enough to see if average conditions have shifted in the past century or two.

One of the longest continuous climatic records available to us comes from Great St. Bernard Pass at the crest of the Alps, where the Augustinian friars have recorded temperatures since the 1820s and snowfall since the 1850s (Fig. 14.2). Between I860 and I960, temperature and snowfall fluctuated approximately in phase, with intervals of cool temperature corresponding to times of above-average snowfall. Although short-term trends in this record persist for about a decade or two, over the entire period of the record there has been a general trend toward warmer temperatures.

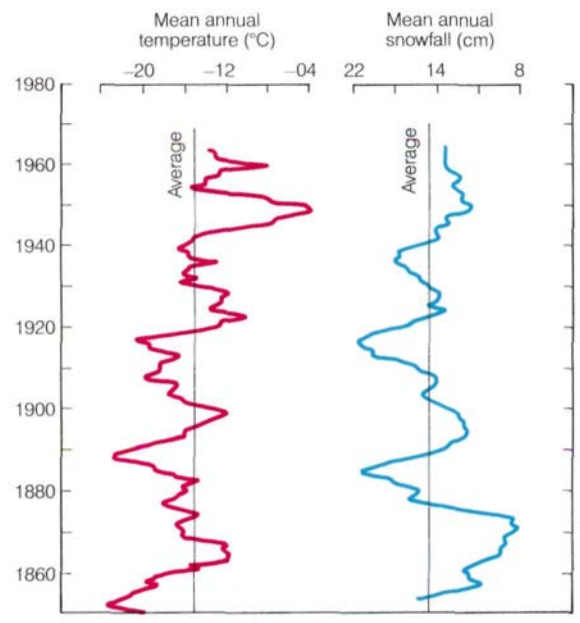


Figure 14.2 Variations of mean annual temperature and snowfall recorded at Great St. Bernard Pass on the Swiss-Italian border since the mid-nineteenth century. The vertical line through each graph shows the long-term average value.

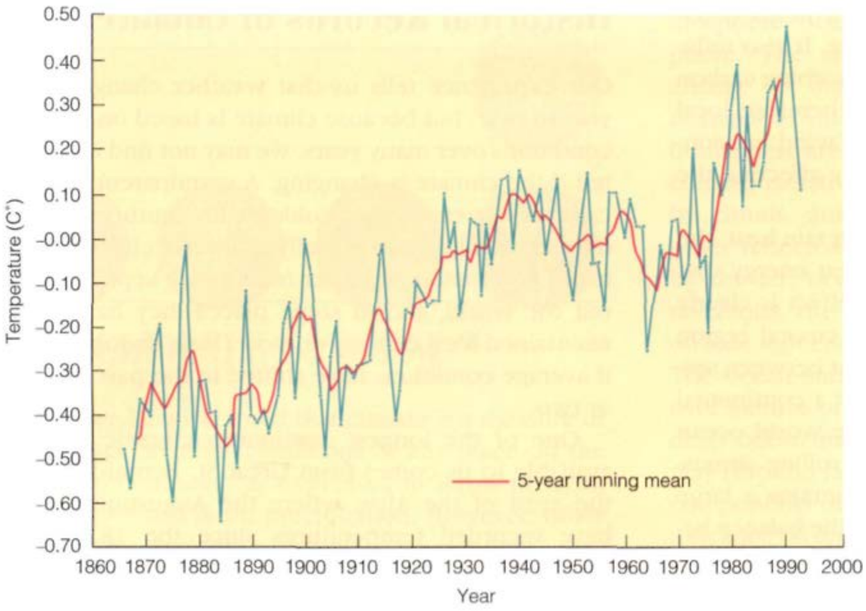


Figure 14.3 Annual mean air temperatures from 1866 to 1992 for the world's land areas. Because data from many different places are included, the annual mean temperatures are expressed as a variation from the average annual mean temperature for the 30 years from 1951 through 1980. From the late 1860s to the early 1940s, the annual mean temperature rose about 0.6°C. From the early 1940s to 1965 the temperature declined about 0.2°C and since 1965 the trend has been upward. The overall rise during the century and a quarter of record is ca. 0.85°C.

The temperature pattern in the Alps is representative of that in other parts of the northern hemisphere, where average temperature experienced a fluctuating rise after the 1880s to reach a peak in the 1940s (Fig. 14.3). Thereafter, average temperatures declined until the 1970s when they again began to rise, and in the early 1990s they reached the highest values yet recorded.

The amplitude of this recent long-term temperature increase, amounting to a little less than 1°C (1.8°F) seems small, yet its effects were seen widely, especially in high latitudes. During the six decades between 1880 and 1940, for example, mountain glaciers in most parts of the world shrank, some conspicuously (Fig. 14.4), and arctic sea ice was observed less frequently off the coast of Iceland. The biosphere also responded: between 1880 and 1940, the latitudinal limits of some plants and animals expanded slightly to-

Figure 14.4 In the late nineteenth century, Findelen Glacier in the Swiss Alps covered all the bare, rocky terrain seen here in the lower part of its valley. Since that time, the glacier terminus has retreated far upvalley in response to a general warming of the climate.

ward the poles, and an increase in the length of the summer growing season led to a general improvement in crop yields.

That the world's climates can change detectably within a human lifetime is a relatively new realization. With this realization has come increasing concern about the impact of such changes on nature and on society, as well as the possible impact of human activities on the Earth's climate (Chapter 18).

The Geologic Record of Climatic Change

The evidence of climatic change on the Earth comes largely from the geologic record. Scientists have long puzzled over the occurrence of geologic features that seem out of place in their present climatic environment. Abundant fossil bones and teeth of hippopotamus—the same kind that lives in East Africa today—

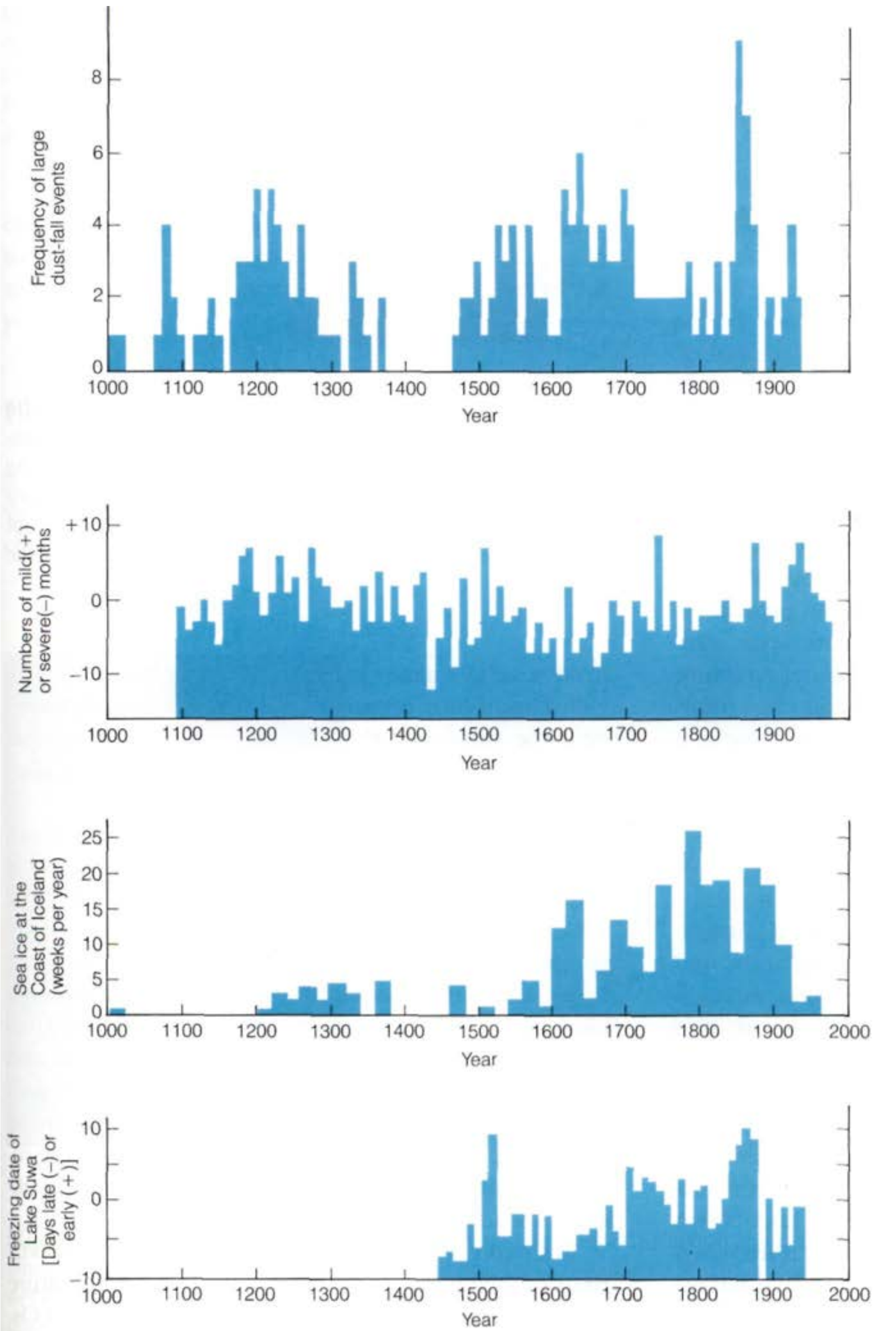


Figure 14.5 Climate proxy records spanning all or part of the last 1000 years: A. frequency of major dust-fall events in China *(Source:* After Zhang, 1982); B. severity of winters in England, recorded as the frequency of mild or severe months *(Source:* After Lamb, 1977); C. number of weeks per year during which sea ice reached the coast of Iceland *(Source:* After Lamb, 1977); D. freezing date of Lake Suwa in Japan relative to the long-term average *(Source:* After Lamb, 1966).

have been recovered from sediments in southeastern England that were deposited about 100,000 years ago under conditions that may have been like those in parts of modern tropical Africa. At many sites beyond the margin of the Great Lakes in north-central United States, plant remains have been found that show this region formerly resembled arctic landscapes like those now seen in far northern Canada. In each case, a significant change in local climate apparently has taken place, so that the biota living in these areas today is very different from the fossil forms we see preserved in the geologic record.

Besides fossils, other anomalous features tell us that the climate has changed: (1) glacial features in temperate lands, (2) desert sand dunes now covered with stabilizing vegetation, (3) beaches of extensive former lakes in dry desert basins, (4) channel systems of now-dry streams, (5) remains of dead trees above the present upper treeline in a mountainous area, and (6) surface soils with profiles that are incompatible with the present climate.

Climate Proxy Records

Scientists attempting to reconstruct former climates can use instrumental records only for the very recent past. To extend the reconstruction back in time, they must rely on records of natural events that are controlled by, and closely mimic, climate. We call these *climate proxy records,* and although lacking the precision of instrumental data, they often can provide us not only with an indication of the year-to-year variability of weather but also with a good general picture of climatic trends. Four of the longest and most informative series are shown in Figure 14.5. Others include (1) the number of severe winters in China since the sixth century A.D., (2) the height of the Nile River at Cairo since A.D. 622, (3) the quality of wine harvests in Germany since the ninth century A.D., (4) dates for the blooming of cherry trees in Kyoto, Japan, since A.D. 812, and (5) wheat prices (a reflection of climatic adversity) in England, France, the Netherlands, and northern Italy since A.D. 1200. Each of these phenomena bears a relationship to prevailing climate and therefore is regarded as a useful proxy for climatic variability.

As we learned in Chapter 10, a further source of paleoclimate (i.e., past climate) information comes from ice cores collected from polar glaciers. Measurements of the ratio of two isotopes of oxygen (^{18}O and ^{16}O) in glacier ice enable us to estimate air temperature when the snow that later was transformed into that ice accumulated at the glacier surface. Cores obtained from the Greenland and Antarctic ice sheets, as well as

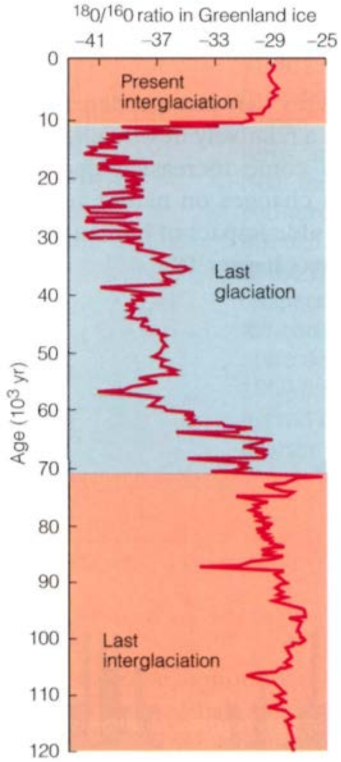


Figure 14.6 Variations in the oxygen-isotope ratio through the Greenland Ice Sheet. The zone of strong negative values beginning about 70,000 years ago and ending about 10,000 years marks the last glaciation. The sharp shift in values about 10,000 years ago marks an abrupt change from glacial to interglacial climate at the end of the glaciation.

from several smaller ice caps at lower latitudes, provide continuous records of fluctuating temperatures near the surface of these glaciers, in some cases extending back many tens of thousands of years (Fig. 14.6).

Trees offer additional important information about past climates. A tree living in middle latitudes typically adds a growth ring each year, the width and density of which reflect the local climate (Fig. 14.7). Many species live for hundreds of years; a few, like the Giant Sequoia and Bristlecone pine of the California mountains, can live thousands of years. Specialists in tree-ring analysis are able to reconstruct temperature and precipitation patterns from tree rings over broad geographic areas for any specific year in the past. These reconstructions provide both a picture of changing regional weather patterns and a picture of long-term climatic trends. (See the Guest Essay.)

In Chapter 8 we saw how scientists use the growth rings of living corals to reconstruct water temperature oscillations related to the cyclic El Niño/Southern Oscillation, an important element of tropical climate. Be-

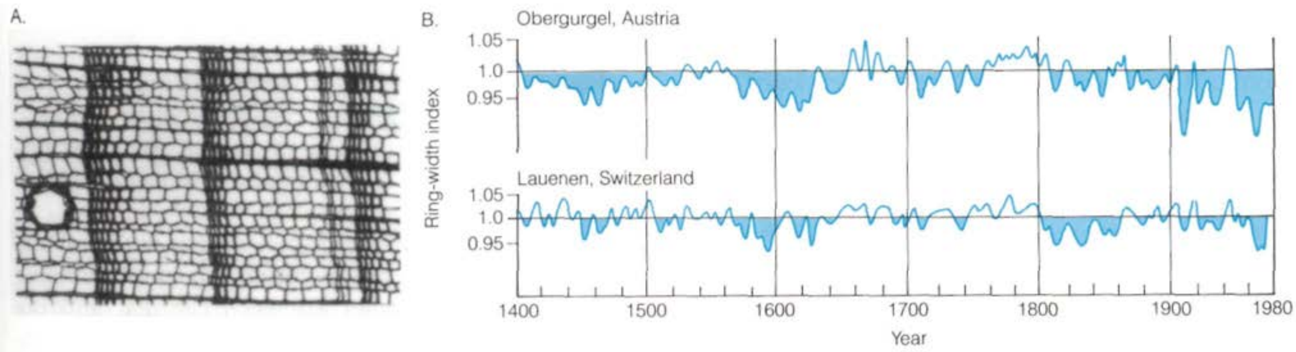


Figure 14.7 Climate and tree rings. A. Enlarged cross section of a 1500-year-old fossil larch tree found in the moraine of a Swiss glacier showing annual growth rings. Early wood of each year consists of large, well-formed cells. Late wood contains smaller, closely spaced cells. B. Tree-ring chronologies based on density measurements of spruce trees at two sites in the Alps that are 200 km apart. A general similarity can be seen among periods of low growth (shaded) that correspond with times of cold climate and glacier expansion.

cause the rings are laid down annually (Fig. C8.2A), we can obtain a high-resolution record of changes in surface water temperature spanning hundreds of years.

Climate of the Last Millennium

A wealth of historical and climate proxy records provide us with an unusually comprehensive picture of climatic variations during the last thousand years. The varied evidence from the northern hemisphere shows that an episode of relatively mild climate during the Middle Ages gave way about 700 years ago to a colder period when temperatures in Western Europe averaged 1 to 2°C (2 to 4°F) lower. This cooler climate caused the snowline to drop about 100 m (330 ft) in the world's high mountains, thereby causing glaciers to advance. Geologists refer to this interval of cooler climate and glacier advance as the Little Ice Age. Throughout much of Western Europe and adjacent islands, the Little Ice Age climate was punctuated by **un-**

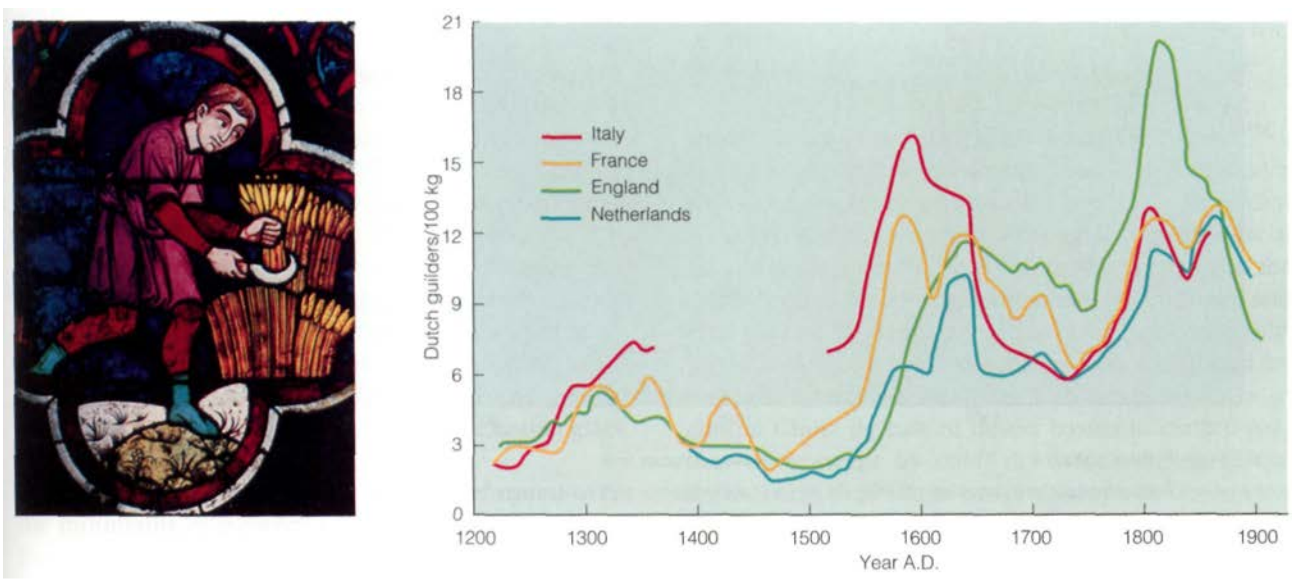


Figure 14.8 Fluctuations in the price of wheat in Western Europe from the thirteenth to nineteenth century, expressed in Dutch guilders, track the course of climate. Intervals of cool, wet climate were unfavorable for wheat production, causing the price to rise. The two largest peaks, in the early seventeenth and early nineteenth centuries, coincide with the greatest advances of glaciers in the Alps during the Little Ice Age.

usually harsh conditions marked by snowy winters and cool, wet summers, expansion of sea ice in the North Atlantic, and an increase in the frequency of violent wind storms and sea floods in mainland Europe. As summers became cooler and wetter, grain failed to ripen, wheat prices rose (Fig. 14.8), and famine became pervasive. In England the life expectancy fell by 10 years within a century.

By the early seventeenth century, advancing glaciers were overrunning farms in the Alps, Iceland, and Scandinavia. During the worst years of that century, sea ice completely surrounded Iceland, and the cod fishery in the Faeroe Islands failed because of increasing ice cover.

The 1810-1819 decade, the coldest in Europe since the seventeenth century, witnessed renewed advances of glaciers in the Alps and many other mountain ranges. Erratic weather in the nineteenth century led to further crop failures, rising grain prices, epidemics, and famines that resulted in large-scale emigrations of Europeans, especially to North America. Thus, many Canadians and Americans owe their present nationality to the vagaries of Little Ice Age climate.

Little Ice Age conditions persisted until the middle of the nineteenth century when a general warming trend caused mountain glaciers to retreat and the edge of the North Atlantic sea ice to retreat northward. Although minor fluctuations of climate have continued to take place (Fig. 18.11), the overall trend of increasing warmth in middle latitudes brought conditions that were increasingly favorable for crop production at a time when the human population was expanding rapidly and entering the industrial age.

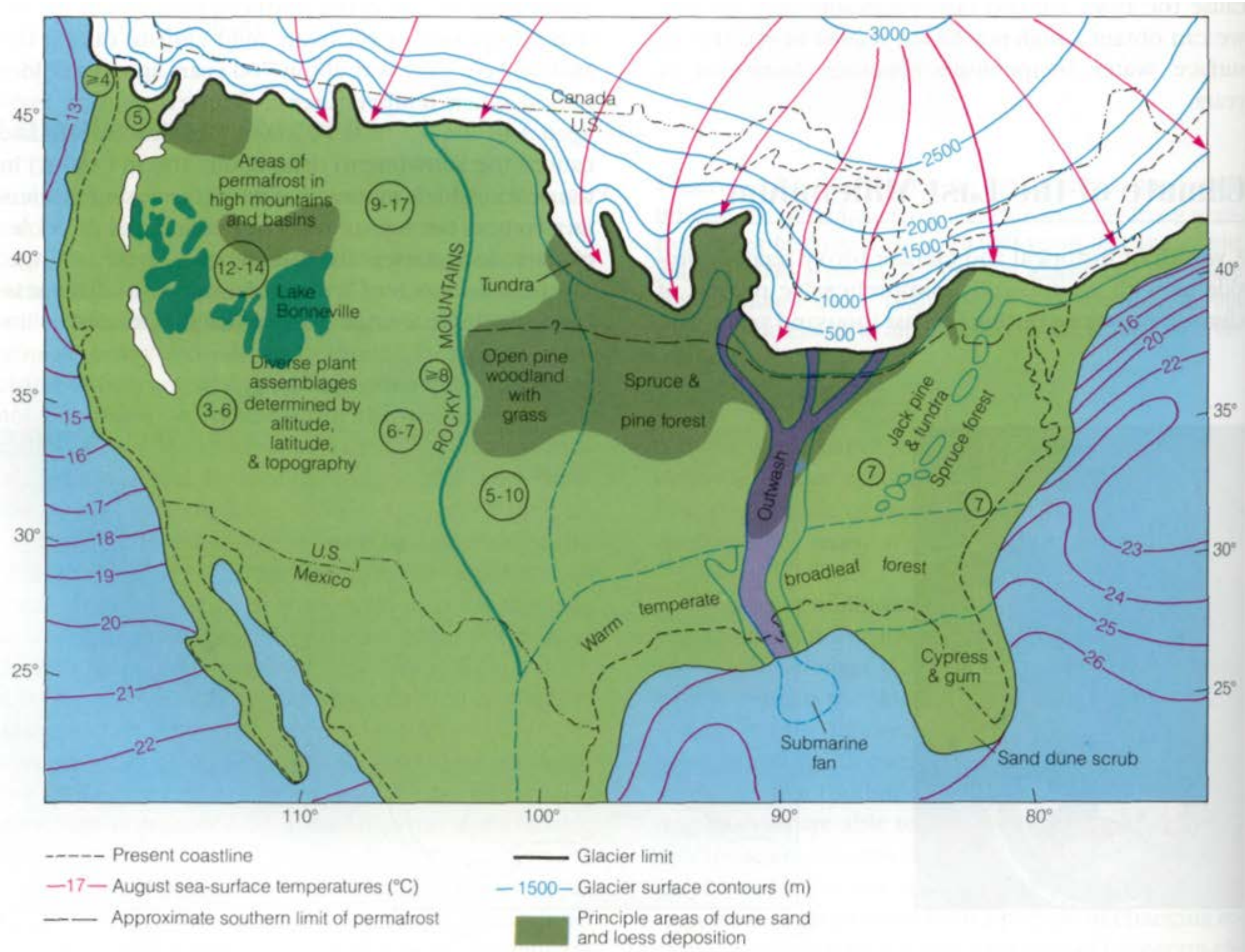


Figure 14.9 Map of central North America during the last glacial maximum, about 20,000 years ago. Coastlines lie farther seaward owing to fall of sea level of about 100 m. Sea-surface temperatures are based on analysis of microfossils obtained from deep-sea cores. Circled numbers show estimated temperature lowering, relative to present temperatures, at selected sites based on various kinds of proxy evidence.

The Last Glaciation

The last time the Earth's climate was dramatically different from what it is now was during the last **glaciation,** an interval when the Earth's global ice cover greatly exceeded that of today. The last glaciation, which culminated about 20,000 years ago, was the most recent of a long succession of glaciations, or ice ages, that characterized the Pleistocene Epoch. To reconstruct the climate of this latest ice age, earth scientists rely largely on sediments and ancient glacier ice that contain fossil and isotopic evidence of ice age conditions.

Glaciers, Permafrost, and Sea Ice

During the last glaciation, the climate of northern middle and high latitudes became so cold that a vast ice sheet formed over central and eastern Canada and expanded southward toward the United States and westward toward the Rocky Mountains (Fig. 14.9). As it moved across the Great Lakes region, the glacier overwhelmed spruce trees growing in scattered groves beyond the ice margin. Ancient logs of that period, now exposed in the sides of stream valleys, are bent and twisted, indicating that they were alive when the glacier destroyed them. Some retain their bark, and some lie pointing in the direction of ice flow, like large aligned arrows. A radiocarbon date for a log tells us the approximate time when the ice arrived and the tree was killed. The ages of buried trees discovered near the southern limit of the ice sheet tell us that the ice reached its greatest extent about 20,000 years ago. Still older wood found farther north pinpoints ages for the ice margin during its southward advance. Dividing the distance between two sample localities along a north-south transect by the difference in age of the logs at these sites yields an average rate of advance of the ice margin across that distance. Results from such pairs of dates suggest that the ice was advancing at an average rate of 25 to 100 m (82 to 330 ft) per year, a speed that is comparable to that of some existing glaciers.

Simultaneously, other great ice sheets formed over the mountains of western Canada and over northern Europe and western Asia (Fig. 14.10). As ocean water was evaporated and then deposited as snow on these growing ice sheets, the world sea level fell. The falling sea level allowed the great ice sheets of Greenland and Antarctica to grow larger as they spread across the adjacent, exposed continental shelves. Large glacier systems also formed in the Alps, Andes, Himalaya, and Rockies, and smaller glaciers developed on numerous other ranges and isolated peaks scattered widely through all latitudes.

We assume that ice shelves also existed under full-glacial conditions, but their size and distribution are not easy to determine. Some geologists postulate that an ice shelf may have covered all of the Arctic Ocean and extended south into the northern reaches of the Atlantic Ocean, thereby linking the major northern ice sheets into a continuous glacier system that covered nearly all of the arctic and much of the subarctic regions of the planet. Other geologists concede that ice shelves very likely were present in favorable places, just as they are today around Antarctica, but suggest that the polar sea was largely covered by much thinner sea ice that extended far south of its present limit into the North Atlantic.

With the southward spread of ice sheets on the northern continents, periglacial zones were displaced to lower latitudes and lower altitudes. In Russia permafrost extended 1000 km (620 mi) or more south of the ice margin. However, in North America evidence of full-glacial permafrost is restricted largely to Alaska, to a narrow belt adjacent to the southernmost limit of the ice sheet in the northern Great Plains and Great Lakes regions, and to the high mountains of the American West, especially the Rockies. The contrast may largely reflect the fact that, whereas the massive Eurasian glacier lay north of 50° latitude, the ice sheet over central North America extended south of 40° into more temperate latitudes. The periglacial zone was therefore much narrower in the United States because the north-to-south gradient of climate there was far steeper.

The Dusty Ice-Age Atmosphere

At the height of the glacial age, the middle latitudes were both windier and dustier than they are today. We infer this from several lines of evidence. Glacial age loess deposits found south of the ice limit in the midwestern United States become both thinner and finer east of major glacial meltwater channels, implying that the dust was picked up and distributed by strong westerly winds. The thick loess deposits of central China lie east of desert basins in central Asia that were swept by cold, dry winds during glacial times. Loess deposits in eastern Europe lie downwind from extensive meltwater sediments lying between the Alps and the southern limit of the great north European ice sheet, and they contain fossil plants and animals consistent with cold, dry conditions, implying that loess deposition is characteristic of glacial times. In each of these regions, successive sheets of loess are separated by soils formed during **interglaciations,**

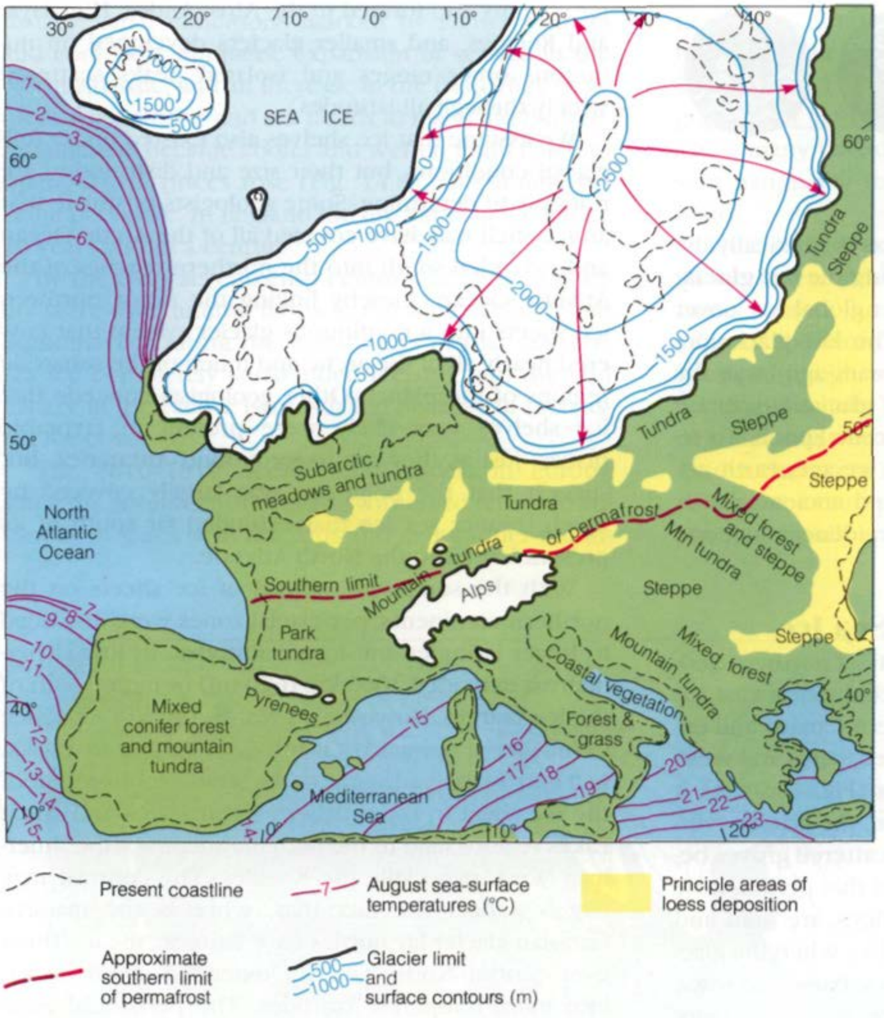


Figure 14.10 Geography of Western Europe about 20,000 years ago during the last glaciation. According to this reconstruction, a vast continental ice sheet extended across the North Sea between Britain and Scandinavia and was separated from the glacier-covered Alps by a frigid periglacial zone. Cold polar waters extended far south of their present limit in the North Atlantic. Because of lower sea level, areas now covered by waters of the English Channel and the North Sea were dry land.

times when both the climate and the global ice cover were similar to those of today (Fig. 11.27).

That glacial times were both windy and dusty is also shown by studies of fine dust found in ice cores from the Greenland Ice Sheet. The percentage of wind-blown dust rises significantly in the part of the cores that corresponds to the last glaciation. Because Greenland and much of northern North America were ice-covered at that time, much of the dust likely originated along the valleys of braided meltwater streams that crossed windy periglacial zones bordering the ice sheets in North America and Eurasia.

Sea-Level and Lake-Level Changes

The fall of world sea level that accompanied the buildup of glaciers on land changed the shape of the continents as broad areas of shallow continental shelf were exposed (Figs. 14.9 and 14.10). The fall in sea level also changed the gradients of the downstream segments of major streams, causing them to deepen their valleys as they reestablished equilibrium profiles. Stream sediments that had been dumped on the inner continental shelf where a river formerly entered the ocean were now transported across the exposed shelf and deposited at the shelf margin. From there, the accumulating detritus could be carried swiftly down the continental slope by turbidity currents.

In many arid and semi-arid regions of the world, including the Sahara, the Middle East, southern Australia, and the American Southwest, the shift to glacial age climates resulted either in the enlargement of existing lakes or the creation of new ones. For example, during the last glaciation the basin of Great Salt Lake in the western United States was occupied by a gigantic water body that geologists refer to as Lake Bonneville (Fig. 14.9). More than 300 m (985 ft) deeper than Great Salt Lake, Lake Bonneville had a volume comparable to that of modern Lake Michigan. Beaches, deltas of tributary streams, and lake-bottom sediments provide the evidence (Fig. 14.11). Although we might guess that expansion of the lake was caused by increased precipitation, evidence in fact points to *reduced* precipitation during glacial time. Lake Bonneville and other lakes of the American

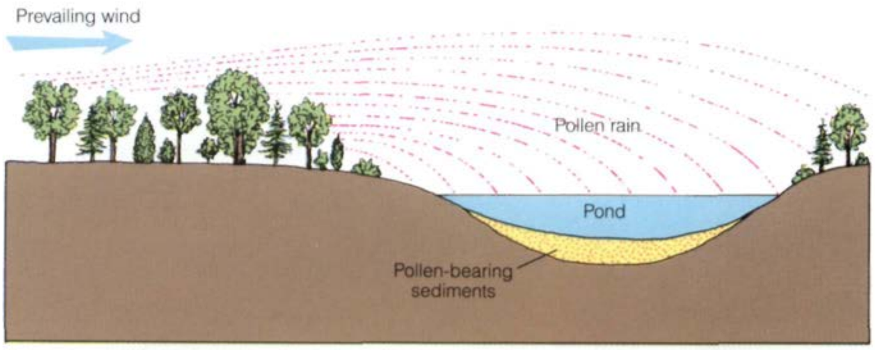
Figure 14.11 Horizontal benches at several levels above the surface of Great Salt Lake, Utah, mark shorelines of Lake Bonneville, a vast Pleistocene lake. At its maximum extent and depth during the last glaciation, the surface of Lake Bonneville stood more than 300 m above that of the present lake.

Southwest formed in a region where present-day evaporation rates are very high, and so an alternative explanation is that lake expansion may have resulted primarily from lower glacial age temperatures that led to reduced water loss by evaporation.

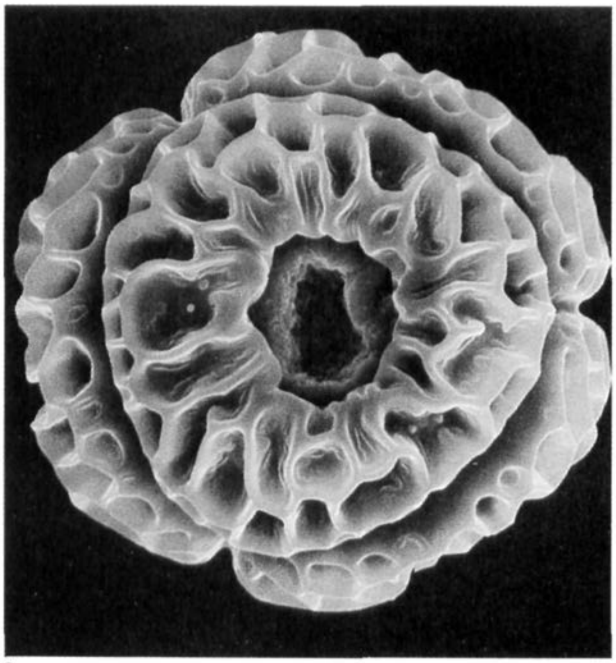
Ice Age Vegetation

Much of our knowledge of climatic conditions outside the great ice sheets during glacial times is based on interpretation of plant fossils. Large plant fragments permit identification of individual species, but they are far less numerous than fossil pollen grains, which possess a hard, waxy coating that resists destruction by chemical weathering. Most pollen is transported by the wind and settles into lakes, ponds, and bogs, where, protected from destructive oxidation in the wet environment, it slowly accumulates (Fig. 14.12). A sample bog or lake sediment yields a vast number of pollen grains that can be identified by type, counted, and treated statistically. At any given level in a core, the pollen grains reveal the assemblage of plants that flourished near the site when the enclosing sediment layers were deposited (Fig. 14.13). If we can find a modern vegetation assemblage that has a composition like that implied by the fossil pollen, then the precipitation and temperature at the site of the modern assemblage can be used to estimate climatic conditions represented by the fossil assemblage.

The vegetation pattern in eastern North America prior to European settlement consisted of several nearly parallel belts that mainly reflected the gradual increase of temperature from pole toward equator. Superimposed on this latitudinal pattern was a change from moist forest in the east to dry grassland in the west. In the Far West, a complex mosaic of vegetation assemblages existed, with patterns determined by latitude, altitude, topography, and distance from the Pacific Ocean and Gulf of Mexico. Pollen studies show us that in glacial times the vegetation distribution was quite different from this recent distribution (Fig. 14.9). About 20,000 years ago, a belt of tundra existed immediately south of the glacier margin with spruce and pine forest to the south of it, consistent with a colder climate. Today's grassland country of the Great



A.



B.

Figure 14.12 Fossil pollen used to reconstruct past vegetation and climate. A. Windborne pollen grains from trees fall into a nearby pond where they are incorporated as part of the accumulating sedimentary strata. B. Scanning electron microscope photograph of a grain of *Drymis winterii* pollen, having a diameter of 42 microns.

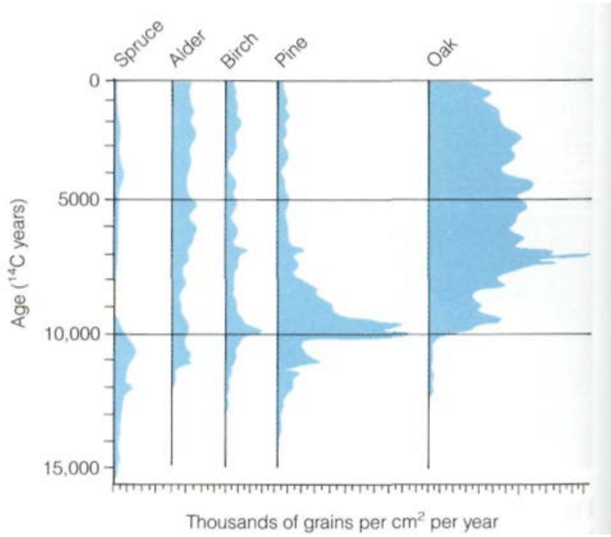


Figure 14.13 Simplified pollen diagram prepared from data collected from Rogers Lake, Connecticut. Variations in pollen influx as a function of time show changes in forest composition. A major change occurred about 10,000 years ago when spruce/pine forest was replaced by a forest dominated by pine and deciduous trees.

Plains was then mostly open pine woodland. It was once supposed that as the ice sheets spread and then retreated, vegetation zones seen on today's map crept gradually southward and then back northward, each maintaining its own character. However, pollen studies show that vegetation changes accompanying the advance and retreat of the great ice sheets were dynamic and far more complicated. Species were displaced in various directions, forming new plant communities that are unknown on the present landscape (Fig. 14.14).

In Europe the vegetational response to glaciation was similar (Fig. 14.10), but with one major difference. In North America migrating plants driven south by the ice could inhabit the relatively warmer lowlands that extended to the Gulf of Mexico. But in Europe, the glacier-clad Alps, 800 km (500 mi) long and 150 km (93 mi) wide, constituted a high, cold barrier north of the Mediterranean Sea. Many species were trapped between the large ice sheet to the north and the Alpine glaciers to the south and were driven to extinction. Thus, Western Europe, which before the glacial ages began had an abundance of tree types, now has only 30 naturally occurring species. By contrast, North America, with no Alpine barrier standing between the Great Lakes and the Gulf of Mexico, has 130 species.

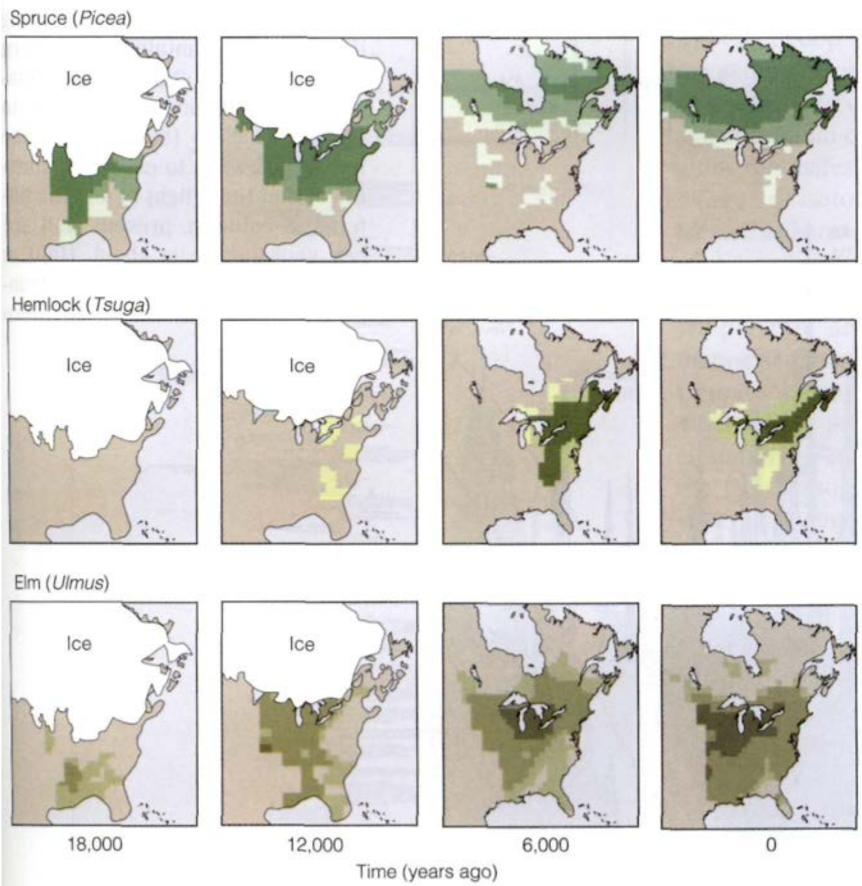


Figure 14.14 Changing distribution of spruce, hemlock, and elm trees in eastern North America at 6000-year intervals between the last glaciation (18,000 years ago) and the present day based on fossil pollen data. The color intensities indicate relative abundance for each species, with the darkest shade of green being the highest and the lightest shade the lowest.

Changes in Temperature and Precipitation

In the popular imagination, glacial ages were times when temperatures were very cold, perhaps rivaling those in the middle of Antarctica today. Although such extreme cold did exist in some regions, in other places *average* temperatures at the culmination of the last glaciation were little, if any, different than they are now. The evidence for glacial age temperatures is varied, and it comes from both the land and the ocean basins.

Estimates of temperature lowering on the land span a range of values. In mid-latitude coastal regions, temperatures were generally reduced by about 5 to 8°C (9 to 14°F), whereas in continental interiors reductions of 10 to 15°C (18 to 27°F) occurred (e.g., Fig. 14.9). Some of the ways such estimates are derived include:

1. By comparing the snowlines of ice age glaciers with those of modern glaciers, a value for snowline depression can be obtained (Fig. 14.15). An estimate of temperature lowering can then be determined by assuming that present and past average rates for the upward decrease of temperature in the atmosphere are similar (6°C/km). When this average rate is applied to the calculated snowline difference, the resulting depression in temperature can be found. Such estimates err, however, in disregarding the effects of precipitation, which also control snowline altitude (Chapter 10).
2. By studying fossil pollen grains, ice age vegetation assemblages can be reconstructed. Using the contemporary range of temperatures for these assemblages, past temperatures can be inferred. This approach works only where an assemblage reconstructed from fossil pollen matches a modern assemblage. A similar approach can be taken using animal fossils, such as beetles, which give comparable results.
3. By obtaining measurements of the oxygen-isotope ratio in ice cores that penetrate ice of the last glacial age, former surface air temperature can be estimated. The measurements show a marked change in isotope values at a level coinciding with

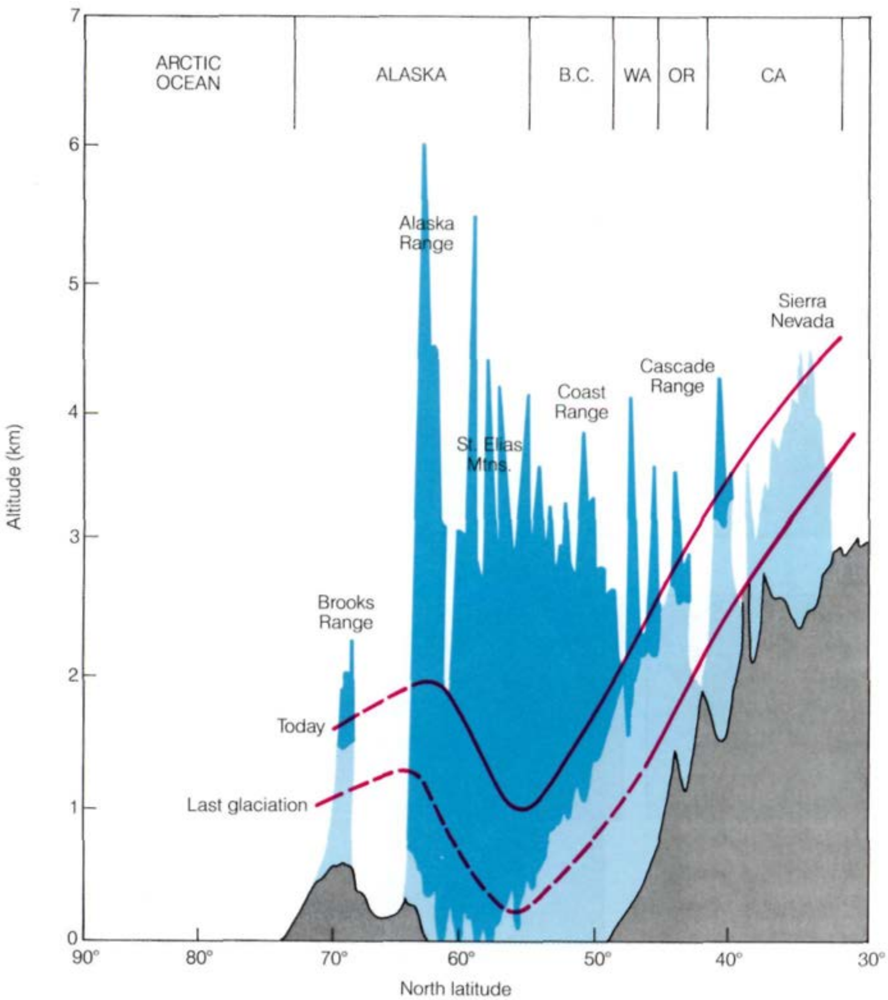


Figure 14.15 Transect along the coastal mountains of western North America showing the relationship of the present snowline to existing glaciers (blue) and of the ice age snowline to expanded glaciers of that time (light blue). The difference between present and ice age snowlines was about 1000 m along the southern part of the transect and about 600 m in northern Alaska.

the transition from mild (Holocene) interglacial climate recorded in the upper parts of the cores to cold ice age temperatures below (Fig. 14.6).

4. By noting the distribution of periglacial features, which indicate the former presence of permafrost in areas now lacking it, an estimate can be made of the minimum temperature change that has taken place. At present, permafrost exists mainly in areas where the mean annual air temperature is below —5°C (23°F). If, for example, evidence of former permafrost is found at a place where the annual temperature is now 4°C (39°F), then the former periglacial climate is inferred to have been at least 9°C (16°F) colder.

5. By examining in deep-sea sediment cores fossils of microorganisms that lived near the ocean surface and by comparing the fossil assemblages with those now living in surface waters, sea-surface temperatures during the last glaciation can be reconstructed. (See "A Closer Look: Reconstructing Ice Age Ocean-Surface Conditions").

From these and other types of evidence obtained on land and from the oceans, we have learned an important fact: the changes accompanying a shift from interglacial to glacial conditions did not affect the whole world equally; the environments of some regions apparently changed little if at all, whereas others experienced profound changes.

Successive Pleistocene Glacial and Interglacial Ages

Until only a few decades ago, it was thought that the Earth had experienced only four glacial ages during the Pleistocene Epoch. This assumption was based on studies of ice sheet and mountain glacier deposits, and it had its roots in early studies of the Alps where geologists identified stream terraces they thought were related to four ice advances. This traditional view was discarded when studies of deep-sea sediments disclosed a long succession of glaciations dur-

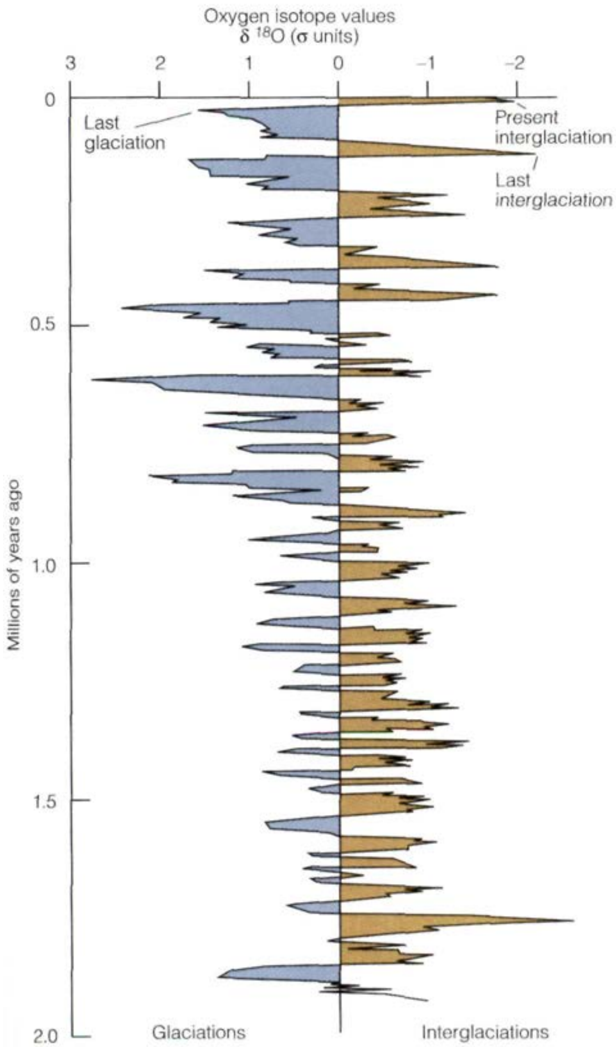


Figure 14.16 Curve of average oxygen-isotope variations during the last 2 million years based on analyses of deep-sea sediment cores. The curve illustrates changing global ice volume during successive glacial-interglacial cycles of the Quaternary Period.

ing the Pleistocene, the most recent of which was shown by radiocarbon dating to equate with deposits of the last glaciation on the continents. Paleomagnetic dating of deep-sea cores (Chapter 6) shows that the most recent glacial-interglacial cycles recorded in the sediments average about 100,000 years long and that during the last 800,000 years alone there have been about eight such episodes. For the Pleistocene Epoch as a whole, about 30 glacial ages are recorded rather than the traditional four. The implications are clear: whereas seafloor sediments provide a continuous historical record of climatic change, evidence of glaciation on land generally is incomplete and interrupted by many unconformities.

The seafloor evidence is of three kinds. First, with increasing depth in a core, the biologic component of the sediments shows repeated shifts from warm interglacial biota to cold glacial biota. Second, the percentage of calcium carbonate in cores from some ocean regions fluctuates in much the same manner. Third, the ^{18}O to ^{16}O isotope ratio fluctuates with a pattern similar to that shown in the biologic and mineral fractions of the sediments. Whereas the isotopic variations in ice cores are believed to represent fluctuations in air temperature near the glacier surface, in Pleistocene marine sediments they are thought primarily to reflect changes in global ice volume. During glacial ages, when water is evaporated from the oceans and precipitated on land to form glaciers, water containing the light isotope ^{16}O is more easily evaporated than water containing the heavier ^{18}O. As a result, Pleistocene glaciers contained more of the light isotope, whereas the oceans became enriched in the heavy isotope. Isotope curves derived from the sediments therefore give us a continuous reading of changing ice volume on the planet (Fig. 14.16). Because glaciers wax and wane in response to changes of climate, the isotopes also give a generalized view of global climatic change.

Over most of the last million years, peaks in the isotope curve that represent times of high global ice volume are nearly equal in amplitude. This suggests that during each glaciation the amount of ice on land was about the same. The interglacial peaks are also nearly equal, indicating that during earlier interglacial ages the planetary ice cover likely was similar to that of today.

A record of ocean-surface temperatures, based on oxygen isotope values in deep-sea cores that penetrate Cenozoic sediment, shows that the oceans have grown colder over the last 50 million years (Fig. 14.17). During one pronounced cooling event about 35 million years ago, surface ocean temperatures declined by nearly 5°C (9°F) within only about 100,000 years. In concert with the long-term cooling trend, glaciers spread from highlands in Antarctica and reached the sea. About 12 to 10 million years ago, ice volume increased, and an ice sheet formed over Antarctica as temperatures continued to fall. The presence of such a large polar ice mass reduced average temperatures on the Earth still further and caused a substantial drop in sea level. From that time onward, large glaciers occupied mountain valleys of Alaska and the southern Andes. Although the evidence is still sketchy, it appears that large ice sheets did not form in northern middle latitudes until about 2.5 million years ago. If this inferred history is correct, glaciation has gradually affected more and more of the Earth's land surface during the Cenozoic: first the Antarctic,

A Closer Look

Reconstructing Ice Age Ocean-Surface Conditions

The upper parts of drill cores taken from the seafloor throughout the world's oceans consist of soft sediments that commonly contain multitudes of tiny fossils. Most of the fossils are of microorganisms that live in the surface waters and whose shells rain down on the seafloor in vast numbers (Fig. C14.1). The rate of sedimentation is extremely slow, however, so that it may take more than a thousand years for a single centimeter of sediment to accumulate. Because the assemblage of organisms that live in the surface waters is closely related to water temperature, the fossil remains in the sediment provide a record of changing conditions at the ocean surface.

In many deep-sea sediment cores, the fossil content changes downward, typically shifting back and forth from predominantly warm-water (interglacial) to cold-water (glacial) forms. By identifying the species present at any level in a sediment core and comparing that assemblage with modern ones, it is possible to infer what the surface ocean temperature must have been when the shells were settling to the seafloor. In practice, geologists can select a level in a core that represents the peak of the last glaciation and determine, from the contained fossils, the surface water temperature at that time. Information from hundreds of cores scattered widely over the oceans has been used to derive a global map of sea-surface temperature for the last glacial maximum (Fig. C14.2).

Surprisingly, the *average* global difference between present and ice age sea-surface temperatures is only about 2.3°C (4°F), but this figure is somewhat misleading. In some large regions, like the subtropics, little or no change in temperature is detected. In others, such as the North Atlantic, sea-surface temperatures were locally as much as 14°C (25°F) colder than now. In this region, cold polar water that is now found mainly north of latitude 60° descended far south at the peak of the glaciation to reach the shore of northeastern United States and the Iberian Peninsula in Western Europe.

The greatest ocean temperature decreases (up to 14°C) occurred in the North Atlantic around which large continental ice sheets were located, in enclosed seas of the northwestern Pacific, and also near the equator, where cold water that welled up off the coasts of Africa and South America spread westward across the equatorial Atlantic and Pacific, respectively. However, over vast areas within the North and South Pacific midocean gyres, sea-surface temperatures changed hardly at all.

Figure C14.1 Skeletons of calcareous foraminifera (smooth globular objects), siliceous radiolaria (delicate meshed objects), and siliceous rod-shaped sponge spicules from a deep-sea ooze, photographed by scanning electron microscope. The fossils are from a sediment core collected in the western Indian Ocean during a Deep Sea Drilling Project cruise.

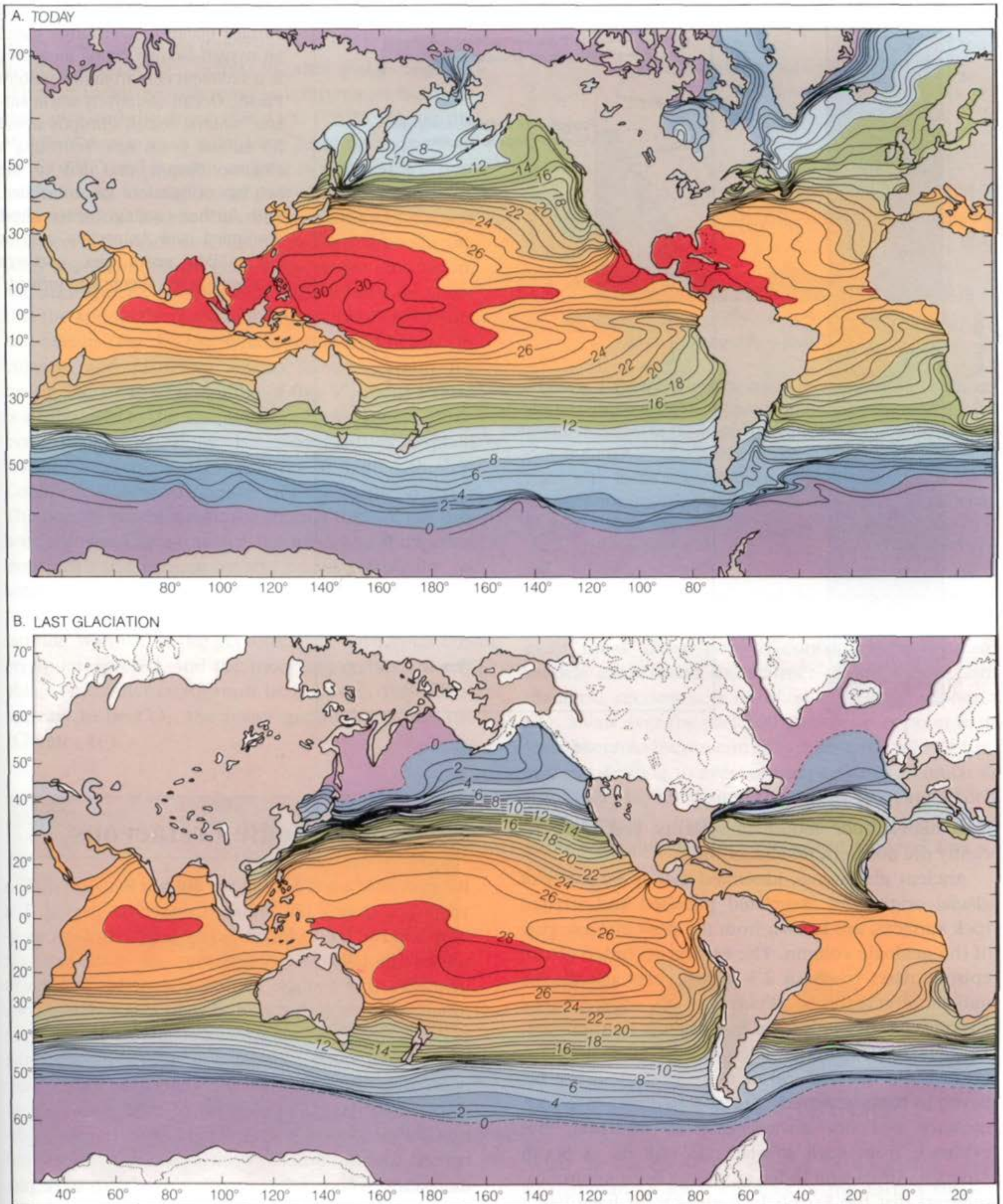


Figure C14.2 Present and past temperatures of surface waters in the World Ocean. A. Map showing modern August sea-surface temperatures (°C). B. Map showing reconstructed August sea-surface temperatures during the last glaciation, about 18,000 years ago. Cold polar water extended far south of its present limit in the North Atlantic, and plumes of cold water extended westward in the equatorial Pacific and Atlantic.

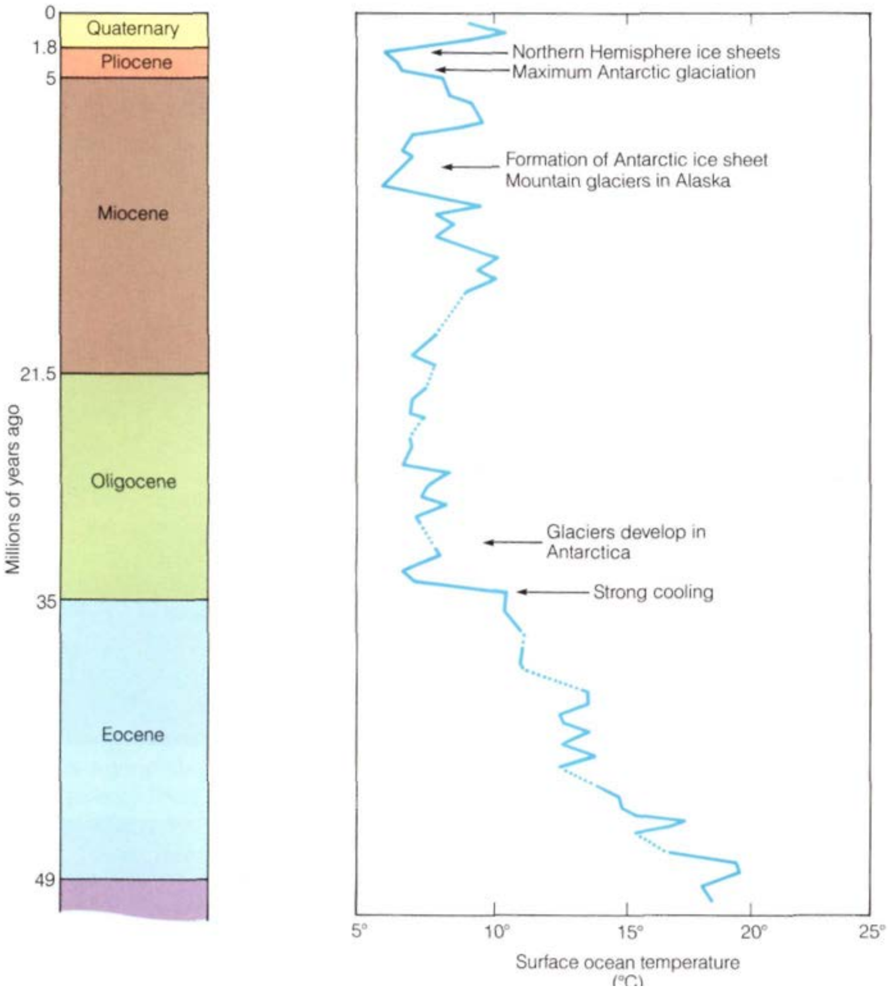


Figure 14.17 A long record of surface ocean temperatures based on oxygen-isotope ratios measured in a sediment core from the western Pacific Ocean. Relatively warm surface waters cooled abruptly about 35 million years ago, reflecting a climatic change that led to the buildup of glaciers in Antarctica. With further cooling, an ice sheet developed over Antarctica, and by 2.5 million years ago, northern hemisphere ice sheets had formed.

then high-latitude mountain systems, and more recently the northern middle latitudes.

Ancient glaciations, identified mainly by rocks of glacial origin and associated polished and striated rock surfaces, are known from the pre-Cenozoic part of the geologic column. The earliest recorded glacial episode dates to about 2.3 billion years ago, in the middle Precambrian. Evidence of other glacial episodes has been found in rocks of late Precambrian, early Paleozoic, and late Paleozoic age. During the latest of these intervals, 50 or more glaciations are believed to have occurred. The geologic record is fragmentary and not always easy to interpret, but evidence from such low-latitude regions as South America, Africa, and India, as well as from Antarctica, suggests that the Earth's land areas must have had a very different relationship to one another during the late Paleozoic glaciation than they do today. In the Mesozoic Era, glaciation of similar magnitude apparently did not occur, for geologic evidence points to a long interval of mild temperatures both on land and in the oceans.

The Warm Middle Cretaceous

It's probably a good thing we did not live 100 million years ago during the Middle Cretaceous Period. Not only was the world inhabited by huge carnivorous dinosaurs, but also the climate was one of the warmest in the Earth's history. Evidence that the world was much warmer in that period than it is today is compelling (Fig. 14.18). Warm-water marine faunas were widespread, coral reefs grew 5° to 15° closer to the poles than they do now, and vegetation zones were displaced about 15° poleward of their present positions. Peat deposits that would give rise to widespread coal formations formed at high latitudes, and dinosaurs, which are generally thought to have preferred warm climates, ranged north of the Arctic Circle. Sea level was 100 to 200 m (330 to 650 ft) higher than today, implying the absence of polar ice sheets, and isotopic measurements of deep-sea deposits indicate that intermediate and deep waters in the oceans were 15 to 20°C (27 to 36°F) warmer than now. Based on such evidence, average global temperature is estimated to have been at least 6°C (11°F) milder than today and possibly as much as 14°C (25°F), with the greatest difference being in the polar regions. Whereas today the difference in temperature between the poles and the equator is 41°C (74°F), during the Middle Cretaceous it may have been no more than 26°C (47°F) and possibly as little as 17°C (31°F).

Computer simulations of past climates provide insights into the Middle Cretaceous world and suggest that several factors were likely involved in producing such warm conditions: geography, ocean circulation, and atmospheric composition. The simulations show that the Middle Cretaceous arrangement of continents and oceans (Fig. 14.18), which influenced ocean circulation and planetary albedo, could account for nearly 5°C (9°F) of warming; of this 5°, about a third is attributable to the absence of polar ice sheets. However, geography alone is inadequate to explain warmer year-round temperatures at high latitudes. Could the poleward transfer of heat be the answer? The oceans now account for about a third of the present poleward heat transfer, but modeling shows that even with the geography and ocean circulation rearranged as they were in the Middle Cretaceous, oceanic heat transfer cannot explain the greater high-latitude warmth. If the geologic data have been correctly interpreted, and the modeling results are reliable, some other factor must be involved. This factor appears to be CO_2, the major greenhouse trace gas (Chapter 18).

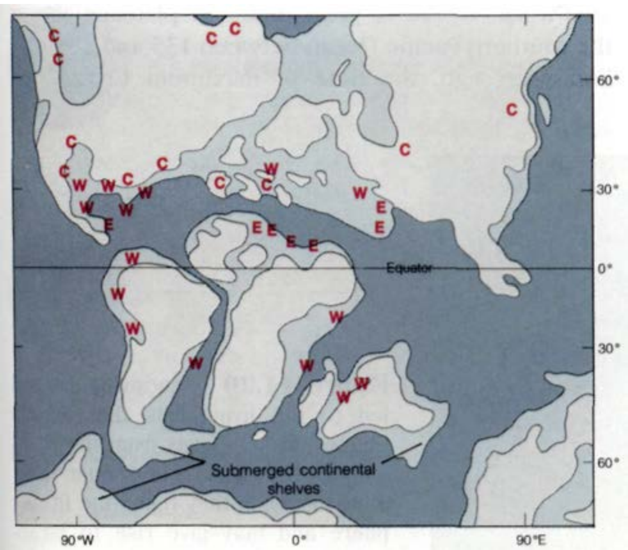


Figure 14.18 During the Middle Cretaceous period, sea level was 100 to 200 m higher than now and flooded large areas of the continents, producing shallow seas. Warm-water animal assemblages (W) and evaporite deposits (E) were present at low to middle latitudes, while coal deposits (C) developed in northern latitudes, implying warmer year-round temperatures.

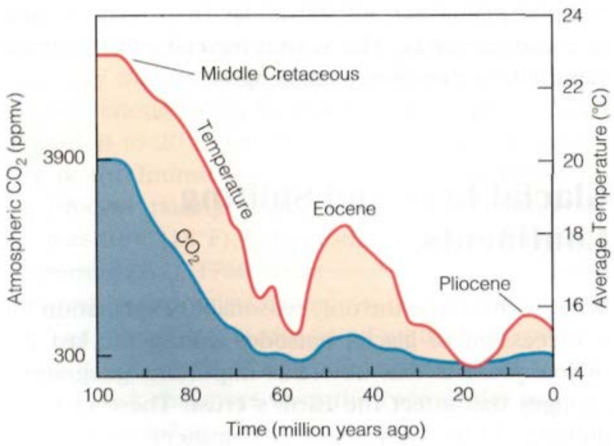


Figure 14.19 A geochemical reconstruction of changing atmospheric CO_2 concentration and average global temperature over the past 100 million years. High CO_2 values and high temperatures in the Middle Cretaceous contrast with much lower present values. Other intervals of higher temperature and CO_2 occurred during the Eocene and the Middle Pliocene.

The simulations show that, by rearranging the geography and also increasing carbon dioxide six to eight times above present concentrations, climate models can explain the warmer temperatures. Geochemical reconstructions of changing atmospheric CO_2 levels over the past 100 million years point to at least a tenfold increase in CO_2 during the Middle Cretaceous, leading to average temperatures as much as 8°C (14°F) higher than now (Fig. 14.19). Under such conditions, it is easy to see why ice volume on the Earth was unusually low and world sea level was so high.

WHY CLIMATES CHANGE

What factors cause the climate to warm and cool, bringing about great changes in the Earth's surface processes and environments? The search for an answer has proved difficult because climate changes on different time scales, ranging from decades to many millions of years, and several quite different mechanisms appear to be responsible for these changes. Furthermore, these mechanisms involve not only the atmosphere, but also the lithosphere, the oceans, the biosphere, and extraterrestrial factors, all interacting in a complex way. The search for causes of climatic variability is therefore a challenging one.

Glacial Eras and Shifting Continents

What seems to be the only reasonable explanation for a succession of glacial episodes during the last 2.3 billion years is the slow but important geographic changes that affect the Earth's crust. These changes include (1) the movement of continents as they are carried along with shifting plates of lithosphere, (2) the creation of high mountain chains where plates collide, and (3) the opening or closing of ocean basins and seaways between moving landmasses.

How such movements affect climate is illustrated by the fact that low temperatures are found, and glaciers tend to form and persist, in two kinds of situations: high latitudes and high altitudes, especially in places where winds can supply abundant moisture evaporated from a nearby ocean. The Earth's largest existing glacier is centered on the South Pole, where temperatures are constantly below freezing and the land is surrounded by ocean. The only glaciers found at or close to the equator lie at extremely high altitudes.

Abundant evidence now leads us to conclude that the positions, shapes, and altitudes of landmasses have changed with time (Chapter 6), in the process altering the paths of ocean currents and atmospheric circulation. As landmasses and ocean basins have shifted position, occasionally they have assumed an arrangement that was optimal for widespread glaciation in high latitudes. Where evidence of ancient ice sheet glaciation is now found in low latitudes, we invariably find evidence that such lands formerly were located in higher latitudes. Although this explanation appears adequate for the pattern of glaciation during and since the late Paleozoic, information about earlier glacial intervals is very fragmentary and more difficult to evaluate.

Why Was the Middle Cretaceous Climate So Warm?

Interspersed with ancient glacial intervals were episodes of exceptionally warm climate, like that of the Middle Cretaceous. If CO_2 was an important factor in Middle Cretaceous warming, as suggested earlier, we still are faced with explaining how this gas increased so substantially. The most likely explanation is volcanic activity, which today constitutes a major natural source of CO_2 entering the atmosphere. Most of this CO_2 is generated by slow, noneruptive degassing of magmas in the upper crust.

Geologic evidence points to an unusually high rate of volcanic activity in the Middle Cretaceous. Rates of continental drift were then about three times as great as now, implying increased extrusion rates at spreading ridges. In addition, vast outpourings of lava created a succession of great undersea plateaus across the southern Pacific Ocean between 135 and 115 million years ago, the time of maximum Cretaceous

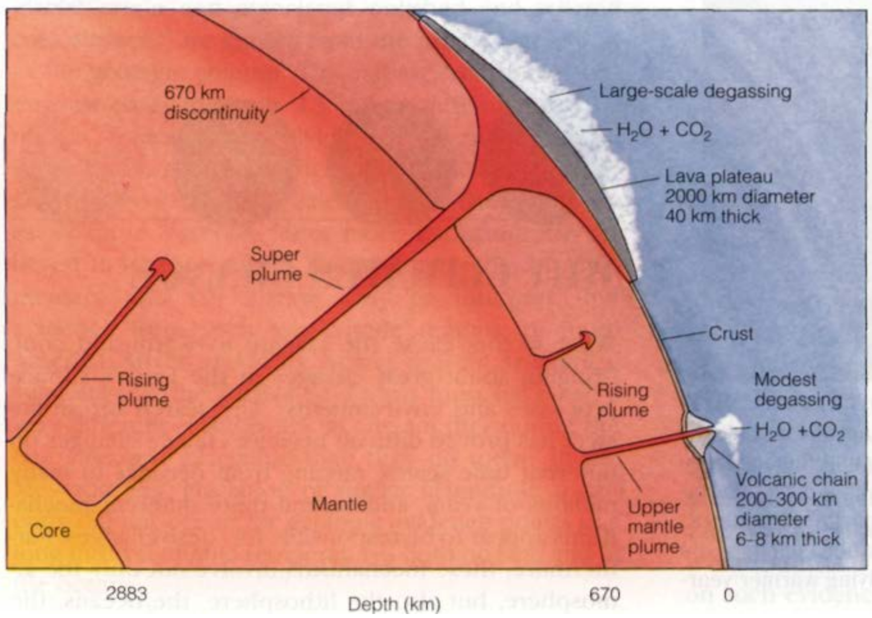


Figure 14.20 New evidence has led to the hypothesis that superplumes, rising slowly from the core-mantle boundary, build huge lava plateaus when they reach the lithosphere and may give rise to large-scale degassing of CO_2, producing a greatly enhanced greenhouse effect. By contrast, plumes rising from the base of the upper mantle at 670 km produce much smaller hot spots that generate volcanoes like those of the Hawaiian chain.

warmth. One of these—the Ontong-Java Plateau in the southwest Pacific—has more than twice the area of Alaska and reaches a thickness of 40 km (25 mi). Such a massive outpouring of lava likely released massive amounts of CO_2. Could this gas emission have been sufficient to warm the climate to unprecedented levels? By one calculation, the eruptions could have released enough CO_2 to raise the atmospheric concentration to 20 times its natural value at the beginning of the Industrial Revolution (ca. AD. 1760), in the process raising average global temperature as much as 10°C (18°F). Other estimates range from 8 to 12 times the A.D. 1760 value.

Recently, geologists have proposed that each such vast lava outpouring is associated with a *superplume,* which is conceived of as a plumelike mass of unusually hot rock that rises from the base of the mantle. Moving upward at a rate of 10 to 20 cm/yr (4 to 8 in/yr), the hot rock spreads out in a mushroom shape as it reaches shallower depths where confining pressures are lower (Fig. 14.20). Such a superplume would be an efficient mechanism for allowing heat to escape from the Earth's core. If this hypothesis is correct, then the plate tectonic cycle cools the mantle both by heat loss at spreading ridges and by the downward plunge of plates of cool lithosphere, while superplumes cool the core. By this reasoning, the core and atmosphere are linked dynamically, and the warm Middle Cretaceous climate was a direct consequence of the cooling of the Earth's deep interior.

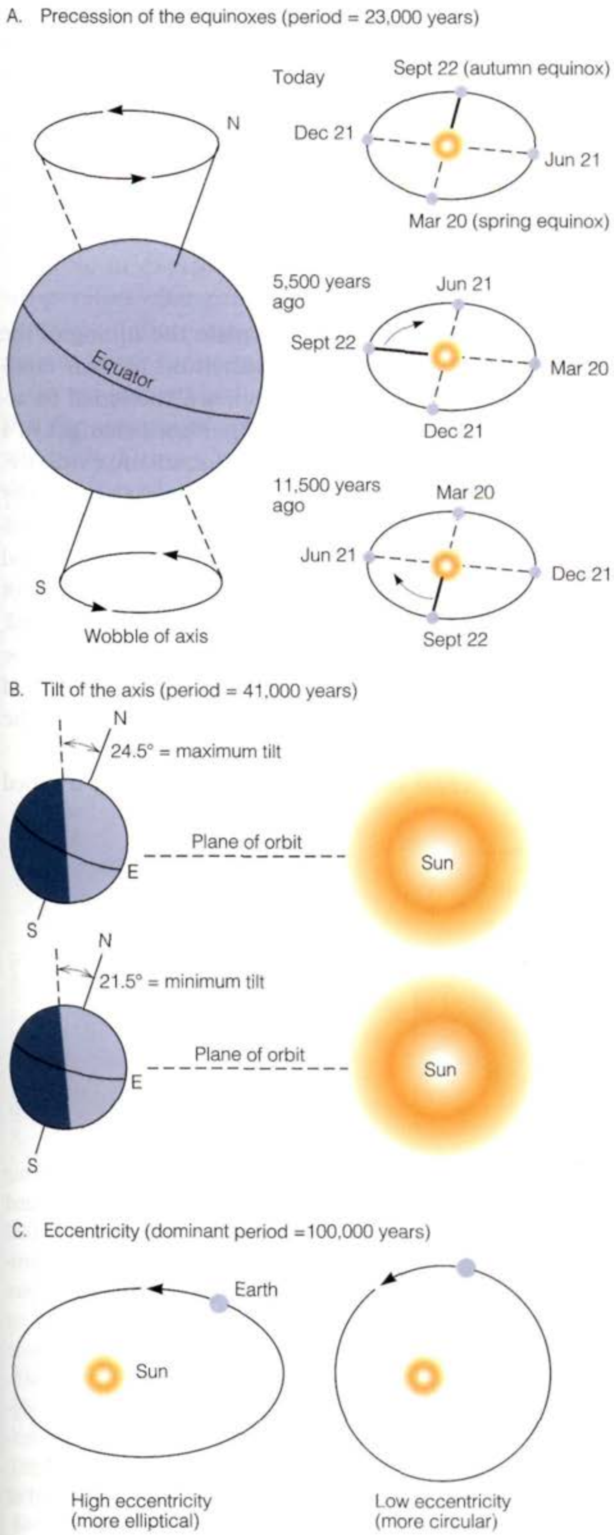


Figure 14.21 Geometry of the Earth's orbit and axial tilt. A. *Precession.* The Earth wobbles on its axis like a spinning top, making one revolution every 26,000 years. The axis of the Earth's elliptical orbit also rotates, though more slowly, in the opposite direction. These motions together cause a progressive shift, or precession, of the spring and autumn equinoxes, with each cycle lasting about 23,000 years. B. *Tilt.* The tilt of the Earth's axis, which now is about 23.5°, ranges from 21.5 to 24.5°, with each cycle lasting about 41,000 years. Increasing tilt means a greater difference, for each hemisphere, between the amount of solar radiation received in summer and that received in winter. C. *Eccentricity.* The Earth's orbit is an ellipse with the Sun at one focus. Over 100,000 years, the shape of the orbit changes from almost circular (low eccentricity) to more elliptical (high eccentricity). The higher the eccentricity, the greater the seasonal variation in radiation received at any point on the Earth's surface.

Ice Age Periodicity and the Astronomical Theory

Determining the cause of the cyclic pattern of glacial and interglacial ages has long been a fundamental challenge to the development of a comprehensive theory of climate. A preliminary answer was provided by Scottish geologist John Croll, in the mid-nineteenth century, and later elaborated by Milutin Milankovitch, a Serbian astronomer of the early twentieth century.

Croll and Milankovitch recognized that minor variations in the Earth's orbit around the Sun and in the tilt of the Earth's axis cause slight but important variations in the amount of radiant energy reaching any given latitude. Three movements are involved (Fig. 14.21).

First, the axis of rotation, which now points in the direction of the North Star, wobbles like the axis of a spinning top (Fig. 14.21A). The wobbling movement causes the North Pole to trace a cone in space, completing one full revolution every 26,000 years. At the same time, the axis of the Earth's elliptical orbit is also rotating, but much more slowly, in the opposite direction. These two motions together cause a progressive shift in the position of the four cardinal points of the Earth's orbit (spring and autumn equinoxes and winter and summer solstices). As the equinoxes move slowly around the orbital path, a motion called *precession of the equinoxes,* they complete one full cycle in about 23,000 years.

Second, the *tilt* of the axis, which now averages 23.5°, shifts about 1.5° to either side during a span of about 41,000 years (Fig. 14.21B).

Finally, the *eccentricity* of the orbit, which is a measure of its circularity, changes over a period of 100,000 years. About 50,000 years ago, the orbit was more circular (lower eccentricity) than it has been for the last 10,000 years (Fig. 14.21C).

The slow but predictable changes in precession, tilt, and eccentricity cause long-term variations of as much as 10 percent in the amount of radiant energy that reaches any particular latitude on the Earth's surface in a given season (Fig. 14.22). By reconstructing and dating the history of climatic variations over hundreds of thousands of years, geologists and oceanographers have shown that fluctuations of climate on glacial-interglacial time scales match the predictable cyclic changes in the Earth's orbit and axial tilt. This persuasive evidence supports the theory that these astronomical factors control the *timing* of the glacial-interglacial cycles.

Amplification of Temperature Changes

Although orbital factors can explain the timing of the glacial-interglacial cycles, the variations in solar radiation reaching the Earth's surface are too small to account for the average global temperature changes of 4 to 10°C (7 to 18°F) implied by paleoclimatic evidence. Somehow, the slight temperature decreases caused by orbital changes must have been amplified into temperature changes sufficiently large to generate and maintain the huge Pleistocene ice sheets. We do not yet know how this amplification was accomplished, but some of the factors involved are likely to be changes in the chemical composition and dustiness of the atmosphere and changes in the reflectivity of the Earth's surface.

The chemical composition of air bubbles trapped

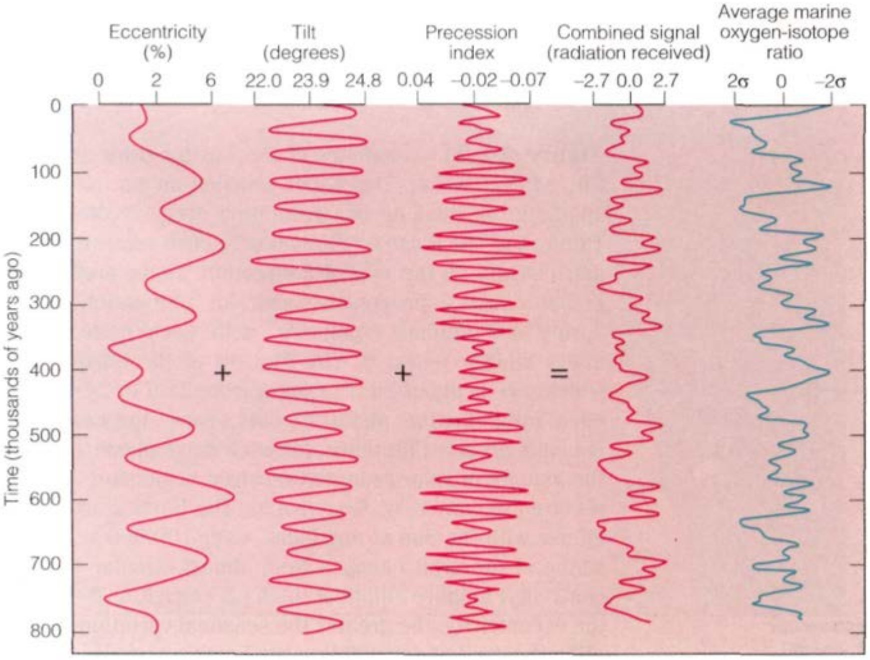


Figure 14.22 Curves showing variations in eccentricity, tilt, and precession during the last 800,000 years. Summing these factors produces a combined signal that shows the amount of radiation received on the Earth at a particular latitude through time. The frequency of oscillations in the combined orbital signal closely matches that of the marine oxygen isotope curve (on right), which constitutes a proxy record of changing global ice volume.

in polar glaciers indicates that during glacial times the atmosphere contained less carbon dioxide and methane than it does today (Fig. 14.23). These two gases are important greenhouse gases (Chapter 18). If their concentration in the atmosphere is high, they trap radiant energy emitted from the Earth's surface that would otherwise escape to outer space. As a result, the lower atmosphere heats up and the Earth's climate becomes warmer. If the concentration of these gases is low, as it was during glacial times, surface air temperatures are reduced. Calculations suggest that the low levels of these two important atmospheric gases during glacial times can account for nearly half of the total ice age temperature lowering. Therefore, the greenhouse gases likely play an important role in explaining the *magnitude* of past global temperature changes. Although we know that the percentages of these gases fell during glacial times, we do not yet know for certain what caused them to drop.

As we learned earlier, ice core studies have shown that the amount of dust in the atmosphere was unusually high during glacial times. The fine atmospheric dust scattered incoming radiation back into space, which would have further cooled the earth's surface.

Whenever the world enters a glacial age, large areas of land are progressively covered by snow and glacier ice. The highly reflective surfaces of snow and ice scatter incoming radiation back into space, further cooling the lower atmosphere. Together with lower greenhouse gas concentrations and increased atmospheric dust, this additional cooling would favor the expansion of glaciers.

Changes in Ocean Circulation

As we discussed in Chapter 8, the circulation of the world ocean plays an important role in global climate. The thermohaline circulation system links the atmosphere with the deep ocean. Warm surface water moving northward into the North Atlantic evaporates, and the remaining water becomes more saline and cools. The resulting cold, saline water is dense and sinks to produce cold North Atlantic Deep Water. Heat released to the atmosphere by evaporation maintains a relatively mild interglacial climate in northwest Europe. Consider what would happen, however, any time this system closed down.

The rate of thermohaline circulation is sensitive to surface salinity at sites where deep water forms. Studies have shown that during times of reduced salinity, thermohaline circulation is also reduced. We therefore can postulate that as summer radiation decreased

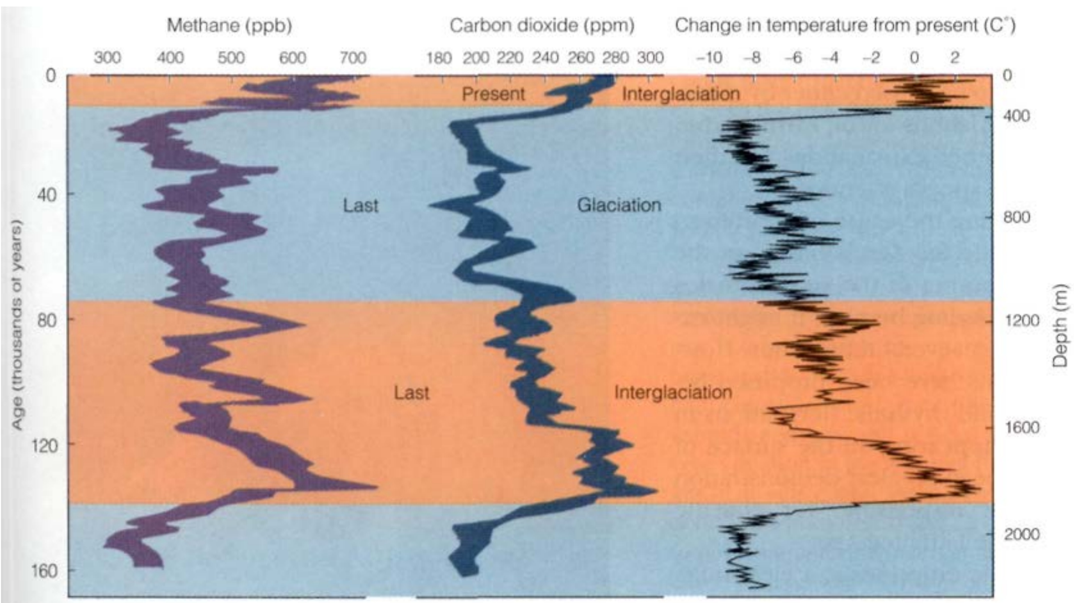


Figure 14.23 Curves comparing changes in carbon dioxide and methane with temperature changes based on oxygen-isotope values in samples from a deep ice core drilled at Vostok Station, Antarctica. Concentrations of these greenhouse gases were high during the early part of the last interglaciation, just as they are during the present interglaciation, but they were lower during glacial times. The curves are consistent with the hypothesis that these gases contributed to warm interglacial climates and cold glacial climates.

at the onset of a glaciation, the high latitude ocean and atmosphere cooled, decreasing evaporation and leading to expansion of sea ice. The resulting freshening of the high-latitude surface waters would have halted the production of dense saline water, thereby shutting off thermohaline circulation. Reduction of high-latitude evaporation, significantly reducing the release of heat to the atmosphere, would have maintained cold air masses moving eastward across the North Atlantic. Further cooled by an expanding sea-ice cover in the North Atlantic and extensive ice sheets on the adjacent continents, the climate of Europe became increasingly cold, causing permafrost to form in a broad zone beyond the ice sheet margin (Fig. 14.10).

Thus, a change in the ocean's thermohaline circulation system provides a means of amplifying the relatively small climatic effect attributable to astronomical changes. Furthermore, it may help explain why the Earth's climate system appears to fluctuate between two relatively stable modes, one in which the ocean conveyor system is operational (interglaciation) and one in which it has shut down (glaciation).

Solar Variations and Volcanic Activity

Climatic fluctuations measured in centuries or decades were responsible for the Little Ice Age and similar episodes of glacier expansion. However, such fluctuations are too brief to be caused either by movements of continents or variations in the Earth's orbit, and require us to seek other explanations for their cause. Two have received special attention.

One hypothesis regarding the cause of short-lived glacial events like the Little Ice Age is based on the concept that the energy output of the Sun fluctuates over time. The idea is appealing because it might explain climatic variations on several time scales. However, although correlations have been proposed between weather patterns and rhythmic fluctuations in the number of sunspots appearing on the surface of the Sun, as yet there has been no clear demonstration that solar variations are responsible for climatic changes on the scale of the Little Ice Age.

Large explosive volcanic eruptions can eject huge quantities of fine ash into the atmosphere to create a veil of fine dust that circles the globe (Fig. 14.24). Like other types of dust, the fine ash particles tend to scatter incoming solar radiation, resulting in a slight cooling at the Earth's surface. Although the dust settles out rather quickly, generally within a few months to a year, tiny droplets of sulfuric acid, produced by the interaction of volcanically emitted SO_2 gas and water vapor, also scatter the Sun's rays, and such droplets remain in the upper atmosphere for several years. The major eruptions of Krakatau (A.D. 1883) and Tambora (A.D. 1815) in the East Indies lowered average surface temperatures in the northern hemisphere by about 0.3 and 0.7°C (0.5 and 1.3°F), respectively. A far greater eruption of Toba volcano about 74,000 years ago, the largest known prehistoric explosive eruption, may have lowered surface temperatures in the northern hemisphere by 3 to 5°C (5 to 69°F).

Figure 14.24 The explosive eruption of Mount St. Helens in 1980 produced a rapidly rising column of ash and gas that reached the stratosphere. There, upper-level winds transported the eruptive products eastward across the United States, and eventually around the world. Although the climatic effects of this relatively modest eruption were unimpressive, much larger explosive eruptions during the last 200 years have cooled temperatures in the northern hemisphere by 0.3 to 0.7°C.

Guest Essay

Is Our Climate Changing? Tree-Ring Records Can Put Extreme Weather into Context

Lisa J. Graumlich uses her interdisciplinary training in biology and climatology to develop long-term proxy records of climatic variation and vegetation dynamics. She received a B.S. in Botany and an M.S. in Geography at the University of Wisconsin. She received a Ph.D. in Dendrochronology from the College of Forest Resources at the University of Washington. She is currently an Associate Professor in the Laboratory of Tree-Ring Research and the Director of the Institute for the Study of Planet Earth (ISPE). In her role as Director of ISPE, she is developing computer-based instructional tools to teach Earth System Science concepts to undergraduate and graduate students.

In the last several years, weather has moved from the back pages of our daily newspapers to the front-page headlines. During the summer of 1988, record heat scorched the Midwestern United States and associated drought reduced the crop yields in our nation's breadbasket by almost 50% over that of previous years. In 1988, 1989, and again in 1990, climatologists reported that global temperatures had reached the highest levels ever recorded in a composite 120-year record. The "greenhouse effect" (see Chapter 18) became a household term as citizens around the world worried that human-induced changes in the composition of the atmosphere had significantly and irrevocably altered the Earth's climate. More recently, in 1993 record rains drenched the Midwest causing record flooding in the Mississippi River, once again disrupting food production and commercial trade. The subsequent winter of 1994 saw record winter cold, forcing the Wall Street financial markets and the federal government to close for days at a time. In the public's mind, the question emerges: What is happening to our weather? To the scientific community, a related question has gained wide-spread attention: Are these weather events a harbinger of large-scale climatic variation or simply a part of natural climatic variability?

These questions, while urgent, are not new. When I was in graduate school in the late 1970s, weather was also in the news. At that time, the west coast of the United States was experiencing its worst drought episode since the dust bowl years of the 1930s and droughts were causing widespread famine in the Sahel of Africa. In my climatology classes, professors expressed dismay and articulated a question that has simultaneously haunted and intrigued me ever since: Are our weather records, which for most of the world extend back only to the early 20th century, adequate to understand the complex dynamics of the climate system? As I puzzled over this question, one professor suggested that I might find an answer to that question by looking at the record of climate recorded in tree rings, a record that extends for several hundred, or, in some cases, several thousand years back in time. With her encouragement and support, I visited the Laboratory of Tree-Ring Research at the University of Arizona, the oldest and largest tree-ring lab in the world, to take an intense seminar in using tree-ring data to reconstruct climate. That short trip was the first step in a life-long odyssey that has taken me to some of the most beautiful mountain ranges of the world in search of the old trees that are providing me the answers to current questions about climate change.

The field of tree-ring research, or dendrochronology, is based on several key concepts. During the growing season, trees produce xylem cells, specialized for the upward transport of water and nutrients, along the outermost circumfrence of the bole. Early in the growing season the tree produces cells that are large, have thin walls, and appear light in color, tater in the growing season, the tree produces cells that are small, have thick cell walls, and appear dark in color. The distinctive alteration of light and dark cells marks a single annual growth ring, often visible to the naked eye when we examine a freshly cut stump. Luckily, tree-ring scientists do not have to cut down trees to study the patterns of the rings. We use a coring device that removes a core about the size of a pencil from the tree, a process that causes no damage to the tree. Back in the laboratory, we sand the core samples to a fine finish and use a light microscope, interfaced with a computer-aided measuring device, to measure the ring widths. (See Fig. 14.7.)

What do the tree rings tell us about the history of climate? The answer to this question depends on the environmental conditions where the trees are growing. If the tree is growing at a site characterized by cold temperatures and a short growing season (e.g. high-latitude or high-elevation sites), year-to-year variation in summer warmth and growing season length will cause year-to-year variation in the width of the ring. If the tree is growing at a site characterized by hot temperatures and a deficit of soil moisture (e.g., trees growing at the edge of semi-arid environments), variations in precipitation will cause variation in the width of the ring.

In my own research, I use these different types of sites to obtain different types of climatic histories. When I sample spruce trees at the northernmost limit of tree growth in the Brooks Range of Alaska, the samples indicate the history of summer temperature fluctuations. Working with archeologists from the region, I'm comparing my temperature reconstructions to the cultural history of the native peoples of this region in order to understand the role climate has played in the development of their subsistence strategies over the past several hundred years. Similarly, tree-ring samples from 1000-year-old pines growing near the crest of the Sierra Nevada of California reveal the history of temperature fluctuations and help us understand the degree to which the warm temperatures of recent years are truly anomalous in the context of the natural variability of climate. Alternatively, my cores from juniper trees at the margins of the Gobi desert, high on the Tibetan Plateau of China, allow me to reconstruct a 1600-year history of the fluctuating rainfall. Working with Dr. Lonnie Thompson and our Chinese colleague Dr. Yao Tandong (see chapter 10), we are comparing the ice core and tree-ring records to understand the complex dynamics of Asian monsoons.

Summary

1. Climate, the average weather conditions over a period of years, is determined by such factors as temperature, precipitation, cloudiness, and windiness. Former climates are determined mainly from fossil plants and animals, sedimentary deposits, and isotopic studies.
2. Changes in climate within the last 100 to 200 years are recognized in instrumental records. Although decadal-scale fluctuations characterize the climatic record, the overall recent trend of average temperature has been upward, with values in the early 1990s reaching record levels.
3. Climatic changes during the last 1000 years are established (with less accuracy) by proxy records. In northern middle latitudes, an interval of mild climate during the Middle Ages was followed by cooler conditions during the Little Ice Age.
4. During the last glaciation, land-surface temperatures were 5 to 15° C lower than today's. Sea-surface temperatures locally fell as much as 14°C, with the greatest changes occurring in the North Atlantic, the North Pacific, and the equatorial zone. Temperature changes are determined from fossils on land and in deep-sea cores, from isotopic measurements of ice cores, from evidence of lowered snowlines, and from the distribution of periglacial features.
5. Glacial ages have alternated with interglacial ages in which temperatures approximated those of today. Studies of marine cores indicate that during the last 800,000 years there have been eight glacial-interglacial cycles and that during the entire Pleistocene Epoch there have been at least 30 such cycles.
6. Glacial ages are discerned in many parts of the geologic column; their record extends back at least 2.3 billion years.
7. During the Middle Cretaceous Period, world temperatures are estimated to have been 6 to 14°C warmer than now and sea level was 100 to 200 m higher, indicating greatly reduced global ice volume. Computer modeling suggests that the increased warmth can be explained by a six- to eightfold increase in atmospheric carbon dioxide, possibly released during an episode of unusually intense volcanic activity.
8. Long intervals in Earth history marked by repeated glacial-interglacial cycles probably are related to favorable positioning of continents and ocean basins, brought about by movements of lithospheric plates. The timing of the glacial-interglacial cycles appears to be closely controlled by changes in the orbital path and axial tilt of the Earth, which affect the distribution of solar radiation received at the surface.
9. Climatic variations on the scale of centuries and decades may be related to fluctuations in energy output from the Sun, from injections of volcanic dust and gases into the upper atmosphere, or from a combination of these factors.

Important Terms to Remember

glaciation (p. 369)
interglaciation (p. 369)

Questions for Review

1. How did discovery of the frozen body of a Neolithic man in the Alps provide evidence of Holocene climatic change?
2. Describe two types of climate proxy records and explain how they are related to changing climate.
3. Evidence indicates that climates during glacial times were drier than they are today. If true, how can you explain the expansion of lakes in closed basins in such regions as the American Southwest and North Africa during glacial times?
4. How might we try to obtain an estimate of ice age land-surface temperature by studying fossil pollen grains? isotopes in glacier ice? evidence of permafrost distribution?
5. What evidence gained from study of deep-sea cores indicates that glacial-interglacial cycles occurred repeatedly during the Pleistocene Epoch?
6. Why do oxygen-isotope measurements of deep-sea sediments provide evidence of changing global ice volume?
7. What factors may have contributed to making the Middle Cretaceous climate so much warmer than the present climate?
8. Describe the three orbital motions of the Earth that contribute to variations in the distribution of solar radiation reaching any point on the land surface over the course of a glacial-interglacial cycle.
9. What effect can large volcanic eruptions have on climate?

Questions for A Closer Look

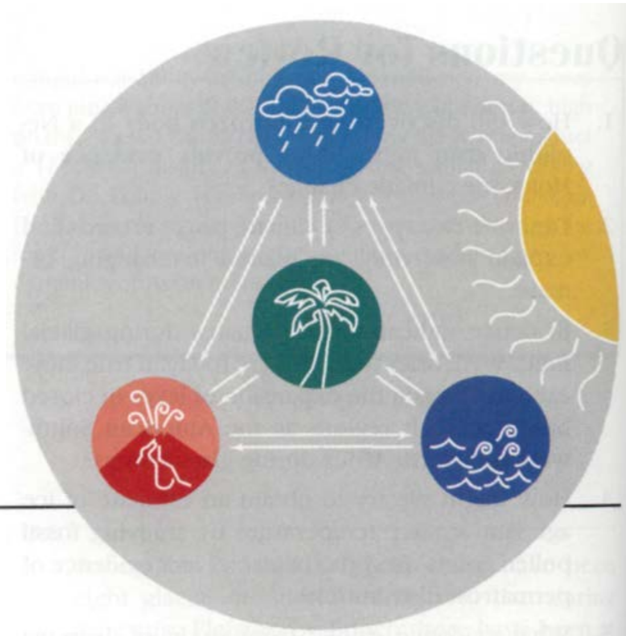
1. How did sea-surface temperatures at the peak of the last glaciation differ from those of the present? How can differences in the degree of change among different oceanic regions be explained
2. If the average sedimentation rate at a site in the Atlantic Ocean is 1 cm/1000 years, how many meters of deep-sea sediment must be extracted to obtain a record spanning the entire Quaternary Period (the last ca. 1.8 million years)?
3. How can sediments accumulating at the *floor* of the deep sea provide information on *surface* water conditions?

Questions for Discussion

1. Name three pieces of evidence that point to a possible change of climate within your lifetime.
2. At the peak of the last glaciation, about 20,000 years ago, the average annual air temperature was about 5 to 7°C lower than now. Describe what environmental effects such a temperature reduction produced in the region where you live.
3. Obtain or compile a record of annual temperature for your community (or a nearby location) for the past 50 to 100 years and compare it with the global temperature record plotted in Figure 18.11. How are the two records similar and how do they differ? What might explain any differences between the two records?

PART FIVE

The Dynamics of Life on Earth



How Do Birds Find the Way?

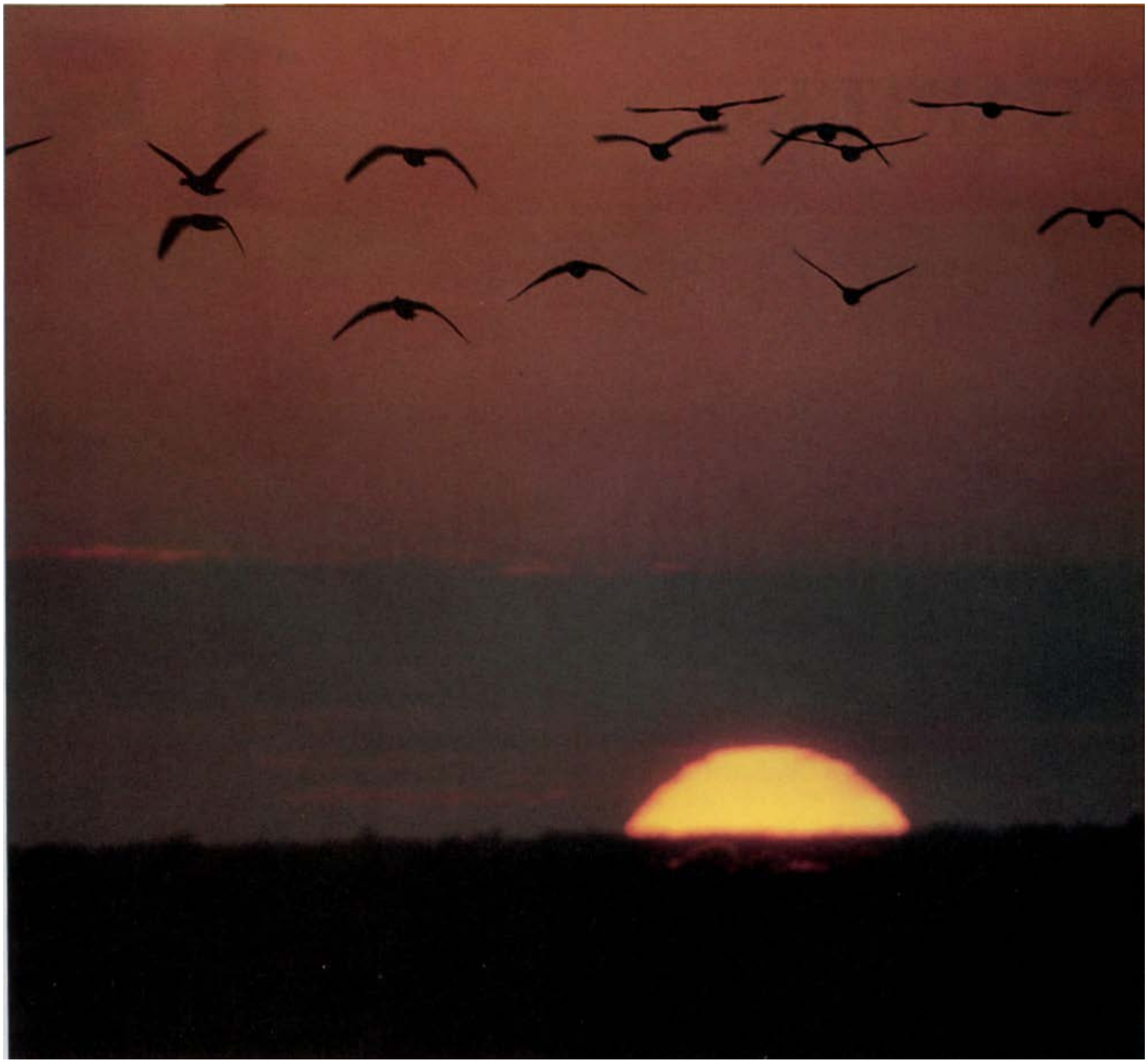
We now consider the fourth part of the Earth system, the biosphere. Refer back to Figures I-1 and I-2 and see that the biosphere is placed at the center of the Earth system. It is accorded this central position because the biosphere is what makes the Earth unique. No other body in the solar system is known to have a biosphere—indeed, no other body in the universe has as yet been proven to have a biosphere.

The many ways organisms in the biosphere employ the resources of the Earth system are fascinating. Almost every property of the other geospheres is employed by some plant or animal. Consider the migration of birds, for example. Because many birds migrate at night and others fly long distances over the sea, it is apparent that at least some birds must rely on something other than visual sighting of familiar land features to find their way. Researchers have found that birds, like the ancient Polynesian navigators discussed in the essay that opens Chapter 8, seem to be able to use a mixture of subtle clues from the other spheres.

Scientists who investigate bird migrations have carried out tests by building large bird cages in which shining lights, magnetic fields, and other features can be varied in strength and position. Long ago, such experiments proved that some night-migrating birds have the ability to use the stars for navigation. Many migration paths are north-south, and some night-migrating birds have apparently learned to fly toward constellations near the poles. Northward-migrating birds, for example, are guided by Ursa Major (also called the Big Dipper) because that constellation is close to the North Pole.

Some birds also use the Earth's magnetic field to help them find their way. Experiments in which the direction and strength of the magnetic field is varied provide evidence that certain birds can sense the magnetic field. When the direction of the field is changed during an experiment, the birds change their direction of flight. How birds sense the field is still a puzzle. Recent research by Joseph L. Kirschvink of the California Institute of Technology has shown that the brains of many birds and animals (including humans) contain minute crystals of magnetite. The magnetite crystals don't seem to help humans sense the field, but it is possible that the crystals may help birds to use the magnetic field to find their way.

Crystals in the brain cannot be the only answer to



How do they find their way? Snow geese (Chen hyerborea) migrating from Delaware to northern Canada always manage to return to the same spot year after year.

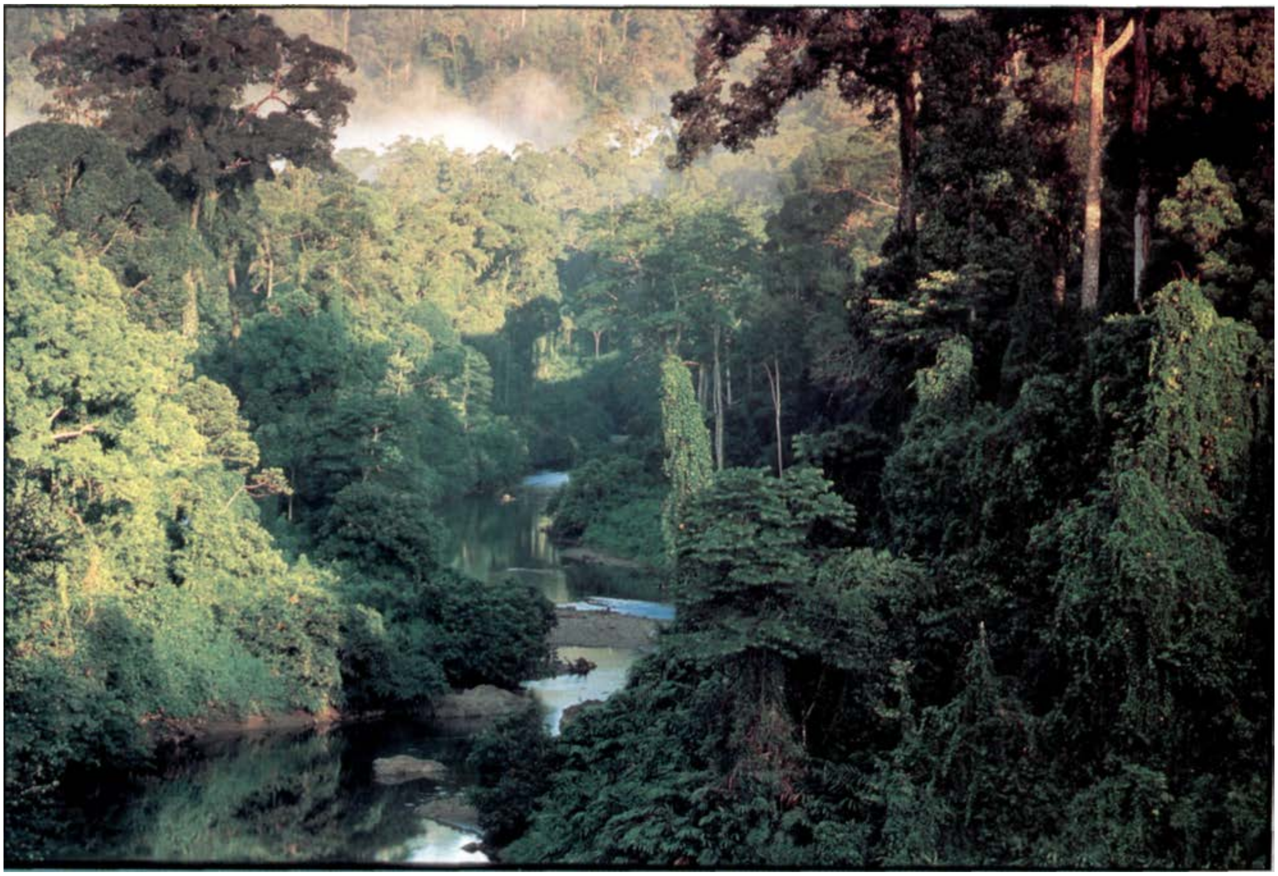
sensing the magnetic field, however. In a 1993 paper published in the British journal *Nature,* scientists reported experiments on tiny Australian birds called silvereyes. It was found that silvereyes can detect the magnetic field, but they can only do so in the light, not in the dark. The reason, the scientists hypothesize, is that light causes some chemical compound in the birds' eyes to lose or gain electrons and thereby to develop magnetic properties.

Another fascinating 1993 paper, also in *Nature,* suggests how certain birds might use sunlight to calibrate their sense of direction. If you look at the sky through a pair of polaroid glasses, you will see that the light intensity varies as you move your head. The reason is that the same scattering of light by the atmosphere that makes the sky blue also causes light to be polarized. The direction of polarization is a plane exactly perpendicular to the incoming direct sunlight. If you look in that direction through your polaroid glasses, the sky will appear dark. Certain birds apparently have natural polaroids in their eyes and can also see the dark band. They can therefore navigate by sunlight without having to look at the sun.

Scientists apparently still have a lot to learn about the myriad ways the creatures of the biosphere interact with and use the other parts of the Earth system.

CHAPTER 15

Dynamics of the Global Ecosystem



The tropical rainforest, here seen along the banks of the Segawa River in Borneo, is the most diverse part of the terrestrial biosphere. We know so little about the number and kinds of organisms in rainforests, and the diversity is so high, that it is possible some small plant or insect becomes an endangered species every time a tree is cut down.

The Biosphere Under Siege

The richest and most diverse part of the biosphere on land is the equatorial rain forest. Increasingly over the past three decades, the rain forest has been destroyed in order to develop agriculture, exploit lumber and mineral resources, and establish transportation routes. It has been projected that, if the destruction of rain forest were to continue at today's rate until A.D. 2050, there would be none left by then.

Because the equatorial forest is the richest part of the terrestrial biosphere, its destruction must be accompanied by the extinction of a large number of species, the majority of which perhaps we don't even know yet. If things go on as they are, perhaps a million species will have been extinguished by 2050. A million species in, say, a century: 10,000 species a year. The natural background rate of extinction is about eight species a year. The great natural mass extinctions of the past (Permian-Triassic, Cretaceous-Tertiary) accounted for perhaps hundreds of species a year. Even if such estimates are off by a factor of 10, the difference between human-induced and natural extinction rates must still be ten times greater than the error.

The pity of this scenario goes beyond the sadness of impoverishing nature. "Civilized" society is destroying the habitat of "primitive" human communities in the Amazon basin, in Borneo, and in several other rain forest environments: communities that have long lived in harmony with tropical rain forests. By allowing the destruction of rain forest, the "civilized" world is inflicting a great loss on itself, too. A large number of medicinal drugs have been discovered in rain forest plants, and doubtless many more remain to be found, but their number decreases daily as the forest is cleared in the name of "progress." That most of the drugs now in use are synthesized industrially from petrochemicals is no consolation. What are we going to do when oil finally runs out? Without seeming to understand what we are doing, or why, we humans are endangering the Earth system by impoverishing its heart, the biosphere.

ENERGY AND LIFE

Nature's way is to crumble to disorder. Unmaintained buildings fall to ruins, uncultivated fields turn to woods, your desk and sock drawer get messier and messier unless you spend the time and energy to keep things in order. In fact, it takes energy to maintain order in any natural system. Life is no exception. Life makes orderly carbohydrate molecules out of disorderly simpler molecules, and this process takes energy.

Ultimately, all the energy used by organisms comes from the Sun. Organisms that can get this energy directly from sunlight are call **autotrophs,** from the Greek words for self and feed, hence "self-feeders." Most autotrophs are green plants, which use sunlight energy to combine carbon dioxide from the atmosphere with water to form carbohydrate (CH_2O) by photosynthesis:

$$CO_2 + H_2O + \text{energy} = CH_2O + O_2$$

The energy stored in carbohydrate is later released by the organism through fermentation or respiration. Fermentation is a process in which carbohydrate molecules combine to form an alcohol plus carbon dioxide and energy. In respiration, carbohydrate combines with oxygen to form carbon dioxide, water, and energy.

Heterotrophs are organisms that are unable to use the energy from sunlight directly and so must get their energy by eating autotrophs or other heterotrophs. Thus, the biosphere is a consumer hierarchy in which autotrophs (primary producers) are at the base; **herbivores** (plant-eating heterotrophs) are on the next level up, and they in turn are devoured by **carnivores** (meat-eating heterotrophs) and **omnivores** (heterotrophs that eat both meat and plants) at the top (Fig. 15.1).

ECOSYSTEMS

Transfer of energy through the biosphere, from the autotrophic base to the omnivorous top, involves a lot of wasted energy. A given mass of autotrophs is necessary to support a smaller mass of herbivores, which

Figure 15.1 Plants (autotrophs) capture and look up energy from the sun in food molecules. Plants form the base of the food chain. Heterotrophs eat autotrophs to get energy. Zebras (herbivorous heterotrophs) feed on plants while lions (carnivorous heterotrophs) feed on the zebras.

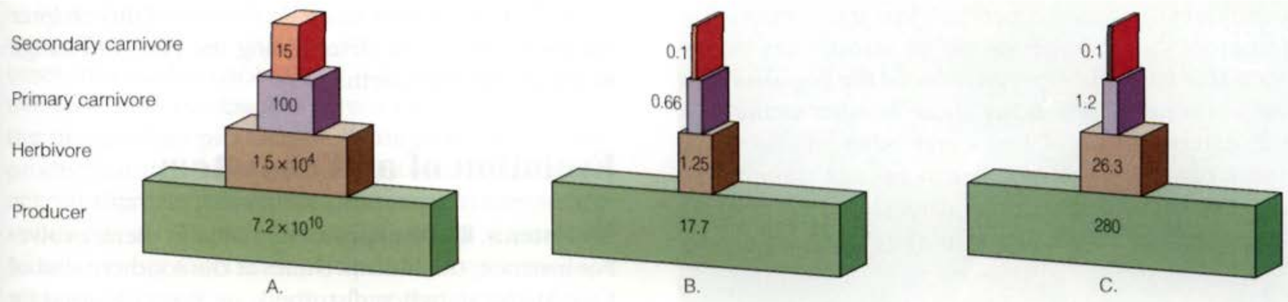


Figure 15.2 Trophic pyramids for a typical ecosystem. A. Individual organisms per square meter. B. The dry weight of biomass in grams per square meter. C. The dry weight productivity in milligrams per square meter per day.

in turn can sustain only a still smaller mass of carnivores. Because of this relationship, the hierarchy of organisms is called a **trophic pyramid** (Fig. 15.2).

The pathways hy which energy (as food) is moved from one trophic level to another are called **food chains.** An example of a food chain is given by the grass (autotrophs) that provides food for rabbits (herbivores), which in turn provide food for foxes (carnivores). Of course, there are many other animals and birds that catch and eat rabbits, so a trophic pyramid consists of many intertwined food chains. Even a simple trophic pyramid has quite a lot of food chains, which together form a **food web** (Fig. 15.3).

A trophic pyramid, together with its habitat (natural abodes of organisms in the pyramid), is called an **ecosystem.** An ecosystem may be as small as a pond or as large as an ocean basin. The *global ecosystem,* covering the whole earth and otherwise known as the *biosphere,* is the sum of all smaller ecosystems. Because both climatic and geographic barriers can also be ecological barriers, the global ecosystem is not as well connected internally as are the smaller ecosystems of which it is composed.

There is a strong interdependence among the corn-

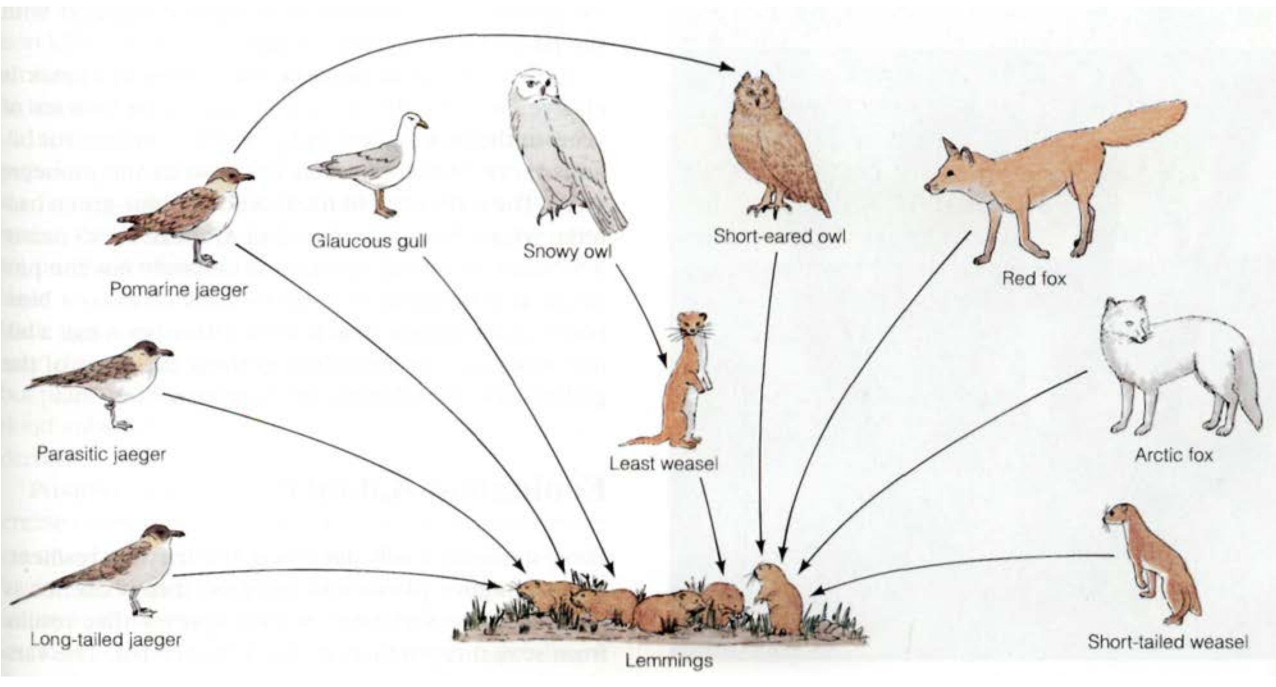


Figure 15.3 A simple food web: Lemmings and their predators in northern Alaska. Lemmings are herbivorous heterotrophs and eat plants that grow in the tundra. The foxes, owls, weasels, and jaegers eat lemmings. Note that Snowy Owls also eat Least Weasels and that Pomeraine Jaegers eat Short-eared Owls.

component will have repercussions on the others. For instance, the widespread use of insecticides on insects that feed on crops has reduced the population of insect-eating birds in many areas. Another example is the extermination of foxes and other predators on many British estates in order to protect game birds; because of the decline of predators, a scourge of rabbits afflicted British farms in the 1940s and early 1950s. See the "Guest Essay" at the end of this chapter for discussion about determining the rates of change in our global ecosystem.

A.

B.

Figure 15.4 The Indiana Dunes. A. Pioneer species start the process of building an ecosystem on a dune. B. The climax community on a dune.

Evolution of an Ecosystem

Ecosystems, like the species that inhabit them, evolve. For instance, the Indiana Dunes at the southern end of Lake Michigan, when disturbed, are first colonized by a few species of grasses that are able to tolerate dry and unstable conditions. These plants begin to stabilize the sand, allowing access to species that can tolerate drought but not a shifting foundation. Humus, the organic portion of soil, builds up in the sand, stabilizing it further, reducing its permeability, and slowing the rate at which water drains away; drought-intolerant plants are thereby able to grow. As plant diversity increases, animals, birds, and insects join the ecosystem. The culmination of this process is a growth of hardwood trees with all the attendant components of a hardwood-forest ecosystem.

The first organisms to establish themselves in an ecosystem, in this case the grasses, are called *pioneers;* the balanced community that ultimately develops is called the *climax* (Fig. 15.4). The process can be followed by observing dunes of different ages. In the Indiana Dunes, the time required to establish climax is on the order of 10,000 years. In this case, the evolution of the ecosystem is rapid compared with the pace of evolution in species.

For the global ecosystem, the notion of climax is elusive because the biosphere has been billions of years in the making and will probably continue for billions more. We don't even know what the pioneers were. The most ancient fossils known, blue-green bacteria, whose fossils are found in Archean rocks nearly 3.5 billion years old, are almost certainly not the pioneers. It is no more realistic to think of today's biosphere as the climax than it would have been half a billion years ago for a trilobite to think the same of the global ecosystem during the Cambrian.

Ecological Balance

Ecosystems are easily disturbed, but they are resilient. Among higher plants and animals, this resilience is rooted in the variability within species that results from sexual reproduction (see Chapter 16). The variability ensures that, barring catastrophes, every species has some members capable of surviving environmental change and that some of these survivors will be adapted to the new conditions.

In a climax ecosystem, the populations of the various species are in balance, or nearly so. In many cases, this *ecological balance* is maintained by negative feedback. Feedback is how a product influences the process that produces it. In **negative feedback,** production decreases with the growth of the product, so that the amount of product is consequently stabilized. A simple negative feedback control is found in every toilet tank. The product (water) is supplied from a valve to which a float is attached. As the water rises in the tank the amount of product increases, the float moves up and gradually shuts off the valve, stopping the production of more product.

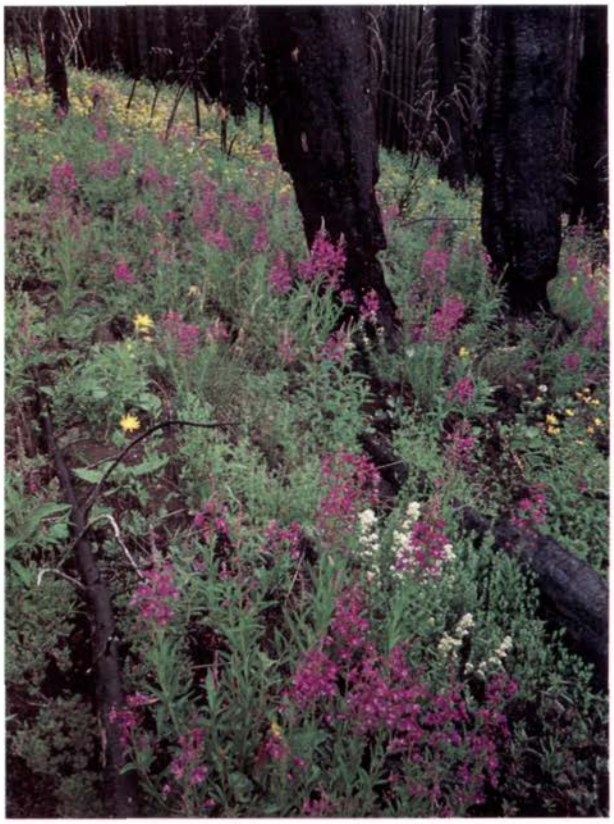
In ecosystems, population growth in each species is generally limited by certain intrinsic factors that provide negative feedback, so that eventually the population levels off or, as we say, a balance called **homeostasis** is reached. Food supply is an example of an intrinsic factor that limits population. Consider a population of rabbits on an island. The population increases until the limit of the food supply is reached, at which point the population stops growing because new additions starve. Other *intrinsic factors* besides competition for food include competition for living space, predation, and (in some cases) disease.

Extrinsic factors also operate in any ecosystem; examples are climatic change and certain catastrophic events. These do not provide any direct feedback control; the explosion of Mount St. Helens in 1980 is an example of an extrinsic factor. The explosion killed *all* trees within reach of the blast. It wouldn't have mattered whether there had been ten trees or ten million.

Intrinsic factors are more effective regulators of population than are extrinsic ones. However, extrinsic factors can exercise a good measure of control when they occur periodically and are not too extreme. Hard winters, for example, can reduce populations to levels from which they are ready to grow back again in spring; the annual flooding of a river may disrupt the floodplain ecosystem temporarily, but plant and insect populations will return when the flood subsides. Forest fires periodically burn out underbrush, but it grows back (Fig. 15.5).

Positive feedback occurs when production increases the growth of the product. Compound interest on a bank account is an example of positive feedback. The bigger the bank balance, the more you get in interest. Unchecked population growth is also an example of positive feedback. If a couple produces four offspring (two male, two female) that survive to reproductive age, then (setting aside the question of incest) their descendants in the tenth generation could amount to as many as 2^{10}, or 1,024, not to men-

A.



B.

Figure 15.5 Extrinsic events can greatly change an ecosystem. A. Yellowstone National Park after the great fire of 1988. B. Vegetation re-establishing itself. Yellowstone National Park two years after the great fire.

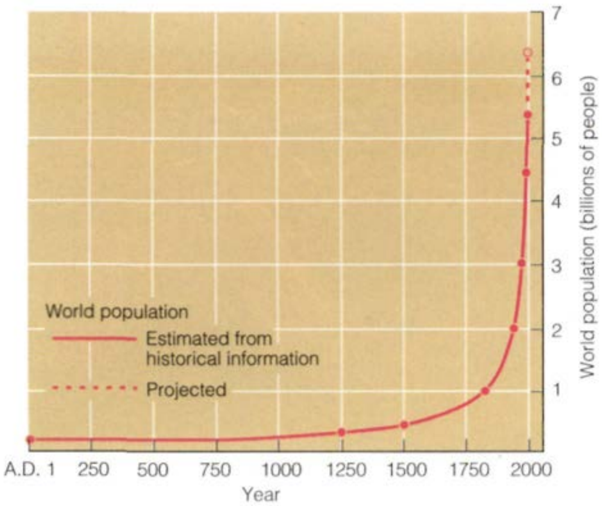


Figure 15.6 Growth of the human population.

tion any survivors from the nine generations along the line.

In the foregoing example of a growing population, the growth is said to be *exponential*, and in exponential growth the *rate* of increase at any instant is equal to the size of the population times a constant. Therefore, the bigger the base population, the bigger the rate of increase. If we write N for the size of the population, k for the constant, and N' for the rate of increase (say the number of individuals per year), then

$$N' = k \cdot N$$

In nature, the limited resources of an ecosystem place a limit on the population the ecosystem can carry. This limit is called the **carrying capacity.** As a population approaches the carrying capacity of its ecosystem, its growth is determined by intrinsic factors, particularly competition and predation, and homeostasis sets in.

The Earth's human population (Fig. 15.6) has grown by positive feedback because agriculture and medical science have permitted the survival of large numbers of people to reproductive age. How long this growth can continue is uncertain. Some scientists believe that agricultural resources are sufficient to support nearly double the present world population; but the effective use of resources to this end is subject to political and economic considerations that may prove difficult to get around. Efforts to relieve famine on the African continent, for example, have not thus far met with great success.

Population Cycles

If an ecosystem is subject to some extrinsic influence (for example, a harsher climate), the population will fall because of diminished resources (food) or exposure. If such an extrinsic factor operates periodically, the growth curve will rise and fall with the same frequency. There will be *population cycles.* In a predator-prey relationship, population cycles in the prey species will generate cycles in the predator population with the same period but sometimes with a delay in phase (see Fig. 15.7).

DIVERSITY

The *diversity of an ecosystem* is the number of species that inhabit it. A species is a population of individuals that can interbreed to produce offspring that are, in turn, interfertile with each other. Sometimes, however, species are hard to count, especially in fossil communities. In such cases, a higher unit in the taxonomic hierarchy (Table 15.1), such as the

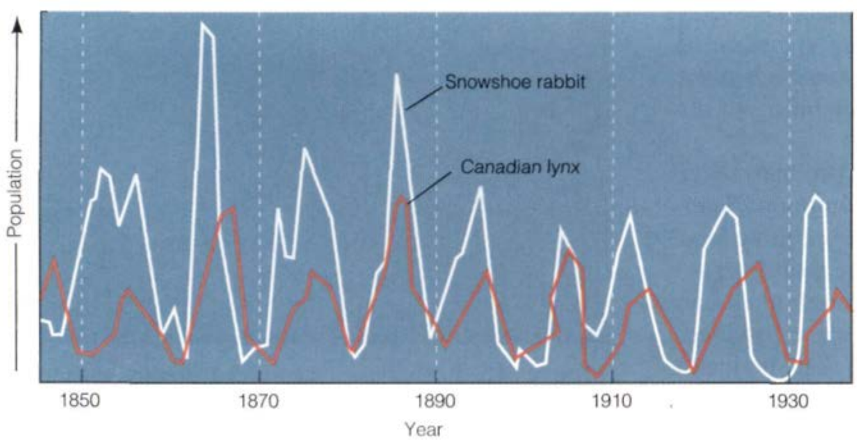


Figure 15.7 Population cycles in the Canadian lynx and its prey, the snowshoe rabbit.

Table 15.1
The Taxonomic Hierarchy

The systematic arrangement of plant and animal organisms according to their similarities. The example is a tiger.

Taxon[a]	Example	General Characteristics and Selected Representatives
Kingdom (a group of phyla)	Animalia	Multicellular, heterotrophic organism. Elephant, snail, human, tiger.
Phylum (a group of classes)	Chordata	Bilaterally symmetrical animals with a dorsal nerve chord. Elephant, eel, human, tiger.
Class (a group of orders)	Mammalia	Warm-blooded animals with hair and mammary glands. Elephant, mouse, tiger, human.
Order (a group of families)	Carnivora	Flesh-eating placental mammals. Dogs, foxes, lions, tigers.
Family (a group of genera)	Felidae	Rounded-headed carnivores walking on toes; retractable claws. Lions, tigers, bobcats.
Genus (a group of species)	Panthera	Large cats, 50 kg (110 lb) or more. Lions, tigers, leopards, jaguars.
Species	*Panthera tigris*	Striped coat, usually over 140 kg (310 lb). Bengal tiger, Siberian tiger.

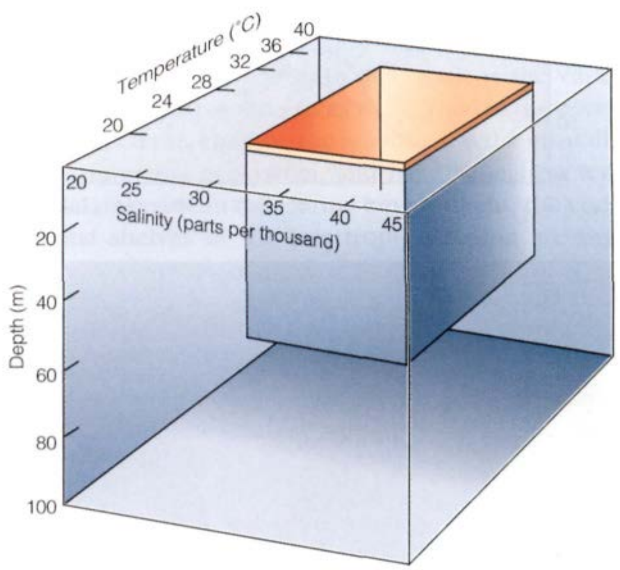
[a]A convenient way to remember the taxonomic order is to make up a mnemonic device using the first letters of each word: for example, *King Philip cares only for gifted singers.*

genus or family, is taken as the unit of diversity, but when possible there is an advantage in considering species, as will be seen below.

Ecological Niches

An **ecological niche** is the sum of the conditions that allow an organism and its offspring to sustain themselves and breed. We can use the common North American robin as an example. Robins feed on worms and grubs—you often see them pulling worms out of lawns; they can stand cold winters but not severe winters; and they need suitable trees and bushes for nest sites that can't be reached by snakes, cats, and other predators. A niche has many components and can be pictured as a multidimensional "resource space" in which each dimension is defined by one of the organism's interactions with its environment. A source of suitable worms and grubs in gardens and woodlets is one of the dimensions of a robin's resource space.

Multidimensional space can be represented only mathematically and is difficult to visualize, but two- or three-dimensional abstractions can often be used to illustrate the relationships between two or three selected variables that define a niche (Fig. 15.8). In such diagrams the niche always appears as a space, not a point, because the organism tolerates a *range* of values of each variable. In studying geometric representations of niche ranges, it is important to remember that they are addressing only a small number of the possible variables that define a niche.



Figurc 15.8 Geometric representation of a niche. This example illustrates the temperature-salinity-depth tolerances that bound the niche of a reef-forming coral. The graph tells us that the water temperature ranges from 16 to 40°C, for instance, but coral thrives only between 20 and 36°C. The water salinity ranges from 20 to 45 parts per thousand, but the coral lives only when the salinity is between 29 and 41 parts per thousand.

Competitive Exclusion

There is more anatomical and physiological uniformity within species than between them; and since anatomy and physiology are adaptive characteristics, it is unlikely that two species will be equally adapted to a particular ecological niche. Zebras and giraffes both live on the same open, grassy plain with scattered trees, but they have developed very different physical characteristics in order to fill their special niches. Therefore, in a given ecosystem, one species generally occupies one niche to the exclusion of

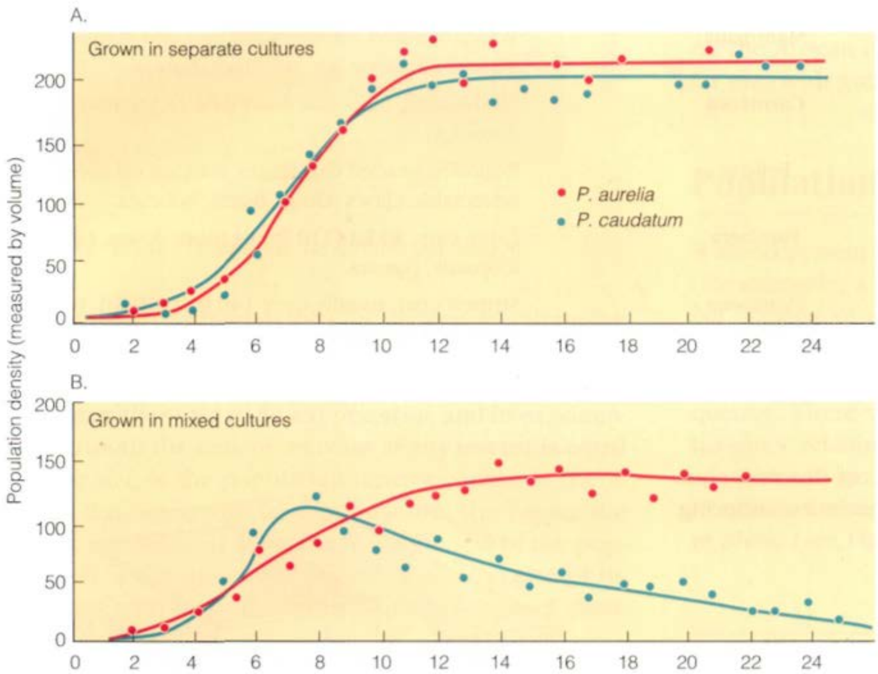


Figure 15.9 Growth behavior of two species of Paramecium. A. Grown in separate cultures so there is no niche competition. B. Mixed and grown in the same culture where they are competing for the same ecological niche. C. An example of a Paramecium (Paramecium cordata). Paramecium are tiny, ranging in length from 0.07 to 0.30 mm.

C.

other species. This reasoning, which is confirmed by observation, is generally known as the principle of competitive exclusion. We thus arrive at a new idea of species: a population that, in a particular ecosystem, occupies one, and only one, niche. Because of the principle of *competitive exclusion,* the best estimate of diversity in an ecosystem is to be found at the species level, as stated above.

Because of the large number of variables that make up a niche, niches may be quite narrowly defined. Swifts and swallows both prey on flying insects, as do insectivorous bats, yet all three species can be found in the same ecosystem. The two bird species hunt by day but at different altitudes, while the bats hunt at night. There are three niches here, all alike save for one detail; in the case of the two bird species, the differences in niche are hunting altitude, and in the case of the bat, the difference in niche is the time of day. Figure 15.9 illustrates competitive exclusion. It shows population growth in two species of microscopic unicellular organisms called Paramecium, (a) in separate cultures and (b) in the same culture. In (a) there is no competition between the two species, which grow about equally well until homeostasis sets in; in (b) the two species are competing for the same niche, and after a brief struggle one emerges victorious while the other dwindles.

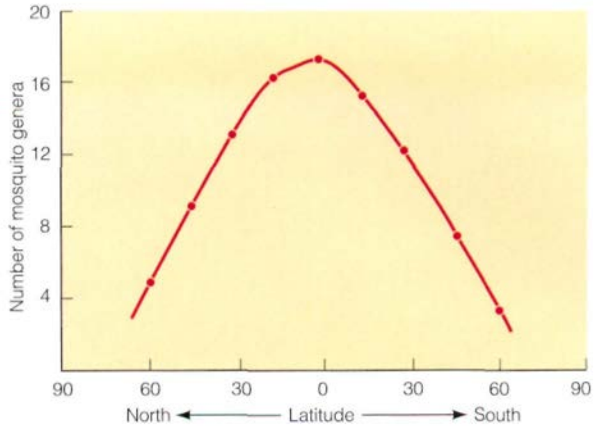


Figure 15.10 The number of mosquito genera as a function of latitude. Comparable patterns are observed in clams, turtles, parrots, foraminifera, termites, snails, frogs, snakes, lizards, crocodiles, reef-forming corals, amphibians, butterflies, and palms.

Diversity in the Global Ecosystem

Because diversity in an ecosystem is the number of species in it and because each species occupies a particular niche, the diversity of an ecosystem is the number of occupied niches in it. If one or more of the variables that define a niche strays outside the tolerance of the species that occupies it, that species will be stressed and possibly wiped out. Species with a wide range of tolerance (generalists) occupy large (i.e., broadly defined) niches; those with a narrow range (specialists) occupy small (narrowly defined) ones. The larger the niches in an ecosystem, the fewer the species in that ecosystem (in other words, a low diversity system) because each niche takes up more of the total resource space; the smaller the niches, the more species in the ecosystem (high diversity).

The Influence of Climate

Environments with a varying climate (which, for example, would cause a seasonal fluctuation in food supply) require wide tolerances in the species that live in them (a winter diet perhaps different from the summer one); environments with a constant climate can support species with narrow tolerances. In low latitudes, where climate conditions are more or less constant the year round, one might expect small niches and high diversity. In fact, such is the case, both on land and in the sea: as we learned in the essay that opened this chapter, the rain forest is the most diverse terrestrial ecosystem, and has the highest terrestrial diversity. In the marine environment, the continental shelves of the intertropical region are the most diverse ecosystems, and there we find the highest marine diversity. At higher latitudes, because the climate varies over a broad range, diversity falls off, and near the poles, it becomes seasonal, reaching a minimum in the polar winter when many birds and mammals have migrated to warmer latitudes. In a general way, then, diversity decreases from a maximum at the equator to minima at the poles (Fig. 15.10).

Even around the equator, however, there are substantial variations in diversity, especially among marine organisms of the continental shelves. Proximity to landmasses can produce marked differences in the "climate" of the ocean. Examples are the shape of upwellings of cold, deep ocean water, the influx of fresh water and sediment from rivers draining the continents, and the frequency and severity of monsoons. For instance, there are no coral reefs at the mouth of the Amazon, which lies on the equator: the temperature is right, but with all the fresh water from the river, the salinity is too low for coral growth. The Australasian shelf (latitude 20°S), however, is remote from continental influences and has a remarkable reef system with the highest diversity in the equatorial belt (Fig. 15.11). Some of this diversity even spills

Figure 15.11 The Great Barrier Reef on the continental shelf of northeastern Australia is one of the world's most diverse marine ecosystems.

over onto the shelves of the Indian Ocean, in spite of unstable climate there (monsoons) and heavy influx of detrital sediment.

The Influence of Provinciality

It might be imagined that, if the equatorial climate extended over the whole Earth, global diversity would be greater than it is today. The high latitudes would be lush with subtropical vegetation (as they were in the Late Cretaceous Period). But this reasoning does not take account of the possibility that, in expanding poleward, the equatorial habitats might support not new species but simply more of the same two million or so species that live there at present. Species that now live in colder climates, such as polar bears, penguins, skuas, moose, caribou, and the rest, would not exist, and global diversity would be impoverished by their absence.

Here, then, another factor in global diversity, as important as climatic stability, comes into play. It is **provinciality,** which is the extent to which the global ecosystem is divided into subsystems by barriers to the migration of organisms. These barriers can take the form of climatic gradients, seaways or mountains between landmasses, or land between seas.

Provinciality is high at present as a result of three of these barrier types, and so is global diversity. Recall that when the two Paramecium species were grown in separate cultures (Fig. 15.9), both flourished because, although they occupy the same ecological niche, they were prevented from competing for it. Thus, the more the Earth's ecosystems are separated from each other, the greater will be the global diversity because a given niche in each ecosystem can be occupied by a different species. For instance, there are several species of "anteater" (they really eat termites) living in South America, South Africa, southwestern Asia, and Australia because they are kept apart by geographical barriers and cannot compete with each other (Fig.15.12).

The Influence of Sexual Reproduction

Diversity at any time is controlled by the rate at which new species evolve and the average existence span of a species. Evolution is driven by mutations (see Chapter 16); sex merely mixes up the genes in a breeding population. In doing so, however, it confers on the population (species) a variability that gives it resilience to environmental vicissitudes, and this resilience helps to prolong the existence span of the


A.


B.


C.


D.

Figure 15.12 So-called anleaters (usually they eat termites) from different parts of the world are different species; they fill the same niches but are not competitive because they do not come in contact. A. Short-beaked echidna (Australia) B. Tamandua (Central America) C. Pangolin (Malaysia) D. Giant anteater (Venezuela)

species. This prolonged existence of itself increases global diversity, but it also increases the chances of adaptive mutations and, hence, the evolution of new species.

Even allowing for the incompleteness of the fossil record, diversity throughout most of the Precambrian was low. Organisms were making replicas of themselves, with an occasional mistake (mutation) that must, more often than not, have been lethal. What evolution did occur had to be accomplished by a few adaptive mutations among many failures. The potential for sexual reproduction arose with the advent of nucleated cells in the Late Proterozoic, after sufficient oxygen had accumulated in the atmosphere to allow

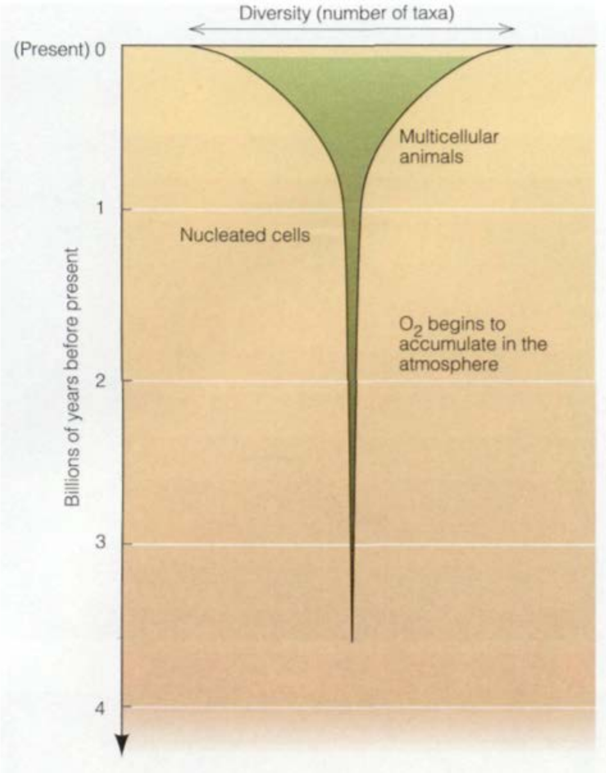


Figure 15.13 A simplified representation of global diversity through geologic time.

respiration. The marked increase in global diversity that followed (Fig. 15.13) was probably no coincidence.

Diversity in Major Taxonomic Groups

The history of taxonomic diversity is the history of niche occupancy, which can be understood in terms of evolutionary innovations by successful mutations on the one hand and availability of niches on the other.

Evolutionary Innovations

Through evolutionary innovations, a group may be enabled to move into hitherto unoccupied niches. Until late in the Silurian Period, for example, fish had no jaws. They lived by dredging sediment for their food. Then one group of them (the acanthodians) developed jaws from their anterior gill cartilages and so became able to graze on algae or eat other aquatic animals (Fig. 15.14). The "invention" of jaws opened to these fish an array of previously unoccupied niches, which they proceeded to fill in a rush that produced the most diverse of all the vertebrate classes.

In the Devonian Period, to take another example, the amphibians, by developing limbs from fins, brought vertebrates halfway onto the land. With skins permeable to water and with fishlike eggs and young to look after, however, amphibians could not fully adapt to this new habitat, and they remained essentially aquatic. The amniotic egg, in which an embryo is enclosed in a fluid-filled membrane, made the descendants of the amphibians, the reptiles, fully independent of the aquatic environment (although many returned to it and exploited it more effectively than the amphibians could). Reptiles quickly occupied, on land and in the air, niches that until then had stood vacant. Diversification of the flying reptiles is well illustrated in the fossil record. Skulls show that different species were adapted to different kinds of diet, even though all species (since their fossils are found in lake sediments) appear to have skimmed the water for their prey (Fig. 15.15).

Availability of Niches

Immediate diversification results when a new "invention" (jaws in fish, the amniotic egg in reptiles) has adaptive advantages for a new way of life. However, a modification that adapts a species to a niche that is already occupied may confer little immediate advantage. Even though the new species may be adaptively superior to the incumbent one, possession is nine-tenths of the law.

Such was the case with the mammals. Originating in the Triassic Period, mammals were from the start more attuned to the niches they now occupy than

Figure 15.14 Acanthodian fish were the first animals to develop jaws. This fossil acanthodian (Cheiracanthus murchisoni), found in Devonian aged rocks in Banffshire, Scotland, lived about 400 million years ago. The lower jaw is visible at the lower right-hand edge of the specimen. The fossil is about 5 cm long.

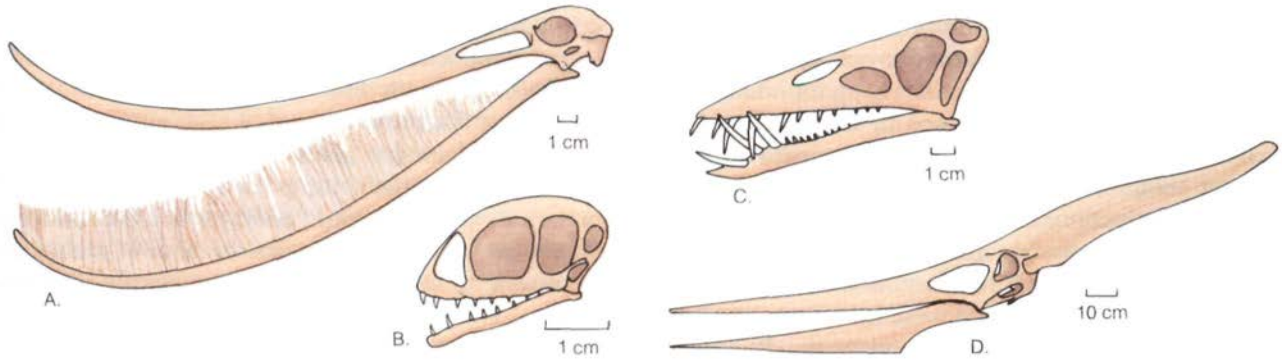


Figure 15.15 Diversity of Jurassic-aged pterosaur skulls. A. A species that was probably a filter feeder living on plankton. B. A species that was an inscclivore. C. and D. Carnivorous species. Fish bones have been found in the rib cages of the fossils of both species.

were the established occupants, the reptiles. With more capable brains, faster metabolism, a uterus to shelter the fetus, milk for postnatal nourishment, and parent-offspring bonding, the early mammals had greater potential to exploit reptilian niches than had the reptiles themselves. The great reptiles had command of the food supply, however, and the mammals, instead of growing larger than their therapsid ancestors, became smaller, perhaps small enough not to interest carnivorous dinosaurs (Fig. 15.16). And so they remained for 150 million years, until fate (an extrinsic factor, probably in the shape of a gigantic meteor impact) put an end to the reptile rule at the end of the Cretaceous Period. The vacated niches were occupied by mammals in a burst of diversification that began in the Paleocene Epoch and continued through the Eocene.

Diversity and Plate Tectonics

The evidence for a gigantic meteor impact at the end of the Cretaceous is persuasive but not conclusive. As discussed in the Introduction, a distinctive iridium-

Figure 15.16 Gasosaurus, one of huge flesh-eating Jurassic dinosaurs. Gasosaurus was up to 2 m high and 4 m long. Mammals that lived at the same time as (iasosaurus were small, about the size of mice or rabbits, and probably of little interest to the large dinosaur.

rich sediment is thought to have been formed as a result of a great meteorite impact that coincides with the Cretaceous-Tertiary boundary. A huge impact crater of exactly the right age has been found in Mexico, and the devastation caused by the impact must have been massive. As a result, many scientists today accept such an event as the probable cause of the massive extinction that coincides with the Cretaceous-Tertiary boundary. However, the greatest of all extinctions in the Earth's history, when some 95 percent of known fossil species were lost, marks the Permian-Triassic boundary at the end of the Paleozoic Era.

There is no convincing evidence that the Permian-Triassic boundary extinction had an extraterrestrial cause. Although both terrestrial and marine life were severely affected by the Cretaceous-Tertiary extinction, it was mainly marine species that suffered in the Permian-Triassic one. This earlier extinction is associated with the assembly of Pangaea, the supercontinent that existed briefly from Late Permian to Late Triassic time. Because the assembly of Pangaea came about as a result of plate tectonic movement, the Permian-Triassic extinction is sometimes referred to as a plate tectonic extinction. The plate tectonic movement in question was the assembly of the supercontinent Pangaea and the consequent loss of diversity led to the extinctions.

The Permian-Triassic extinction was associated with a marked drop in sea level caused by the dwindling of the ocean ridges that had driven the Paleozoic continents together. Consider the effect of these developments—continents massed together and sea-level drop—on the continental shelves. A simplified model (Fig. 15.17) shows that, if four continents of equal size are stuck together to make a single supercontinent, there is a 50 percent loss of coastline and, consequently, of continental shelf. A 100-meter (110 yd) drop in sea level will approximately halve the remaining shelf area so that the total shelf area is reduced to about 25 percent of what was there before: a drastic reduction in living space for shelf organisms. Furthermore, with four separate continents there must be at least four ecological shelf provinces, and more if the climatic gradient is taken into account. (Remember that the south polar region was glaciated in the Late Carboniferous Period.) When the four continents are joined together, the number of niches decreases because the shelves that remain become contiguous; Pangaea lay across the equator, forcing the warm, westward-flowing equatorial current into high latitudes, bringing heat from equator to poles. In other words, the Triassic climatic gradient was much less steep than the Carboniferous gradient had been (or, for that matter, than the present one is), reducing provinciality still further.

whether the reduction in shelf area and provinciality that attended the assembly of Pangaea can alone account for the Permian-Triassic extinction is uncertain. (There have been quite a lot of extinctions since the Precambrian that were not, of course, associated with supercontinent formation.) But it is reasonable to suppose that supercontinent formation must have made a major contribution to the extinction.

Another line of thought connects plate tectonics with diversity in reptiles and mammals. Reptiles began diversifying in the Late Carboniferous Period and had achieved most of their full diversity, some 9 of the 13 or so orders usually recognized, by the end of the Triassic Period 90 million years later. This amounts to about 0.1 order per million years. As noted earlier, the mammals originated in the Triassic Period but did not get a chance to diversify until the Cenozoic Era, producing about 30 orders by the end of the Eocene Epoch, a mere 30 million years later: 1.0 order per million years, or ten times the reptilian rate. It is reasonable to ascribe some of the mammals' diversity to the different circumstances in which rep-

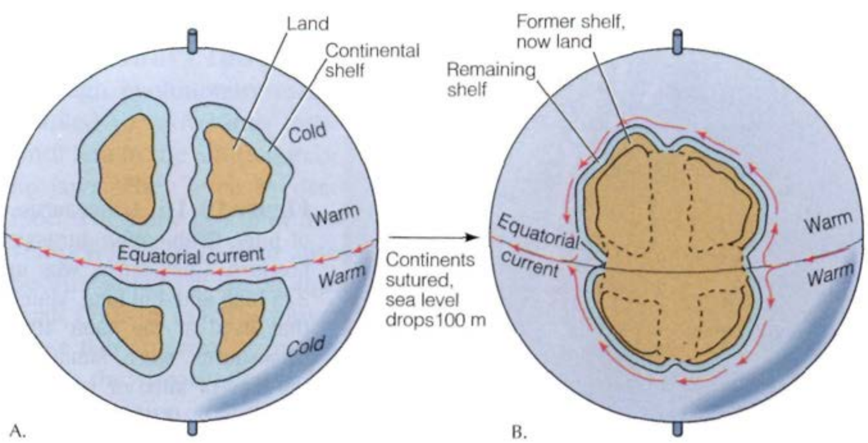


Figure 15.17 Idealized diagram of a hypothetical Earth with A. Four separate continents and an equatorial seaway and B. the four continents sutured together plus a 100 m drop in sea level. The resulting supercontinent lies across the equator, forcing the equatorial current poleward. The area of the continental shelf is drastically reduced, and the climatic gradient is weakened.

tiles and mammals diversified. Reptiles diversified on Pangaea, a single continent with a weak climatic gradient and therefore low diversity. Mammals diversified on seven separate continents following the breakup of Pangaea and faced a strong climatic gradient as the late Cenozoic ice age approached. The influence of plate tectonics on provinciality through the assembly and fragmenting of continents is nowhere better seen than in Australia, with its zoo of marsupial mammals that occupy niches that are elsewhere occupied by placental mammals.

A further example of diversity reduction by plate tectonics is found in South America, which was invaded by placentals from North America after the Isthmus of Panama was built by volcanic arc activity in the Late Pliocene Epoch. The placentals usurped the niches of indigenous marsupials (and a few more primitive placentals such as the liptoterns). One marsupial survivor, the redoubtable possum, made its way to the northern continent where, despite never having figured out the rules of the road, it holds its own comfortably in this placental stronghold by means of a high birth rate and omnivorous diet.

LINKS **WITH** THE **OTHER** GEOSPHERES

The Atmosphere

Among the geospheres, the atmosphere has had the greatest influence on the biosphere and has in turn been the geosphere most affected by the biosphere.

Oxygen

The Archean atmosphere provided the **anaerobic** (without oxygen) conditions necessary for starting life. Then, when life became autotrophic, the oxygen it produced created the oxygenated atmosphere that we know today. This last step was of the most profound significance for life. Without respiratory metabolism, it is unlikely that any of the attributes of "higher" life could have developed: organelles, which are specialized cell parts such as the cell nucleus; multicellular bodies with differentiated organs; and sexual reproduction. And without the ozone (O_3) of the upper atmosphere, there would have been insufficient protection from short-wave radiation for life to get a foothold on the land.

Nitrogen

Oxygen can be thought of as life's gift to itself, but the atmosphere's main gift to life has been nitrogen, a relatively unreactive gas that seems to have been a minor constituent of the atmospheres of all the planets. Nitrogen's main role in life is in proteins, where it plays a key part in the chemical reaction by which amino acids polymerize to form these most fundamental of all life's molecules. Much energy is needed to split N_2 molecules into usable reactive atoms; some of this energy is provided by lightning bolts, around which oxides of nitrogen form and are rained onto the Earth's surface. The only organisms that can split N_2 molecules are the so-called *nitrogen fixers,* mostly bacteria that live in a **symbiotic** relationship (two or more dissimilar organisms living in a mutually beneficial relationship) with certain plants (especially legumes such as clover, alfalfa, and peas), providing them with nitrogen in assimilable forms. When organisms die and decay, the nitrogen is returned to the atmosphere, partly as molecular nitrogen (N_2), *by denitrifying bacteria* which regain the activation energy that split the original molecules.

Carbon Dioxide

A third atmospheric gas vital to the biosphere is carbon dioxide, CO_2, the source of the carbon used by autotrophs to make carbohydrates by photosynthesis. It is present in today's atmosphere in a concentration of about 300 parts per million, and it is gradually increasing because of the burning of fossil fuels. The primary source of this CO_2 is the mantle, from which it issues through volcanoes. How much is recycled from subducted surface materials and how much is new is uncertain. Nearly all of the Earth's CO_2 is locked up in carbonate rocks (limestone and dolostone) and as fossil carbon in detrital sediments. (If you look ahead at Fig. 18.7, you will see a fine example of the kind of limestone in which CO_2 is locked up.) The little that remains in the atmosphere is vital to the autotrophs at the base of the global food web. Together with water vapor, it is also a climate-moderating influence ("greenhouse gas") without which the climatic gradient would be too steep to allow the existence of the sorts of temperate-climate ecosystem we know today.

The Hydrosphere

Water is indispensable for the biosphere. The hydrosphere is the water source for terrestrial ecosystems and the habitat of aquatic ecosystems. It mediates most of the gas exchange between aquatic ecosystems and the atmosphere, and it supplies to aquatic systems the essential elements that terrestrial systems derive from the soil. As with terrestrial systems, a good part of the essential elements used by aquatic organisms is recycled from dead organisms. In shallow environments (rivers, most lakes, and the continental shelves), resources are recycled fairly rapidly; in the deep-sea environment, the cycles are longer and slower. For example, phosphorus is supplied by the continental drainage to the coastal seas, where most of it gets locked up in the ecosystems of the continental shelves. Pelagic organisms (those that float or swim) of the deep ocean sink when they die, taking their phosphorus with them to the ocean floor. This phosphoros is returned to the surface only after a long time and in only a few localities where deep ocean water wells up. The shelves teem with marine life; the deep-sea surface, by comparison, is a sort of aquatic desert.

The Lithosphere

Ultimately, of course, the lithosphere contributes all of the minor elements necessary to life and some of the major ones. The most important element it contributes, however, is phosphorus. This element plays two essential roles: in the form of sugar-phosphate units, phosphorus forms the helical framework of the DNA molecule, and as adenosine mono-diphosphate and adenosine triphosphate it serves as the currency for all of life's energy transactions. The biosphere's phosphorus supply is released from the lithosphere by weathering. As the growth of most ecosystems is limited by phosphorus availability, weathering is an important regulator of total biomass. The lithosphere gives up its phosphorus rather grudgingly from relatively insoluble minerals such as apatite (calcium phosphate).

Soil is the base of nearly all terrestrial ecosystems as well as itself an ecosystem of innumerable species and great complexity. Without the lithosphere there would be no soil. The inorganic constituents of soil supply a complete habitat to the organisms of the soil ecosystem. In addition, these inorganic constituents act as a substrate, together with water, nutrients, and essential trace elements, to the larger ecosystem (forest, prairie, or whatever it may be) above.

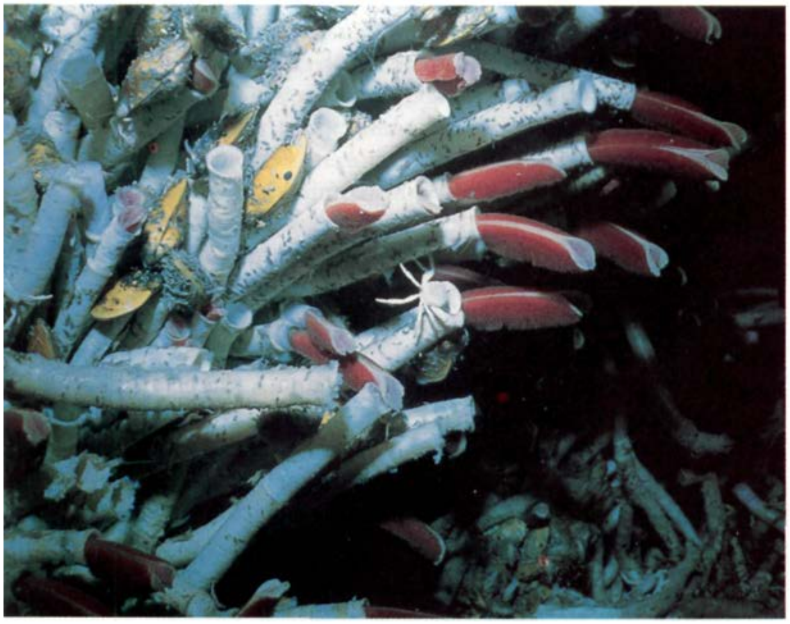


Figure 15.18 An extraordinary ecosystem around a black smoker at a depth of 2500 m in the Pacific Ocean. Bacteria that derive their energy inorganically through the oxidation of H_2S brought up by the smoker are the autotrophs. The helerotrophs that live directly or indirectly on the autotrophs include worms, clams, starfish, crabs, and skates.

The Mantle

Through volcanism the mantle is the source of the carbon that (as CO_2) is the starting point for synthesis of organic compounds in the biosphere. Although today most of this carbon has likely been recycled, the mantle itself must be given credit for the initial supply that made the pre-biotic surface environment rich in CO_2.

Another very interesting contribution of the mantle to life on Earth, discovered not very long ago, is chemical energy for a small group of autotrophs. In an expedition to the Galapagos rift in 1977, geologists in the small research submarine *Alvin* discovered submarine hot springs that are now known as *black smokers*—chimney-like structures of anhydrite ($CaSO_4$) and sulfide minerals from which issued plumes of hot (350°C, or 662°F) water darkened by a suspension of sulfide minerals of iron, zinc, and cop per. Around these vents (and others like them discovered subsequently on other segments of the midocean ridge), at depths up to 3000 meters (1.9 mi), lives a community of beard worms (*Pogonophora*), clams, starfish, crabs, and other invertebrates, as well as deep-water skates (Fig. 15.18). All these animals are chalk white, except for the beard worms, which are reddish pink. Because the deep sea is devoid of sunlight, autotrophs cannot get their energy from sunlight through photosynthesis. The energy comes instead from chemical compounds in the hot water.

Bacteria are the primary producers in the unusual and diverse ecosystem around submarine hot springs. Living in complete darkness, they are **chemoautotrophs,** deriving energy from the oxidation of hydrogen sulfide (H_2S) in the water discharged from the smokers.

Biological diversity has kept growing since it first took off in the Late Cambrian, and it seems reasonable to suppose that it will continue to increase. But when we try to imagine niches yet to be filled, it turns out to be more difficult than we have perhaps expected. Would anyone have predicted the existence of a diverse submarine ecosystem fueled by heat energy from the mantle instead of sunlight?

LINKS WITH HUMAN ACTIVITY

Humans change and perturb the biosphere in many ways. The signs of change are everywhere around us—in the building of cities, the spread of pollutants, the cutting of forests. A few examples will serve to illustrate the changes.

Chemicals in the Food Web

The famous nineteenth-century English biologist Thomas Huxley (1825-1895) portrayed the gardener as perpetually at war with nature. The moment the gardener's back is turned, weeds begin to grow in the garden and stifle what is planted there. So the gardener's job is to be constantly vigilant and to defend the garden against nature's incursions, protecting pampered flowers and vegetables from competition with the wild species that are better adapted to the prevailing natural conditions. On a larger scale, the same is true of the farmer. For the past 50 years, nature has been increasingly kept at bay with herbicides and pesticides for the benefit of food crops and of the people who eat them.

There are difficulties with herbicides and pesticides, however. In moving up the trophic pyramid in a food web, some chemical elements become concentrated in the tissues of organisms. Because some of the elements are toxic, what may appear to be harmless quantities of toxic substances can lead to serious consequences (see "A Closer Look: Analysis of a Disaster").

Soil Erosion

The soil of the continents serves as the base from which nutrition in all the terrestrial ecosystems (including the human one) is derived. "Primitive" human cultures have generally known how to conserve this essential resource (there have been exceptions); modern peoples, going for short-term gain, are wasting it.

The soil is the part of the lithosphere most vulnerable to erosion. James Hutton, at the end of the eighteenth century, said that God designed the rock cycle so that weathering could replace the soil lost by erosion. Before humans began intensive farming and grazing, the average rate of soil erosion was about 10 billion tons a year for all the continents, and soil production and loss rates were in balance.

The present rate of erosion is about 25 billion tons a year (Fig. 15.19). With the average rate of soil formation being 10 billion tons a year, the current ero-

Figure 15.19 Erosion as a result of overgrazing and poor farming practice in Ethiopia.

A Closer Look

Analysis of a Disaster

In 1956 doctors described curious symptoms in residents of Minamata Bay, Japan. The symptoms were indicative of severe toxification of the central nervous system. It was soon discovered that patients afflicted with the disease were all fishermen and their families, and that they regularly ate fish caught in Minamata Bay (Fig. C15.1). Further study showed that mercury was involved and that a chemical plant on Minamata Bay, which produced basic chemicals for the plastics industry, used mercury salts for catalysts and was the source of the problem. The amounts used were tiny, however, so the question then became, how could the fishermen accumulate so much mercury in their bodies?

Organisms eliminate certain poisons at a rate that depends on the concentration of the poison in their tissues. If M is the amount of poison in the organism, M' the rate of elimination, and k a rate constant,

$$M' = -k \cdot M$$

If an organism is given a single dose M_0 of the poison, a graph of M against time looks like Figure C15.2. M decreases in a regular manner that after 100 days one-half of the initial dose M_0 remains; after another 100 days M has again been halved, and so on. The halving period of 100 days is called the *half-life* of the poison in the organism. The rate constant is related to the half-life in a simple way:

$$k = \frac{\text{natural log of 2}}{\text{half-life}} = \frac{0.7}{100}, \text{ or } 0.007 \text{ per day}$$

Suppose now that instead of receiving a single dose, the organism begins a diet in which a constant amount of poison is ingested every day. At the start, the organism's poison content M will rise fairly steeply. As it builds up, however, the rate of elimination will increase until a steady state is reached where M' (rate of elimination) becomes equal to the rate of intake. (See Fig. C15.3 which, you may notice, is the mirror-image of Fig. C15.2.) Strictly, the steady state is reached only after an infinite

Figure C15.1 A victim of pollution. Physical deformity arising from mercury poisoning as a result of eating contaminated fish from Minamata Bay, Japan.

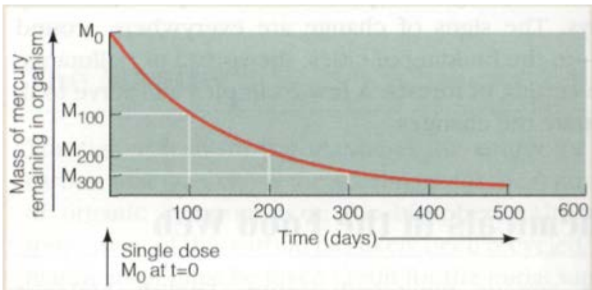


Figure C15.2 Decrease of mercury in the body of an organism that eliminates half of dose present each 100 days.

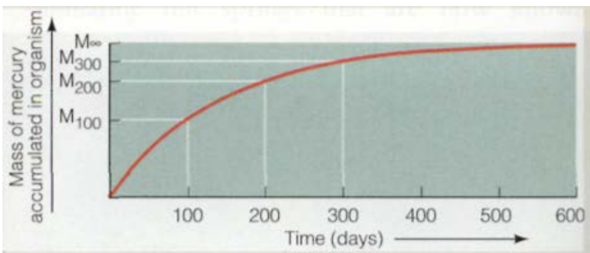


Figure C15.3 Buildup of mercury in the body of an organism that receives a steady diet of mercury each day. As the amount in the body builds up, so the rate of elimination is increased. A balance is reached when the intake equals the elimination rate.

Table C15.1
Increase in Mercury Concentration Up the Food Web

Trophic Level	Mass of Individual	Daily Diet of Individual	Daily Mercury Intake	Steady-State Mercury Content	Mercury Concen./Tissues (parts per billion)
Algae					10
Herbivores	100 g	10g algae	10^{-7} g	1.4×10^{-5} g	140
Primary carnivores	1000g	1 herbivore	1.4×10^{-5} g	2×10^{-3} g	2000
Secondary carnivores	10,000 g	1 primary carnivore	2×10^{-3} g	0.29 g	29,000

time, but it is near enough for practical purposes after about six half-lives.

The poison content of an organism at steady state (assuming the poison has not killed it first) is the rate of intake divided by the rate constant k:

$$M = \frac{\text{daily intake}}{0.007}$$

Imagine an ecosystem with the trophic levels and diets given by the first two columns (mass and daily diet of individual) in Table C15.1. Now suppose that a small amount of mercury is introduced into the system and is taken up by the algae, producing a mercury concentration in them of 10 parts per billion (1 part in 10^8). The daily diet of a herbivore, 10 grams of algae, contains 10×10^{-8} or 10^{-7} grams of mercury. At steady state, the herbivore will contain $10^{-7}/0.007$ or 1.4×10^{-5} grams mercury, and the mercury concentration in its tissues will be $1.4 \times 10^{-5}/100$, or 1.4×10^{-7} (140 parts per billion). We can now complete Table C15.1 by calculating the mercury concentrations in the species at successive trophic levels in this food web.

Thus, in moving three trophic levels up the food web from the algae, the mercury gets concentrated by a factor of nearly 3000 ($2.9 \times 10^{-5}/10^{-8}$), reaching a concentration of 29 parts per million in the secondary carnivores: more than 500 times the acceptable limit of 0.05 ppm for food in the United States

There are some weaknesses in this oversimplified model. For example, the half-life varies from one sort of tissue to another. Furthermore, it is unlikely that all of the prey species would survive the six half-lives necessary to reach the steady state because some would get killed and eaten with lower concentrations of mercury in them. Finally, such an ecosystem might not be closed; fish that swam in after having fed elsewhere could be mercury-free. Nevertheless, in Minamata Bay top-level carnivores were found with as much as 50 ppm of mercury in them. A 60 kilogram Japanese eating 200 grams of such fish a day would accumulate 24 ppm of mercury in her tissue.

sion rate could remove most of the world's topsoil in less than a century. The calculation is crude; but even if it's off by a factor of 5, a century from now there will be much less freedom to choose to grow crops in the world's productive agricultural zones (the humid, midlatitude prairies) than there is today.

Mine Drainage

The subterranean resources of the earth (metals, coal) exist in an oxygen-free environment. When this environment is opened up to the atmosphere by mining, some compounds (mainly pyrite) are oxidized to acids (such as sulfuric acid). These acids then acidify the surrounding environment, especially streams, whose ecosystems are traumatized by the high acidity. In addition, the more acidic water of these streams mobilizes toxic elements, such as lead and cadmium, held in the rocks.

Acid Rain

Coal, mostly formed on ancient floodplains near sea level, contains varying amounts of sulfur (as pyrite derived from marine sulfate) which, on burning, is converted to sulfur dioxide. This latter compound forms sulfurous acid by combining with water, and eventually the sulfurous acid is oxidized by the atmosphere to sulfuric acid, which gets rained out onto the land downwind of the coal-burning installations. Protection of neighboring land is secured by building high smokestacks, and the result is often the export of pollution to distant environments. Thus, the United States exports acid rain to Canada, and Western Europe (especially Britain) exports it to Scandinavia.

Like acid-mine drainage, acid rain traumatizes aquatic ecosystems by acidifying the water. The long-term effects depend on geological conditions. In limestone country, the acid is quickly neutralized by the reactive carbonate rocks, and little harm may result. Unfortunately, Canada and Norway, two of the principal victims of acid rain, have many granitic terranes, where neutralization is slow and releases aluminum (from feldspars and other aluminosilicate minerals) into the water. There it reaches concentrations that are toxic to wildlife.

Acid rain has been reported to have deleterious effects on terrestrial ecosystems as well. Conifers have allegedly been killed in New England, and agricultural crops damaged in a variety of places. Further studies are needed to verify these reports.

Eutrophication of Surface Waters

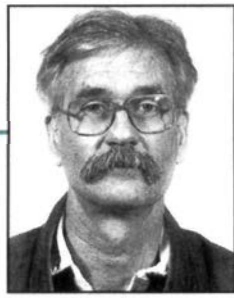
Big coastal cities produce large amounts of sewage. Whether raw or bacterially oxidized, this waste water usually contains fairly high concentrations of phosphorus (mostly from detergents) and nitrogen, in the approximate proportions that are optimal for plant growth. All too often, especially in the lakes or seaways where the cities are located, the result is **eutrophication,** which means bodies of water with a high level of plant nutrients and consequently vigorous growth of algae (Fig. 15.20). As the algae die, they sink to the bottom, where their decay creates an oxygen demand that quickly makes the environment *anoxic* and asphyxiates all aerobic organisms living in it. A new generation of algae follows on the first, and so on, while masses of dead organic matter pile up on the bottom, sometimes thickly enough (as in the case of the Chesapeake Bay in the late 1960s) to obstruct shipping lanes.

Overturn of the water column in winter can cause fish kills as the anoxic water comes to the surface, and whatever resources (fish, oysters, shrimp) the ecosystem once offered to the human community are reduced to nil. In lakes, the rotting vegetation builds up until it reaches the surface and serves as the foundation for a bog ecosystem. In this case, human activity is merely speeding up nature's own ways. Lakes are ephemeral features. If they have an outlet stream, they drain as it erodes its bed; if this does not happen first, they fill with sediment from streams that feed them, and become bogs.

Figure 15.20 Algal bloom due to eutrophication on a pond in western New Jersey.

Guest Essay

Mapping the Earth from Space



Compton Tucker is a physical scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland. His research involves using satellite remote sensing to study desertification, tropical deforestation, and temperate forest issues. He has a B.S. in biology and M.S. and Ph. D. degrees from the College of Forestry, all from Colorado State University.

Changes in vegetation tell earth scientists a great deal about our planet's biosphere. Deforestation in tropical rainforests, for example, can affect factors ranging from biological diversity to concentrations of greenhouse gases in the atmosphere, but the ability to monitor rainforests on the ground or by airplane presents difficult challenges. The spread of deserts causes different consequences but similar challenges for earth scientists: while desertification is often accompanied by droughts and famines, it is difficult to determine desertification trends, or even if desertification exists. However, since the 1970s, scientists have begun to employ remote sensing techniques to monitor the earth from satellites. These information sources allow observation and study of the global biosphere from several hundred kilometers in space.

Although these satellites can be classified into two broad categories, the remote sensing instruments on all perform the same basic functions. These instruments detect vegetation changes in the biosphere by collecting reflected and/or emitted radiation from the earth. All remote sensing instruments, old and new, measure light reflected in the visible and near-infrared spectral regions. Although instruments on newer satellites can measure thermal, or longer-wave infrared radiation (which can be used to determine temperatures), measurements in the visible and near-infrared spectra are most useful for the purpose of observing global biospheric dynamics. These measurements, taken from various points in space, help scientists determine the photosynthetic capacity for specific areas of the biosphere. This information, collected over time, produces "seamless" continental and global estimates for changes in vegetation.

Landsat satellites have been used recently to measure deforestation rates in the Amazon of Brazil, which comprises about one third of the planet's total rainforests. Because deforestation must be measured over small areas (with dimensions of less than 500 m), Landsat satellites, which take measurements over areas with widths of 30 or 80 m, are the only satellites that can provide such focused data with high degrees of accuracy. Estimates of the annual tropical deforestation rate in Brazilian Amazon varied from between 20,000 to 80,000 km^2, and Landsat satellites measuring areas with widths of 30 m provided coverage of this forest's entire 5,000,000 km^2 area. The data the Landsats provided, once processed with a geographic information system, or GIS (a powerful computer program that transforms raw data into graphs and maps), indicated an annual deforestation rate between 1978 and 1988 of around 15,000 km^2/year, substantially lower than previous estimates. This information allowed scientists to estimate more accurately the direct destruction of different species' habitats by deforestation.

In addition, the refinement of this data through a GIS helped scientists discover patches of forest surrounded by deforested areas. Such identification can provide important biospheric information for determining the indirect impacts of deforestation on biological diversity, all of which hinge on the relationship between tropical forest edge zones and areas of deforestation. One indirect impact involves the isolation of species from their former range, caused by deforested areas around patches of forest. Easier access to forested areas from adjacent cleared areas by non-forest plants, animals, and hunters represents another indirect impact on biological diversity.

Remote sensing techniques can be used not only to detect long-term changes in the biosphere, such as deforestation, but also day-to-day rates of desertification. Rather than use highly focused Landsat satellites, research into desertification has been conducted with National Oceanic and Atmospheric Administration satellites (NOAA). These satellites employ advanced very high resolution radiometer (AVHRR) instruments, which measure areas with widths of 1 to 4 km. While these instruments cannot measure with the detail of Landsat instruments, they can make daily measurements of the entire earth.

Earth scientists have used AVHRR data to study the Sahel of Africa which is a broad transition grassland between the Sahara desert to the north and the more humid savannas to the south. This area has attracted scientific interest because of periodic drought and concern over possible expansion of the Sahara to the south. AVHRR techniques have been used to identify grassland total biomass production, which gives an indication of the amount of vegetation in a designated area, and to identify thermal data related to clouds, which gives an indication of precipitation amounts. By taking ten-day averages over a period stretching back to 1981, scientists can now determine specific areas of vegetation with precipitation amounts of less than 10 cm per year. Identifying these areas helps trace the expansion and contraction of deserts, and these studies have shown that desertification is a dynamic, rather than permanent, process.

Such data can be combined with historical information about crop yields, grazing conditions, and severity of droughts to provide early warning for food security purposes. Conditions for areas of interest can be monitored through the growing season and compared to the previous 13 years of historical data to identify areas of food and/or fodder shortfalls. This type of famine early warning is much more objective than depending upon in-country reports. It provides rapid and accurate information, which can help identify where to send relief when required and ensure that the most needy areas are not overlooked. Exactly the same data as those used to investigate desert expansion and contraction are used for food security or famine early warning purposes.

Information collected from remote sensing techniques helps earth scientists determine biospheric changes in the present, perhaps with an eye to the past. However, the previous examples of deforestation and desertification illustrate the substantial predictive power achieved through interpretation of this data. Whether investigating future impacts of decreased biological diversity or the prevention of tragedy caused by famine, the ability to view the earth from space provides an invaluable supplement to everyday human observation.

Summary

1. Autotrophs, which get their energy from inorganic sources, form the bottom of the food chain.
2. Heterotrophs feed on autotrophs or other heterotrophs, thereby creating a trophic pyramid in which energy is moved upward from level to level via the food chain.
3. An ecosystem is a trophic pyramid plus the habitat in which the pyramid exists. Ecosystems can be small, like a small pond, or large, like an ocean basin.
4. The global ecosystem, also known as the biosphere, is the sum of all the smaller ecosystems on the Earth.
5. Ecosystems evolve, being first colonized by pioneer species and eventually becoming a balanced climax community. The concept of a climax is unclear for the global ecosystem because the biosphere is at least 3.5 billion years old and will probably endure for billions more years.
6. Ecological balance is commonly maintained by negative feedback.
7. The resources of an ecosystem limit the population; the limit is known as the carrying capacity. Just what the Earth's carrying capacity is for humans is an open question.
8. The diversity of an ecosystem is the number of different species in it. Each species in an ecosystem occupies a niche. Thus, large niches mean fewer species (low diversity), and small niches mean more species (high diversity).
9. Diversity is influenced by many factors, among them climate, sexual reproduction, evolutionary innovations, provinciality, and niche availability.
10. Plate tectonics, through the assembly and subsequent breakup of the supercontinent Pangaea, played a major role in reducing diversity at the Permian-Triassic boundary.
11. The biosphere influences and interacts with all of the other geospheres. The spheres most directly influenced are the atmosphere and the hydrosphere.
12. Humans continuously change and perturb the biosphere. Examples of change are the clearing of forests, the release of acid-mine drainage, the spread of toxic chemicals through the food chain, soil erosion, and eutrophication of lakes and streams.

Important Terms to Remember

anaerobic (p. 405)
autotrophs (p. 391)
carnivores (p. 392)
carrying capacity (p. 396)
chemoautotrophs (p. 407)
ecological niche (p. 397)
ecosystem (p. 393)
eutrophication (p. 410)
food chains (p. 393)
food web (p. 393)
herbivores (p. 392)
heterotrophs (p. 392)

homeostasis (p. 395)
negative feedback (p. 395)
omnivores (p. 392)
positive feedback (p. 395)
provinciality (p. 400)
species (p. 396)
symbiotic (p. 405)
trophic pyramid (p. 393)

Questions for Review

1. What do organisms need energy for, besides moving around?
2. What is the difference between autotrophs and heterotrophs? Name two different classes of heterotrophs. Is a dog an autotroph or a heterotroph?
3. In what way is carbon dioxide vital to the biosphere?
4. What is meant by an ecosystem? Describe an ecosystem you are familiar with in the area in which you live.
5. What is a climax ecosystem?
6. How does negative feedback regulate population in an ecosystem?
7. What is positive feedback? Is the growth of the human population till now an example of it?
8. If a couple produces two offspring, essentially replacing themselves, how many descendants would they have within five generations?
9. What is meant by *diversity* in an ecosystem?
10. What is an ecological niche? What is the relation between a niche and a species?
11. What is the principle of competitive exclusion? Give two examples.
12. What factors influence the diversity of the *global* ecosystem?
13. What factors influence diversity in major taxonomic groups (classes, for example)?
14. How has plate tectonics influenced global diversity?
15. Compare diversity in Mesozoic reptiles and Cenozoic mammals, and suggest an explanation for any difference you find.
16. Briefly describe the fauna (animal life) of the Galapagos rift. What is the energy source in this ecosystem?
17. Suggest some ways in which humans have changed or are changing the global ecosystem.
18. What is the source of acid rain and what effects does the increased acidity of rainfall cause?
19. Describe the process by which lakes and rivers near big cities become anoxic.

Questions for A Closer Look

1. Why do some toxic elements (such as mercury) become more concentrated in animal tissues as they move up the trophic pyramid?
2. If the half-life of a toxin is one year and the daily intake of aquatic algae leads to a toxin concentration of 100 parts per billion, what will be the toxin level in a herbivore that eats 20 grams of algae a day?

Questions for Discussion

1. The deer population has increased dramatically in many regions of North America over the past half-century. Research the reasons for the increase. Is the increase an example of positive or negative feedback?
2. What chemical elements in the lithosphere are necessary for life? How are the elements released from the lithosphere and made available to the biosphere?
3. Why save the spotted owl? Present both sides of the question.
4. Ecologists have recently become concerned about the currently unexplained reduction in frog and toad populations in many ecosystems around the world. Research this topic and discuss the possible reasons for the reduction. Are the factors involved intrinsic or extrinsic?

CHAPTER 16

Evolution of the Biosphere

The veldt in southern Africa as it would have appeared to early hominids 1.5 million years ago. The zebra-like animals on the right are quaggas (Equus quagga) and the large antelope on the left is a hartebeest (Rabaticeras porrocornutus). Both the quagga and the hartebeest are now extinct.

Back from the Dead: Fact or Fiction?

Two red-headed birds live on the island of Sado, Japan. They face certain extinction because they are the only two of their species still alive and they are now too old to breed. When the birds die, parts of their bodies will be stored in liquid nitrogen. The hope is that the bird's genetic code can be saved until scientists discover how to re-create a crested ibis by implanting the code in a closely related species such as an Asian ibis.

The crested ibis story may seem a bit far-fetched, but, of course, bringing extinct species back to life using samples of genetic codes preserved in fossils is what one of the most successful movies of all time, *Jurassic Park*, is all about. The crested ibis tissue will be stored under near ideal conditions for the preservation of the birds' genetic code, and so, if research on implantation is successful, crested ibises may some day be brought back to life. But is it reasonable to expect that the codes of extinct species are preserved in nature where conditions are anything but favorable, and could those codes be used to raise fossils from the dead?

Genetic codes are recorded in the configurations of atoms in DNA (deoxyribonucleic acid), the huge organic molecule that is present in all living organisms. The configurations, and therefore the codes, are slightly different in each species. When an organism dies, its DNA soon begins to break down into smaller molecules under the attack of water, oxygen, and natural radiation, and so some of the details are inevitably lost. But does it all break down, or are there circumstances where some DNA might be preserved? That question has set in motion some exciting research.

Proof that DNA can last at least 100 years came in 1984. Two scientists at the University of California, Berkeley, discovered DNA in museum samples of a quagga, a zebra-like animal from Africa that became extinct more than a century ago. Not only was some DNA left, but the genetic code it contained was very similar to the codes of related, though living, animals. The similarity of codes reduces the likelihood of contamination from some other organism.

In 1985 a Finnish scientist extracted DNA from a 2500-year old Egyptian mummy; in 1989 the same scientist found DNA in the bones of a 13,000-year-old ground sloth and a 40,000-year-old woolly mammoth. In each case, the DNA that remained was less than perfect. The original large molecules had broken into smaller fragments, but even so, enough information remained in some of the fragments for the codes to be read.

Encouraged by the sloth and mammoth reports, scientists started looking for DNA in much older samples. In 1990 there was a report of DNA fragments from a 17-million-year-old fossil magnolia leaf from Idaho. This was followed in 1993 by astonishing reports of DNA from a fossil bee about 25 million years old and from a fossil weevil about 120 million years old. The bee and the weevil had been trapped and preserved in amber.

Research is proceeding, but many scientists are still skeptical about the very old DNA samples. One probkm, the skeptics argue, is the possibility of contamination. The DNA found in ancient fossils may have come from bacteria or some other much younger contaminant. Another problem is degradation. Even if bits of DNA do remain in some fossils, they are probably insufficient to re-create extinct species, and they may even be insufficient to be sure of the differences between extinct and related living species. The scientific jury is still out on all of this research. If you follow the news over the next two decades, you will probably learn *whether Jurassic Park* is just science fiction, or whether there really are samples of fossil DNA from which extinct organisms may one day be raised from the dead.

WHAT IS LIFE?

This chapter addresses three tantalizing but still unresolved issues. First, we consider the most intriguing question of all: How did the biosphere begin or, to phrase it more simply, how and where did life start on the Earth? The second issue we address is the evolution of life from its humble, microscopic beginnings to the complex biosphere that exists today. The third topic is the effect of the changing biosphere over geologic time on the evolution of the Earth system. We start our discussion by addressing a fundamental question: What do we mean when we say something is alive?

What are the essential differences between living and nonliving matter? Most of us seldom stop to ask this question. A dog runs about and barks, while a stone lies still and silent. What about a dog and a potato? The differences must obviously be something else besides the ability to run and bark. The simplest organisms known are bacteria (Fig. 16.1). How do we know they are alive? We can tell because they do the same basic things complex organisms do. They reproduce and they carry out chemical reactions, collectively called **metabolism,** by which they grow and they maintain themselves. Life as we know it, then, down to the simplest organism, is the ability to grow, to reproduce, and to metabolize.

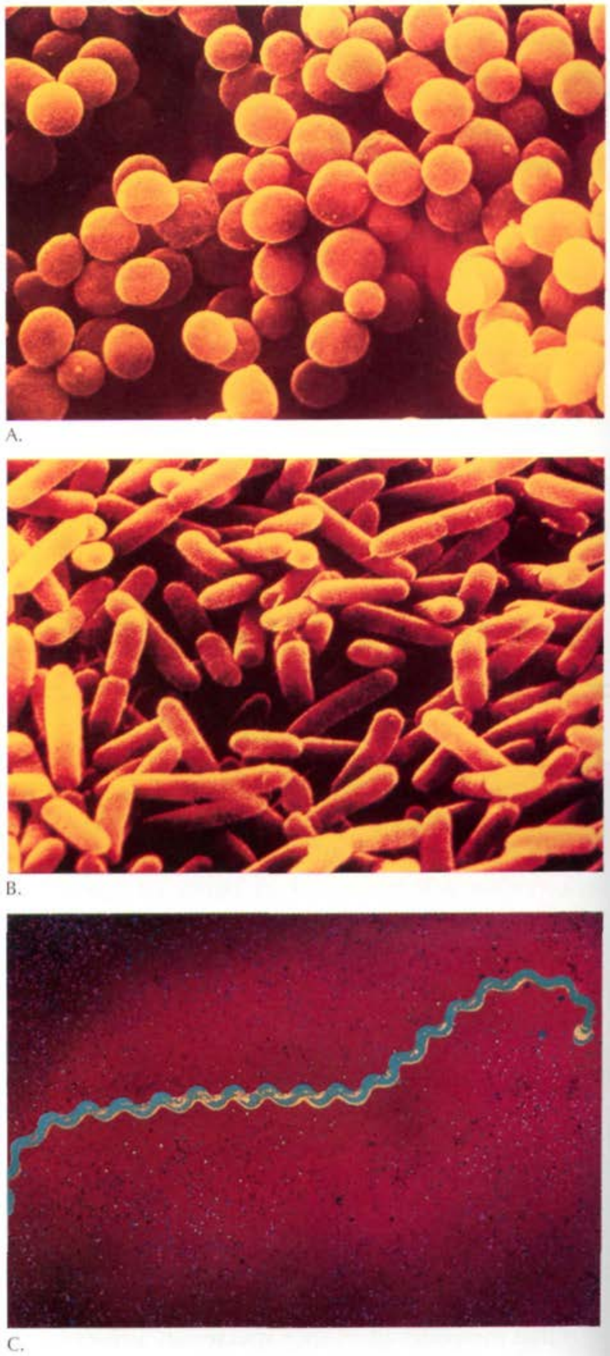


Figure 16.1 Examples of the shapes of bacterial cells. A. Sphere. After dividing, cells sometimes remain attached to one another, producing a chain. B. Rods. C. Spirals.

Growth, one of the abilities of living organisms, involves the ordering and organizing of atoms and small molecules to make larger molecules. Ordering can happen in two ways. The first is by *polymerization*, which is the stringing together of small molecules, like beads on a necklace, to make large chain- or sheet-like molecules. Most of the plastics we use in the home are polymers made in this fashion. The second way of ordering things is by *crystallization*, which, as discussed in Chapter 4, is the packing of atoms or molecules in ordered geometric arrays. Crystallization is what happens when the randomly ordered H_2O molecules in water vapor form ice crystals on a cold window pane.

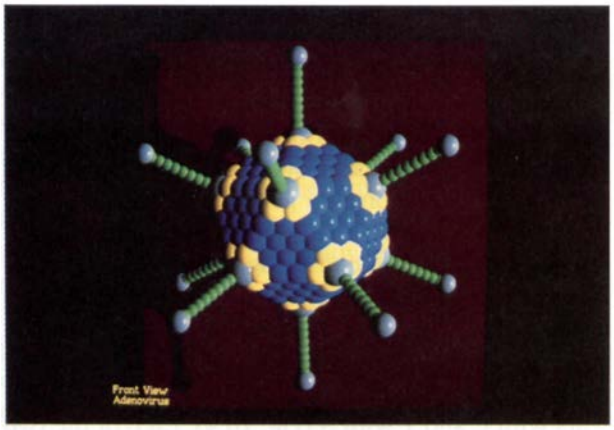


Figure 16.2 Model of a virus. The colored spheres and the rods sticking out like antennae are proteins. The proteins enclose a core of either DNA or RNA.

In both cases, polymerization and crystallization, an ordered pattern of atoms or molecules is replicated throughout the structure. There is an important difference between the two systems, however. Polymerization *absorbs* energy, while crystallization *releases* energy. (This is the latent heat of crystallization discussed in Chapter 12).

Growth in all living matter involves the polymerization of small organic molecules to form large organic molecules *(biopolymers),* and so growth requires a source of energy. The most important group of small molecules where polymerization is concerned is the **amino acids.** The amino acids are organic acids containing an amino (NH_2) group; the large molecules formed through their polymerization are **proteins.** Proteins are essential constituents of all living cells.

Crystals have the power of growth, but unlike living organisms they have no metabolism. Viruses, those nasty things that cause colds and flu and many other diseases, fall in a shadowy area somewhere between living and nonliving matter. Atoms in viruses are ordered like those in crystals (Fig. 16.2), and viruses can reproduce themselves by using the replication machinery of some other organism. But viruses cannot reproduce by themselves, nor can viruses metabolize.

Viruses are "almost" living, and bacteria are the simplest living organisms. Could either one of these, a virus or a bacterium, have been the first living thing? The answer seems to be "no." Life cannot have begun as a virus since a virus requires a living host. Nor is it likely that life began as a bacterium because, even though bacteria are the simplest known organism, they are nevertheless very complex—too complex, in all probability, to have been the first living organism. The complexity of bacteria has been tellingly described in the following way: to build a model of a bacterium using small glass beads for atoms, you will need a large cathedral to work in and to house the finished model, which will weigh as much as an ocean-going ship and take 35 years to assemble with a crew of 1,000 workers. (The real living bacterium assembles itself in about half an hour.) The operating complexity of this simplest organism can be grasped by reflecting that a bacterium sustains life by hundreds of chemical reactions, and these reactions are all essential. If one of them fails, the organism sickens and dies. Thus, bacteria, as we have said, are just too complex to have been the first living organisms.

So, although today the boundary between the living and the nonliving is quite a sharp one, there must have once been some prebacterial form of life that has (so far as we can tell) left no record of itself. Did it metabolize? If so, what and how? If not, was it really alive? The search for evidence about how life began is one of the most intriguing and challenging quests facing scientists.

Another feature of life, perhaps its most characteristic and one that sets it apart from the nonliving world, is that it is in a continual state of change. A crystal of quartz that grows today is the same as a crystal that grew 3 billion years ago. Living organisms, however, change over time. The kinds of animals that grow today are completely different from the primitive organisms that lived 3 billion years ago. Today's organisms have arisen as a result of change. This change affects not only the individual organism in its life cycle of growth, reproduction, and decline, but also the population to which it belongs, which is always responding to environmental fluctuations and occasionally undergoing more radical change in the generation of new species. The changes that organisms and populations undergo through time are called **evolution.**

Because organisms and populations evolve, the biosphere itself must have evolved through time and must still be evolving today. This means that the Earth system must also have evolved. One way evolution of the Earth system can be seen is in the cyclic transfer of chemical elements among the Earth's four reservoirs (the solid earth, hydrosphere, atmosphere, and biosphere). The cycling patterns of some elements have changed dramatically through time, but the most marked changes have been in the recycling rates of biologically important elements such as carbon, oxygen; nitrogen, and phosphorus. Elements essential to—and whose cycles are strongly influenced by—the biosphere are said to have **biogeochemical cycles.**

THE IMPORTANCE OF PROTEINS

The basic structural unit of all living organisms is the cell (Fig. 16.3). Many bacteria are *unicellular* (that is, they consist of one cell), but most organisms are *multicellular* and consist of hundreds to trillions of cells—the larger the organism, the more cells it contains. The growth of new living matter involves the construction of new cells. Cell growth follows specific plans and each kind of cell has its own special plans. Full details of a growth plan are passed from cell to cell, generation after generation. The passed-on information is stored in **deoxyribonucleic acid (DNA)**, a biopolymer consisting of two twisted chainlike molecules held together by organic molecules called *bases.* There are four bases in DNA: guanine (G), cytosine (C), thymine (T), and adenine (A), and the bases hook together to form *base pairs;* G always pairs with C and T with A (Fig. 16.4). Each of the two strands of DNA thus carries four kinds of base pairs, and the sequence of base pairs along the chain can be varied almost infinitely. Like the barcodes used in supermarkets, the sequence of bases in DNA is the code that stores information. The stored information provides a cell with a reference library of how to carry out the activities of life, such as the details of reproduction, growth, and maintenance, including the polymerization of protein. The DNA codes are read and executed by **ribonucleic acid (RNA),** a single-strand molecule similar to one-half of a double DNA chain.

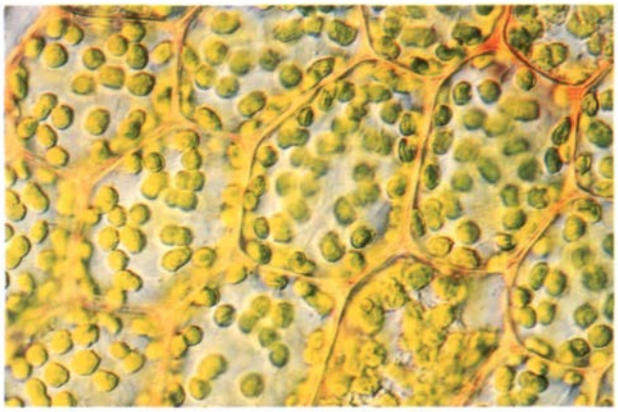


Figure 16.3 Cells from a multicellular organism. The yellow lines are membranes that enclose each cell. The green spheres are organelles called chloroplasts which help the plant to photosynthesize. Each cell is about 0.06 mm across. A multicellular organism is a cooperative grouping of cells, each of which can perform all the activities associated with life.

Proteins have three essential functions: (1) they form tissue such as muscle, ligaments, hair, and blood

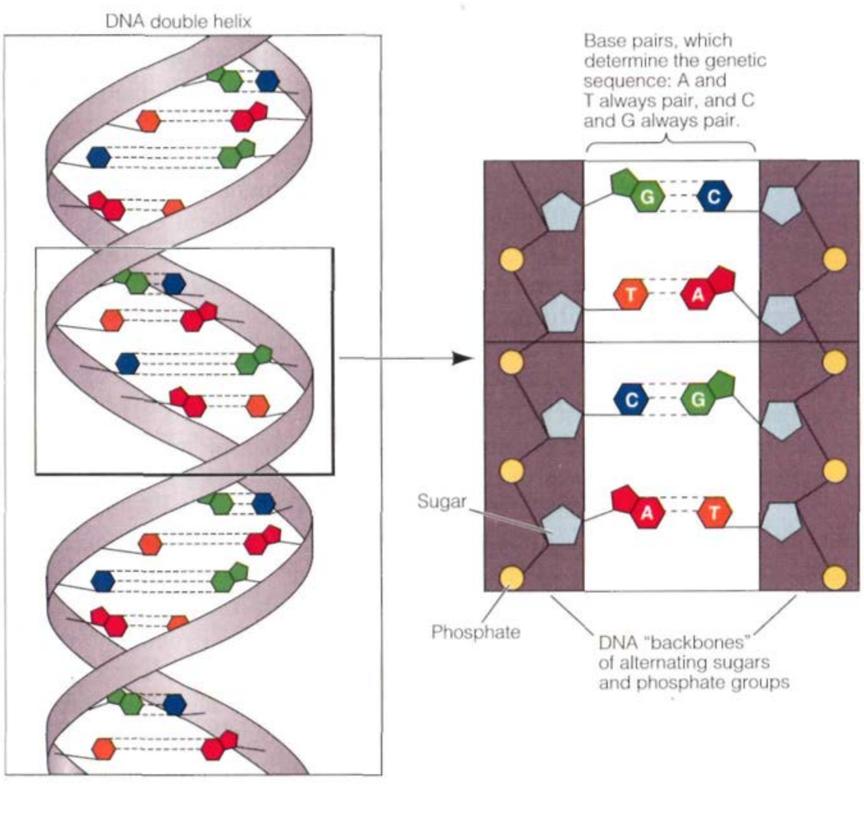


Figure 16.4 The twisted, chain-like molecule of DNA. The two strands of DNA are joined by four organic molecules, called bases: adenine (A), cytosine (C), guanine (G), and thymine (T). The binding molecules always join as base pairs: A always Joins to T, and C always joins to G. The sequence of bases is the code that directs the activities of a cell.

cells; (2) they provide patterns for laying down mineral structures such as shells and bones; and (3) as **enzymes** they catalyze (speed up) all chemical reactions in the organism, including those necessary for the assembly and polymerization of amino acids to form additional proteins. (Metabolic reactions would be far too slow to sustain life in the temperature range it can tolerate, were it not for the catalytic effect of enzymes, of which there are many hundreds in even the simplest organism.)

The key role played by proteins leads to two important conclusions. The first is that the inherited characters of every organism, be it a bacterium or an elephant, are determined entirely by the information encoded in its DNA: these include its anatomy through the structural proteins it makes and its physiology through the enzymes it needs. The second conclusion is that proteins, the primary products of the reproduction process, are (as enzymes) also indispensable to it, because the instructions encoded in DNA can be neither read nor executed without their help. Without enzymes, metabolism cannot proceed either. So how did something as complex as life get started, and which came first—proteins that make growth possible or DNA that carries the plans for growth?

The interdependence of vital processes extends far beyond the chicken-and-egg conundrum of DNA and proteins. It tells us that, organizationally, bacteria must be about as far from the origin of life as we humans are from bacteria. Yet there are fossil bacteria 3.5 billions years old, and geochemical evidence based on biochemical cycles suggests that life goes back at least 3.8 billion years, less than a billion years short of the origin of the Earth itself. Two of the great unanswered puzzles center on how life could have started and then on how it evolved to an organism as complex as a bacterium in less than a billion years.

HOW DID LIFE BEGIN?

Whatever may have been the initial state of matter destined to become alive, we can specify three steps that must have been accomplished on the way to the complex life forms we know today: (1) **chemosynthesis,** which is the synthesis of small organic molecules such as amino acids; (2) **biosynthesis**, which is the polymerization of small organic molecules to form biopolymers, in particular proteins; and (3) the development of all the complex chemical machinery, including DNA and RNA, needed for replication. We will deal with the first two steps in this section. Because the third step covers the entire question of the development of life on the Earth, we will treat selected aspects of it later in this chapter.

Chemosynthesis

We focus on chemosynthesis first because it must have been the first step taken on the way to life. In 1923, a Russian scientist, Aleksandr Ivanovich Oparin, hypothesized that simple organic compounds may have been synthesized from a primitive atmosphere of ammonia (NH_3), methane (CH_4), water vapor (H_2O), and hydrogen (H_2), with energy for photochemical reactions supplied by lightning bolts or the Sun's ultraviolet radiation (Fig. 16.5). Thirty years later, a U.S. scientist, Stanley Miller, carried out an experiment designed to check Oparin's hypothesis. He recovered some amino acids and other simple compounds from such a mixture of gases through which electric sparks had been passed. Subsequent work along the same lines succeeded in synthesizing all 20 of the important protein-forming amino acids, as well as other biologically important compounds (for instance, adenine, one of the four bases in DNA).

Research over the years following the work of Oparin and Miller has exposed a difficulty—it is very doubtful that methane, ammonia, and hydrogen were major constituents of the Earth's primitive atmosphere. As discussed in Chapter 12, it seems likely that the dominant gases of the early atmosphere were carbon dioxide, nitrogen, and water vapor. Recent investigators have therefore turned their attention to the photochemical synthesis of organic molecules from CO_2, N_2, and H_2O with small admixtures of methane and ammonia. The main difficulty encountered in these investigations is that the yield of amino acids is minuscule and that the molecules that do form quickly disappear as a result of oxidation.

Because the basic organic molecules of life would have been hard (but not impossible) to synthesize on the Earth, some scientists hypothesize that organic molecules may have arrived, ready made, from some other part of the solar system or even from the galaxy beyond the solar system. The hypothesis has attracted attention for two reasons. First, astronomers have demonstrated that there are lots of small organic molecules in interstellar space. Second, among the many kinds of meteorites that fall on the Earth there is one class, called *carbonaceous chrondrites,* that contains a liberal amount of small organic molecules. If inter-

Figure 16.5 Is this the environment in which the ingredients of life were made? Lightning bolts discharge through volcanic gases over Mount Pinatubo, Philippines. Oparin suggested that similar discharges in the early atmosphere created small molecule organic compounds from which the large molecules needed for life were subsequently created.

stellar dust, carbonaceous chrondrites, or both, fell on the early Earth, perhaps they provided the starting molecules for life. Wherever the first organic molecules formed, the evidence is clear that there would have been enough around in the early years of the Earth. The big problem is the next step: How did the early molecules polymerize to become "life" molecules?

Biosynthesis: The Central Problem

Oparin believed that the products of chemosynthesis collected as an organic "soup" in the surface waters of the primitive Earth. The "soup" concept works just as well for organic matter from interstellar space or meteorites as it does for organic matter created in the atmosphere. Most scientists therefore agree with at least some aspects of the "soup" hypothesis.

According to Oparin, some of the less soluble organic compounds in the "soup" would clump together as droplets like those of butterfat in milk. Such *droplets,* made in the laboratory from dispersions of oil in water, are hollow microballoons with a double-layered wall enclosing a droplet of water. They can grow and divide, but here their resemblance to living cells stops. They cannot build biopolymers, nor do they show any metabolic activity.

The main problem with the Oparin "soup" scenario is that when amino acids polymerize to make proteins, water is eliminated: in an aqueous environment, this is a difficult, if not impossible, feat. However, if amino acids are dehydrated and heated, polymerization does happen. The resulting polymers, *protenoids* containing up to about 200 amino acids, will aggregate into chains and will divide, although they have no replication machinery. Protenoids may have formed on the early Earth by the drying out of some "soup" along ancient shorelines and subsequent polymerization by solar or volcanic heat. Polymerization might also have occurred if organic molecules had adsorbed onto the surfaces of clay minerals and the reactions had taken place there. Regardless of how polymerization occurred, whether by drying and heating or by adsorption on clay surfaces, protenoids have no metabolism. Whether such protenoids were capable, under some conditions, of further evolution and of becoming "alive" is not known.

Black Smokers

Since the discovery in 1977 of the submarine hot springs called "black smokers" (see Fig. 1518), a number of scientists have suggested these as possible sites for the origin of life. In this hypothesis, the first organic molecules were formed on the surfaces of pyrite grains that form around the vents. In this way, the pyrite rather than clay served as a concentrating mechanism. Prebiotic evolution began with reactions that took place in organic layers as thin as one molecule thick. Reactions in the organic layer were akin to metabolism, and the products were the first cells. When completed, the cells became detached from the mineral substrate. It is important to remember that this "black smoker" origin is simply a hypothesis and that no firm evidence for the process exists. These supposed earliest cells are not to be confused with the unicellular, autotrophic bacteria that live around black smokers today and that now serve as the base of the food chain in this strange ecosystem (Chapter 15).

Panspermia

Because the organic molecules needed for life can be formed in space, some scientists have gone one step further and hypothesized that life may have developed in space and arrived, ready made, from some other part of our galaxy. Biopolymers, it is hypothesized, could perhaps have grown in space from small molecules and formed organisms there. Such organisms, moving on the interstellar winds of fortune, could have reached other parts of the galaxy, including planets. Most would have fallen on stony ground and perished, but conditions on the Earth were hospitable and allowed the further evolution of life.

The hypothesis of the supposed origin of life in space, followed by a diaspora to various parts of the galaxy (including the Earth), is known as **panspermia.** One of the difficulties with hypotheses about life originating through biosynthesis and evolution on the Earth is, as we mentioned earlier, the short time available. The attractiveness of the panspermia hypothesis is that it extends by a few billion years the time available for biosynthesis of the earliest biopolymers because life could have originated before the solar system formed. It is hard to imagine, however, that conditions in interstellar space could have favored biosynthesis. Any location with a suitable temperature range would have been subject to concentrated shortwave electromagnetic radiation unless it was on a planet with an atmosphere that filtered out the radiation. Such radiation, though it favors the formation of small organic molecules, is deadly to biopolymers. Furthermore, the dispersion of matter in outer space is so thin that the formation of polymers is very unlikely in any case. Like most ideas about the origin of life, the panspermia hypothesis is hard to test.

Early Life and Oxygen

We have looked at several ideas about the origin of life, and we have seen that they do not agree about much beyond the general premise that small organic molecules could have been synthesized without much trouble and that somehow and somewhere they polymerized to form the biopolymers needed for building and replicating cells. There is one further point of general agreement: early life was anaerobic (i.e., lived in the absence of oxygen). Ever since the famous French scientist, Louis Pasteur (1822-1895), patiently demonstrated that spontaneous generation of life does not occur today, it has been understood that the conditions for the origin of life must have differed from those in which it now continues because most life today is aerobic (i.e., lives in the presence of oxygen). This is so even though it is well known that life's molecules are vulnerable to oxygen. One of the wonders of aerobic life is that it can use oxygen for metabolism without burning itself up in the process. As we will see farther on in this chapter, there is evidence in the geologic record that the early atmosphere contained little or no free oxygen. Even so, how aerobic life developed from anaerobic life is one of the intriguing and still unsolved puzzles about the biosphere.

THE ORGANIZATION OF LIFE

Earlier in this chapter, we pointed out that the cell is the fundamental organizational unit of life. A cell is a complex grouping of chemical compounds enclosed in a membrane. The membrane separates the materials inside from the environment outside and facilitates the controlled exchange of materials and energy between the cell and its environment.

All organisms are composed of one or more cells. Even though an organism may consist of many cells, an individual cell can be removed and kept alive and healthy in a culture. Such cultured cells have all the mechanisms of life. They can take in and digest nutrients, excrete their wastes, absorb oxygen in order to get energy by respiration, grow, reproduce, and move around.

Procaryotes and Eucaryotes

Cells may be smaller (0.01 to 0.02 mm, or 0.0004 to 0.0008 in.) or larger (0.05 mm to a few centimeters, or 0.002 to 1.5 in. or larger, in rare cases), but large or small, all cells are of two kinds, procaryotic or eucaryotic. **Procaryotic cells** (from the Greek *pro* - before and *karyote* = nucleus, hence before a nucleus) are generally small and simple in structure. These cells house their DNA with all its genetic information in a poorly demarcated part of the cell (Fig. 16.6A). The main body of the cell, which is called the **cytoplasm**, lacks distinctly defined areas in which the various cell functions are carried out. Most importantly, the portion of the cell that houses the genetic information is not separated from the cytoplasm by a membrane. All known procaryotes, living and fossil, are bacteria.

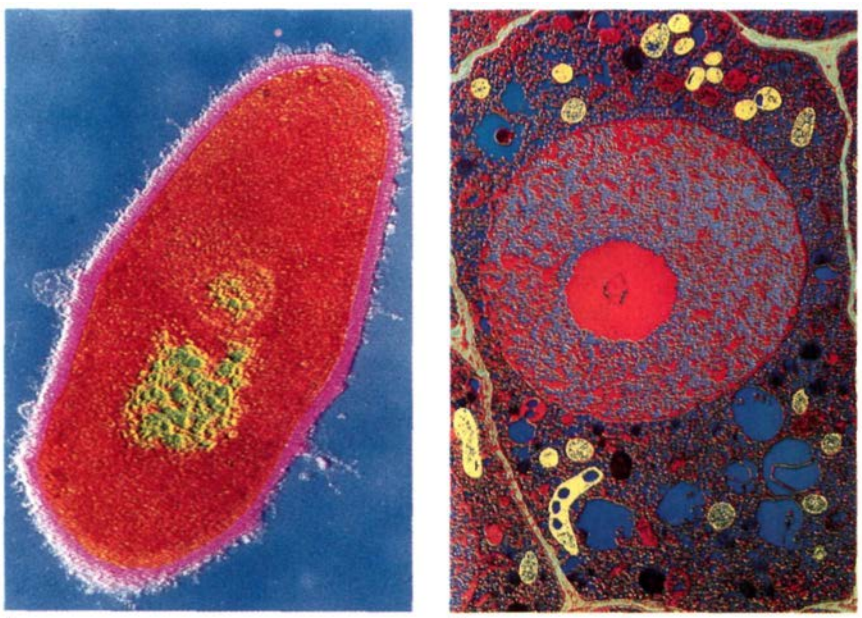
Eucaryotic cells (from *eu* = true, hence with a true nucleus) are larger and more complex than procaryotic cells. Their genetic information is housed in a well-defined nucleus that is separated from the cytoplasm by a membrane (Fig. 16.6B). The cytoplasm in a eucaryotic cell contains a variety of well-defined cell parts, called **organelles,** each having a particular function in the operation of the cell. We humans, and most other living things, are eucaryotes.

Some procaryotes are anaerobic. Because they cannot tolerate oxygen, anaerobic procaryotes derive their energy *by fermentation* of carbohydrates. Fermentation, as mentioned in Chapter 15, is a process by which carbohydrate molecules combine to form an alcohol plus carbon dioxide and water and release energy in the process. All organisms that are not anaerobic (including some procaryotes) are aerobic and obtain energy by *respiration,* which means they use oxygen to oxidize carbohydrate to carbon dioxide, water, and energy. Respiration is more efficient than fermentation because it releases all of the available energy; fermentation, because its end product, alcohol, is still oxidizable, does not use all the available energy.

The Antiquity of the Biosphere

The most ancient fossils that have been found to date are about 3.5 billion years old. Some of the ancient fossils are the remains of microscopic procaryotes (Fig. 16.7); others are structures made up, layer upon layer, of thin sheets of calcium carbonate that were precipitated as a result of blue-green bacteria (also procaryotes) influencing the chemistry of seawater (Fig. 16.8A). The layered structures, which are called *stromatolites,* are not fossils of the actual organisms, but they provide clear evidence of the presence of them because we can see and study similar structures being formed today (Fig. 16.8B). For at least 2 billion years the only life on the Earth was procaryotic. Several different kinds of procaryotes evolved over their 2-billion-year supremacy; then, about 1.4 billion years ago, a profound change occurred, and eucaryotes appeared. How and where the first simple eucaryotes came into being is a matter for speculation.

Just as it is only possible to hypothesize about how the first procaryotes came into being, so, too, are we limited to hypothesizing about the origin of the eucaryotes. We can, however, be reasonably sure that eucaryotes arose from procaryotes. The chemical pathways in the two classes of cells are so similar that it is clear they must be related. Furthermore, the organelles in the eucaryotes so closely resemble some



A. B.

Figure 16.6 A. Procaryotic cell. A bacterial cell devoid of visible organelles and with the DNA concentrated in a poorly defined nucleoid that is not separated from the cytoplasm by a membrane. B. Eucaryotic cell from a plant root with a well-defined, membrane-bound nucleus and varied cytoplasmic organelles. Note that the cells are colored because they have been stained.

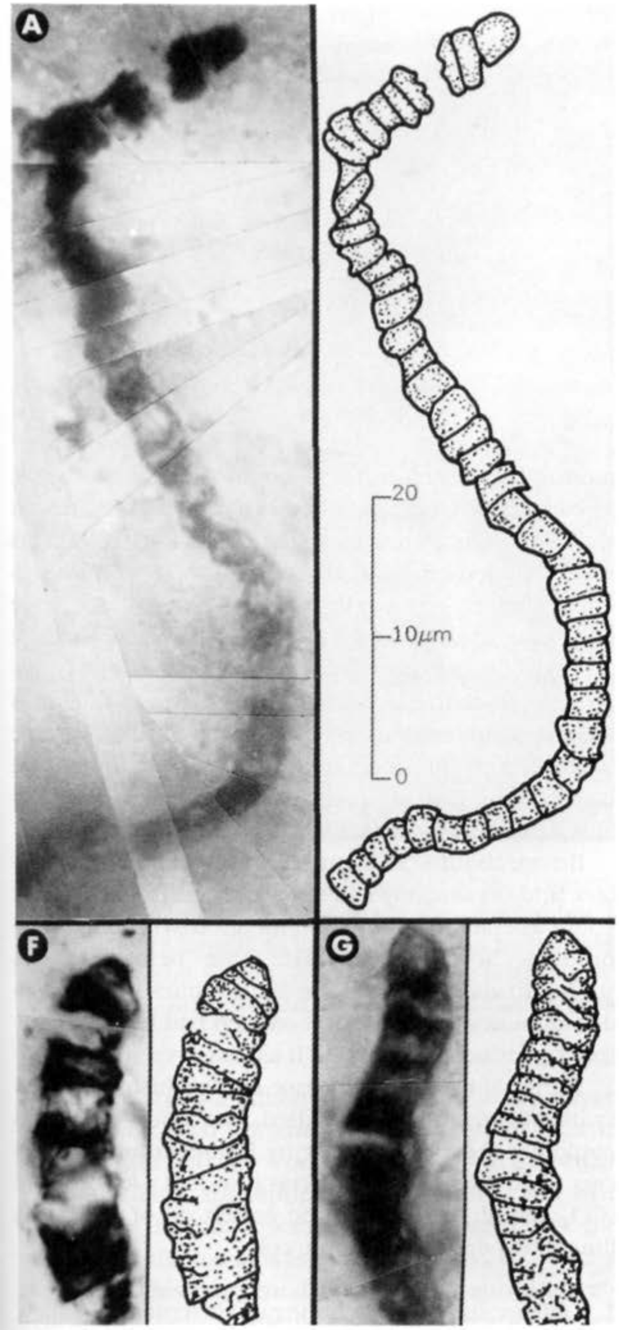
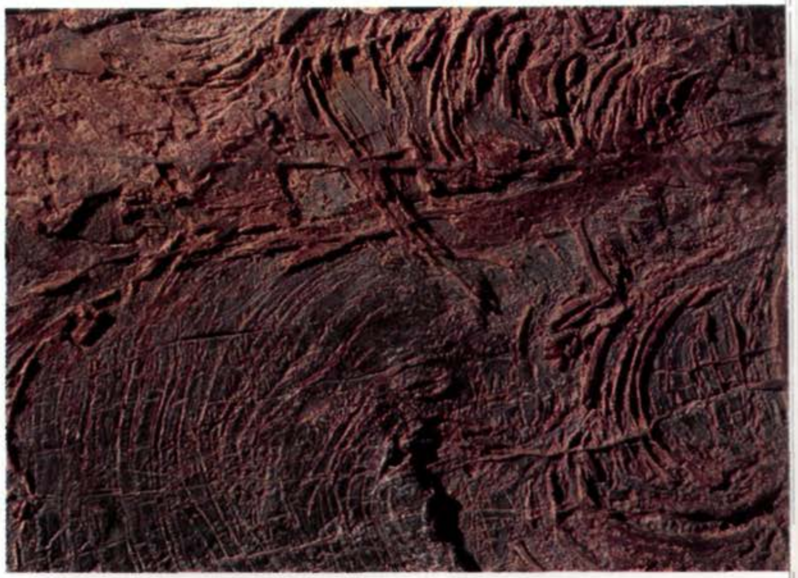
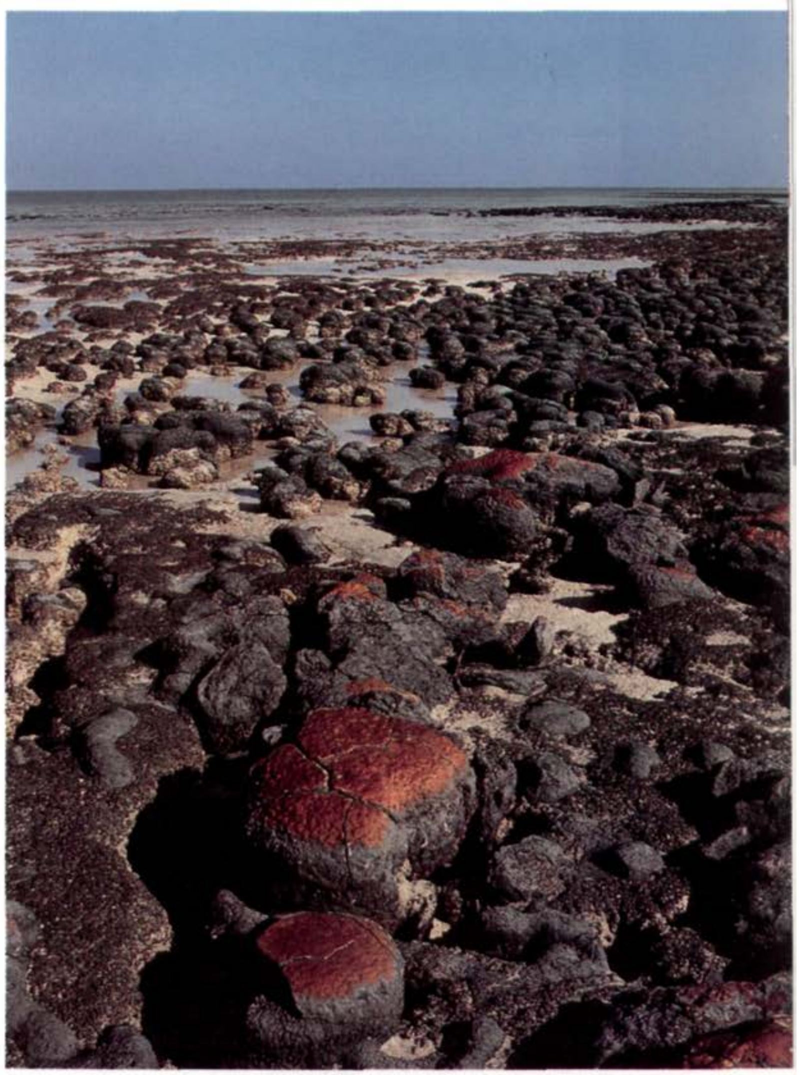


Figure 16.7 Examples of the most ancient fossil procaryotes ever found. 3.5 billion-year old microfossils in chert from Western Australia. Adjacent to each photograph is a sketch. Magnification is indicated by the scale; 10μm is equal to 0.1 mm



A.



B.

Figure 16.8 Evidence of the antiquity of life. Stromatolites are layered growths that form in warm, shallow seas when photosynthetic bacteria cause dissolved salts to precipitate. A. Fossil stromatolites greater than 1.5 billion years old from the northern Flinders Range, South Australia. B. Modern stromatolites forming in the intertidal zone, Shark's Bay, Western Australia.

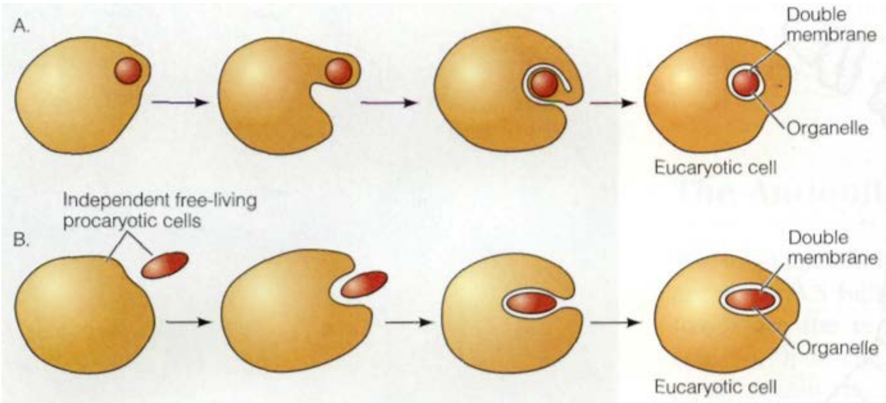


Figure 16.9 Hypotheses for the origin of eucaryotic cells. A. The invagination hypothesis in which a procaryotic cell encloses a portion of itself, which then becomes an organelle. B. The symbiosis hypothesis in which an independent, free-living procaryotic cell invades and lives inside another procaryotic cell. The invader evolves into an organelle.

of the smaller procaryotes that most authorities believe that organelles were once procaryotic bacteria and that eucaryotes somehow arose by larger procaryotes enclosing, and using the chemical processing of, smaller cells. There are two hypotheses about how this might have happened: by *membrane invagination* or by *symbiosis* (Fig. 16.9).

The first hypothesis, membrane invagination, proposes that the ancestral procaryotic cell folded in on itself, forming pockets in which particular enzymes could be concentrated. When the folds were pinched off they were enclosed by and protected in some of the outer membrane of the procaryotic cell and thus became simple organelles. Over a long period of time, such pinched-off bits of cells became the increasingly complex organelles we see today.

The second, and more popular, hypothesis is given the name *symbiosis* because it refers to the close association between different organisms. The hypothesis is that the nucleus and organelles of eucaryotic cells were originally small procaryotic cells that simply invaded larger procaryotic cells and took up residence there.

Why eucaryotic cells arose at all, like so many other questions involving the early biosphere, is not known with certainty, but their origin may involve the rise of oxygen in the atmosphere.

The Rise of Oxygen

There is plentiful evidence to show that before about 1.5 billion years ago the Earth's atmosphere was deficient in oxygen. Water-worn grains of pyrite (FeS_2), for example, turn up in ancient sedimentary rocks. Today, when grains of pyrite get into streams, they are oxidized long before they accumulate in sediments. The most convincing evidence of an oxygen-deficient atmosphere is found in ancient chemical sediments called banded-iron formations (Fig. 16.10). These sediments were laid down in the sea, a sea that must have been able to carry dissolved iron (something it can't do now because oxygen precipitates the iron). It is hypothesized that the ancient banded-iron formations were precipitated in places where photosynthetic bacteria were producing oxygen locally and that the banded-iron formations were thereby relieving the atmosphere of a potentially lethal burden of oxygen.

Before about 1.5 billion years ago, all life was anaerobic and presumably had not reached a point where it could tolerate, let alone use, its photosynthetic waste product. Scientists calculate that before oxygen started to accumulate in the air, 25 times the present-day amount of atmospheric oxygen had been neutralized by reducing agents such as dissolved iron.

During the anaerobic phase of life's history, photosynthesis was well established, but living organisms could get their energy only by fermentation. The energy yield from fermentation is low, and a lot of waste (CO_2 and alcohol) has to be gotten rid of. This puts limitations on the anaerobic cell:

1. A large surface-to-volume ratio is required to allow rapid diffusion of food in and waste out. Anaerobic cells must therefore be small.
2. Anaerobic cells have enough trouble keeping themselves supplied with energy. They cannot afford to deploy energy resources on the maintenance of specialized organelles, including the nucleus. This means that all anaerobic bacteria were procaryotes.
3. Procaryotes need free space around them; crowding interferes with the movement of nutrients and water into and out of the cell. Therefore, they live singly, or are strung end-to-end in chains. They cannot form three-dimensional structures.

Figure 16.10 Banded-iron formation of the Hamersley Range, Western Australia, formed during the Lower Proterozoic Eon. Banded-iron formations are chemical sediments and are thought to have formed when iron in solution in seawater was precipitated as a result of photosynthetic bacteria releasing oxygen. The woman in the foreground is Dr. Janet Watson, a distinguished English geologist.

It took about 2 billion years for the Earth's oxygen sinks to be used up, and during this entire time the procaryotes had the world to themselves. Eventually, an oxygenated atmosphere started to form, and when it did, the biosphere seems to have wasted little time in turning the lethal waste product, oxygen, to its advantage. This happened through the appearance of eucaryotic cells; some of the characteristics of aerobic eucaryotic cells are as follows.

1. They use oxygen for respiration, and because oxidative respiration is much more efficient than fermentation, they do not require as large a surface-to-volume ratio as anaerobic cells do. Such cells are larger.
2. Aerobic cells, because of their superior metabolic efficiency, can maintain a nucleus and organelles.
3. Aerobic eucaryotes are not bothered by crowding so, unlike procaryotes, eucaryotes can form three-dimensional colonies of cells.

With the appearance of eucaryotes and the growth of an oxygenated atmosphere, the biosphere started to grow larger, to change rapidly, and to play an increasingly important role in the Earth system. For an example of some of the complex systems that developed as a consequence of the rise of oxygen, see "A Closer Look: The Biochemical Cycle of Nitrogen."

EVOLUTION AND THE FOSSIL RECORD

Eucaryotes appeared in the middle of the Proterozoic Eon about 1.4 billion years ago. The exact date is not known with certainty, but the fossil evidence clearly shows that by 1 billion years ago the eucaryotes were well established. The first eucaryotes were single-celled green algae. The earliest fossils of larger multicellular organisms appear in rocks about 600 million

A Closer Look

The Biochemical Cycle of Nitrogen

Amino acids are essential constituents of all living organisms. They are given the name *amino* because they contain amine groups (NH_2) and the key element in amines is nitrogen. As a consequence, nitrogen is essential for all forms of life.

Nitrogen in nature exists in three forms. Nitrogen in the atmosphere is present in the elemental form (N_2), but reduced forms such as ammonia (NH_3) and oxidized forms such as nitrate (NO3) also occur in nature. It is only in the reduced form that nitrogen can participate in biochemical reactions.

The key to the nitrogen cycle is understanding how reduction (also called *fixation)* and oxidation (also called *denitrification)* take place and how nitrogen moves between the four major reservoirs—the atmosphere, biosphere, ocean, and soil and sediment. As you read the following discussion, please study Figure C16.1 carefully. The figure shows the reservoirs, the estimated number of grams in each reservoir, and the paths by which nitrogen moves between the reservoirs.

The nitrogen cycle is dominated by the atmosphere and by the fact that N_2 cannot be directly used by organisms. Nitrogen is removed from the atmosphere and/or made accessible to organisms in three ways:

1. By solution of N_2 in the ocean.
2. By oxidation of N_2 to NO_3 by lightning discharges, in which form it is rained out of the atmosphere and into the soil and the sea. Plants can reduce NO_3 to NH_3, thus making nitrogen assimilable to the biosphere.
3. By reduction of N_2 to NH_3 through the actions of nitrogen-fixing bacteria in the soil or the sea. The reduced nitrogen is quickly assimilated by the biosphere.

The nitrogen cycle is interesting because of its complexity but interesting too because parts of the cycle have had to evolve as the atmosphere became oxygenated. Because organisms cannot use N_2 directly, there must have been either some reduced nitrogen available when life arose or the earliest organisms had the ability to reduce N_2. Anaerobic nitrogen-fixing bacteria are certainly very ancient, and the fixation chemistry that evolved with them will not work in the presence of oxygen. Such bacteria must have evolved before the atmosphere contained oxygen. Today these bacteria live only in oxygen-free environments. A few nitrogen-fixing bacteria have developed an oxygen tolerance, even though they still use the old, anaerobic fixation chemistry. They perform this trick by making sure that the sites in their cells where fixation occurs are carefully guarded from oxygen.

As the oxygen content of the atmosphere increased, the amount of nitrate rained into the soil also increased. This opened new niches that were soon occupied by organisms that learned to reduce NO_3 to NH_3. Many of the higher plants have this ability to use nitrate. Those that

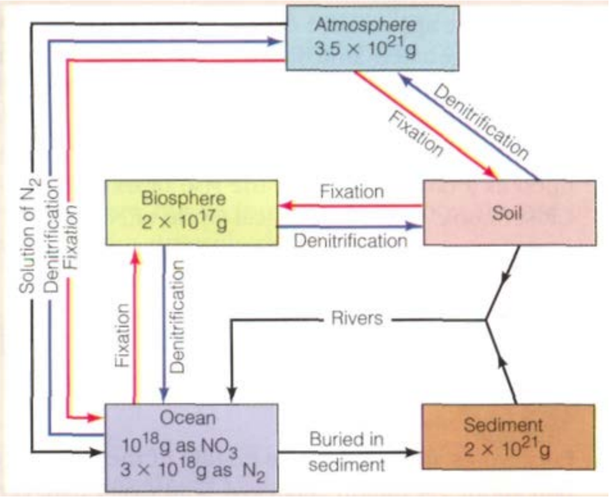


Figure C16.1 The nitrogen cycle. For a discussion, see the text.

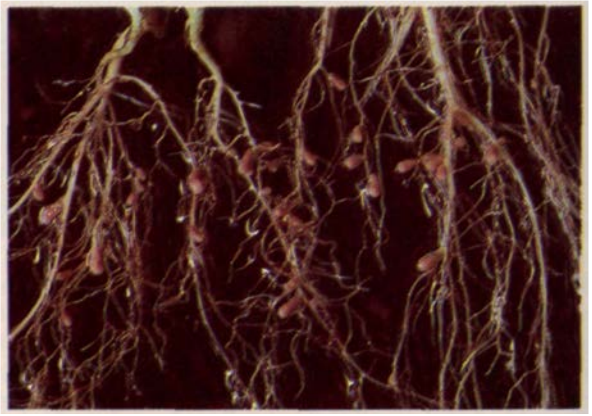


Figure C16.2 Root nodules on white clover produced by colonies of nitrogen-fixing bacteria.

cannot reduce nitrate act as hosts to symbiotic nitrogen-fixing bacteria, to which they supply energy in exchange for fixed nitrogen (Fig. C16.2).

Once reduced, nitrogen tends to stay reduced, remain in the biosphere, and be either reused by other organisms or oxidized back to N_2 and returned to the atmosphere. The main route by which nitrogen returns to the atmosphere, however, is the reduction of nitrate. This route is kept open by bacteria that use the oxygen in nitrate in order to oxidize carbon compounds during metabolism. Denitrifying bacteria must therefore have evolved quite late in the history of the biosphere, after oxygen started to accumulate in the atmosphere. Thus, the simple nitrogen cycle of the early Earth has evolved into today's complex cycle in response to a changing atmosphere.

years old, and thereafter the appearance and spread of multicellular organisms happened very rapidly (Fig. 16.11).

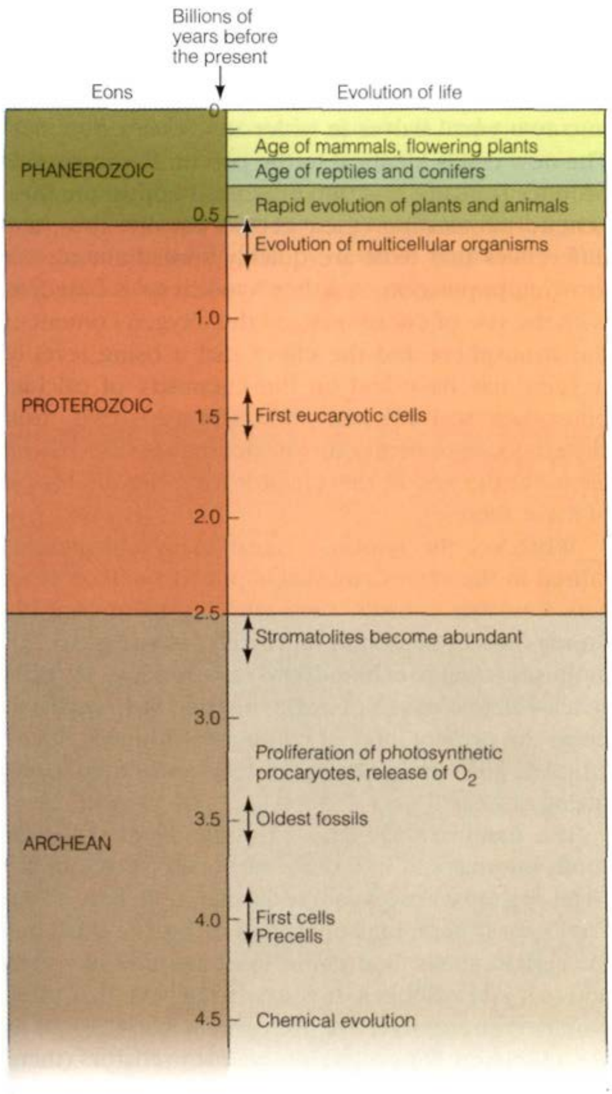


Figure 16.11 The evolution of life on the Earth from 4.6 billion years ago to the present. The rates at which new organisms appear and of biological diversity both increase with time.

Ediacaran Fauna

The earliest fossils of multicellular organisms were first discovered in 600-million-year-old rocks in the Ediacara Hills of South Australia and are known as **Ediacaran animals.** Nearly identical fossils have subsequently been discovered in similar-aged rocks in other parts of the world.

The Ediacaran fauna were animals that lived in quiet marine bays and lacked any hard parts. They seem to have been jelly-like animals without any armor or defense. They probably did not need any defense mechanism because predatory animals had not yet arisen.

Even if the Ediacaran fauna were the first, or at least among the first, animals to evolve, they nevertheless represent a huge jump in complexity from the first unicelled eucaryotes 800 million years earlier. Scientists have not yet been able to discover much about what went on during those 800 million years.

The Ediacaran animals are of three main kinds: (1) disc-like, resembling today's jellyfish (Fig. 16.12A); (2) pen-like, resembling today's sea-pens or soft corals, and (3) worm-like, resembling broad flat worms (Fig. 16.12B). There are some odd features about the animals. For example, the disc-like fossils are not really jellyfish because they lack the central radial structure and peripheral concentric structure of true jellyfish. Furthermore, although it may be a bit too much to read from such an ancient fossil, the "worms" don't seem to have guts. The odd feature shared by most of the Ediacaran animals is that they were flat. All organisms need to exchange material with the environment, and one way to do this efficiently is to have a large surface-to-volume ratio. One way to have a large ratio is to be flat. It is thought that, by having a large surface for feeding, respiration, and excretion, the Ediacaran animals could grow very large. (*Dickinsonia,* a disc animal, was up to a meter long.)

Being flat can have problems: it is difficult to survive in waves and high-energy environments, for example. Therefore animals quickly evolved away from flatness and developed convoluted and branching ex-

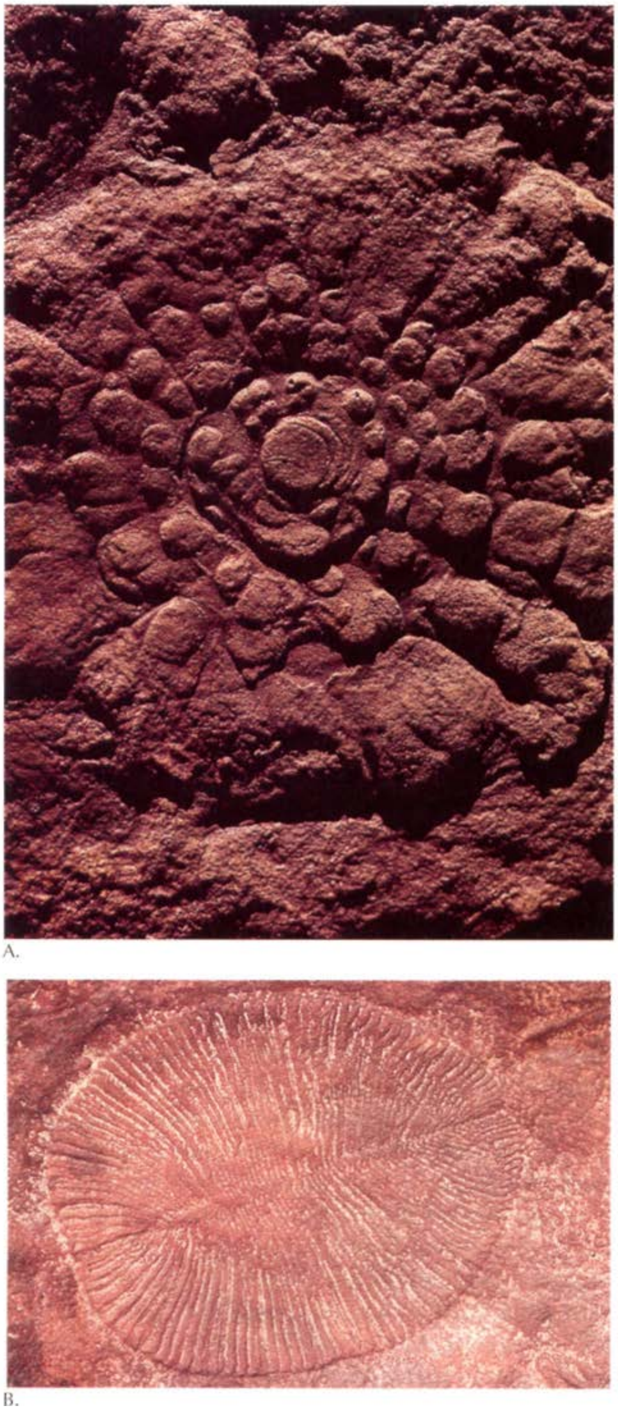


Figure 16.12 Two members of the Ediacara fauna from South Australia. These are the most ancient multi-celled animals that have ever been found. A. *Mawsonia spriggi,* a discoid shape, possibly a floating animal like a jellyfish. B. *Dickinsonia costata,* a curious worm-like creature.

change surfaces. Brains, lungs, gills, guts, and the vascular system are all variations on these which, as we see next, started to develop in the Cambrian Period.

The Cambrian Expansion

The Ediacaran animals arose at the very end of the Proterozoic Eon. The Phanerozoic Eon, which starts with the Cambrian Period, was a time of incredible biological diversity. Figure 15.3 is a generalized diagram of biological diversity through geologic time. It is clear that the Proterozoic biosphere (regardless of what size it was) was not very diverse. This means that most of the ecological niches later to be occupied by multicelled organisms were vacant in the Proterozoic.

Why should biological diversity have exploded in the Cambrian Period? That is another of the great unanswered questions about the biosphere. Many hypotheses have been offered, but none, as yet, is backed by hard evidence. One hypothesis is that sex, which developed with the eucaryotes, caused the explosion. Procaryotes are asexual and reproduce by a process called *mitosis* in which the cell splits its DNA into two equal halves in order to create a new cell. The new cell is identical to the parent. Eucaryotes reproduce sexually, and two organisms contribute their genetic information (their genes) equally; thus, any differences that exist are quickly spread among the growing population. Another hypothesis is based, as with the rise of eucaryotes, on the oxygen content of the atmosphere and the effect that a rising level of oxygen may have had on the chemistry of calcium phosphate and calcium carbonate (the two main skeleton components) in the ocean. See the "Guest Essay" at the end of this chapter for more discussion of these theories.

Whatever the reason, a great many changes occurred in the early Cambrian about 570 million years ago. Compact animals were evolving to replace the floppy ones of Ediacaran times: trilobites (Fig. 16.13), molluscs (clams), echinoderms (sea-urchins), and sea-snails—all (except trilobites) types that have persisted up to the present, and all equipped with gills, filters, efficient guts, a circulatory system, and other space-saving devices.

The Cambrian saw the introduction of skeletons (both internal and external): carbonate skeletons for small worms whose fossils today mark the base of the Cambrian system; phosphate and carbonate skeletons for certain shells; and chitin (a compound like nails and hair) for trilobites. It is easy to believe that there would be strong selective pressure on many groups to get a skeleton: for protection against predators (there were some by now); against drying out; against injury in turbulent water; and so on. Why did skeletons evolve in the Cambrian, however? Perhaps it was simply because no one needed to wear them before that time. Perhaps it was because the rising level of oxy-

Figure 16.13 Fossil trilobite from the Cambrian Period. Trilobites were one of the first animals to develop a hard, chitin covering, presumably as a defense against predators. This sample was collected in Utah.

gen in the atmosphere and the extra metabolic energy were now available to make skeletons and encourage diversity.

The idea of an oxygen control is plausible (but far from proven), for carbonate, phosphate, and chitin skeletons are built only by aerobic organisms. The main difficulty is that there is no way of telling what the atmospheric oxygen content was at the beginning of the Cambrian: it may already have been far above the 10 percent that many hypotheses suggest was needed to develop skeletons. Lest it be thought that skeletons were the rule among Cambrian animals, it is instructive to look at a famous fossil assemblage in the *Burgess Shale* (Middle Cambrian of British Columbia, Fig. 16.14). Note that when all of the soft-bodied forms have been removed, few skeletonized ones remain. The Burgess Shale assemblage is an object lesson in the interpretation of the rest of the fossil record.

The Burgess Shale assemblage shows something else, too: a richness of forms that we no longer see in the contemporary bottom-dwelling marine fauna, rich though it is. The Cambrian was a time of almost unbridled growth and diversity in the marine environment: thousands of niches waited to be filled, and there was no lack of candidates for their occupancy. Most had to fall by the wayside, yielding to the nine surviving phyla that we know today. When we look at this amazing snapshot of Cambrian marine life, with all its extraordinary forms, we cannot resist the feeling that running the ecological race is as good as winning it.

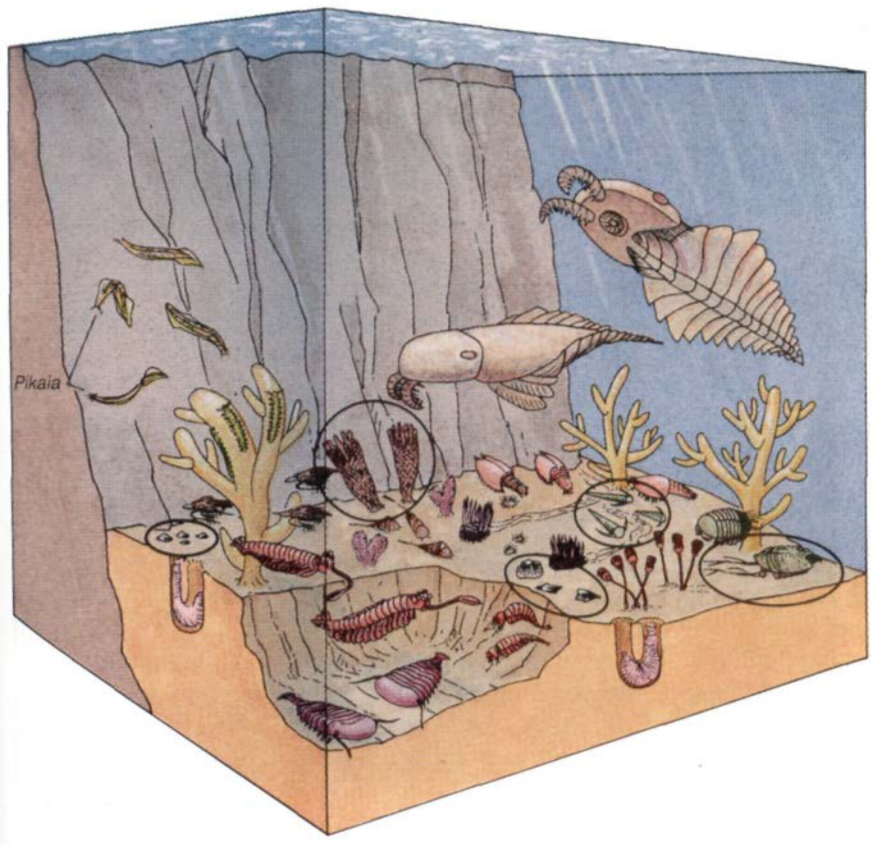
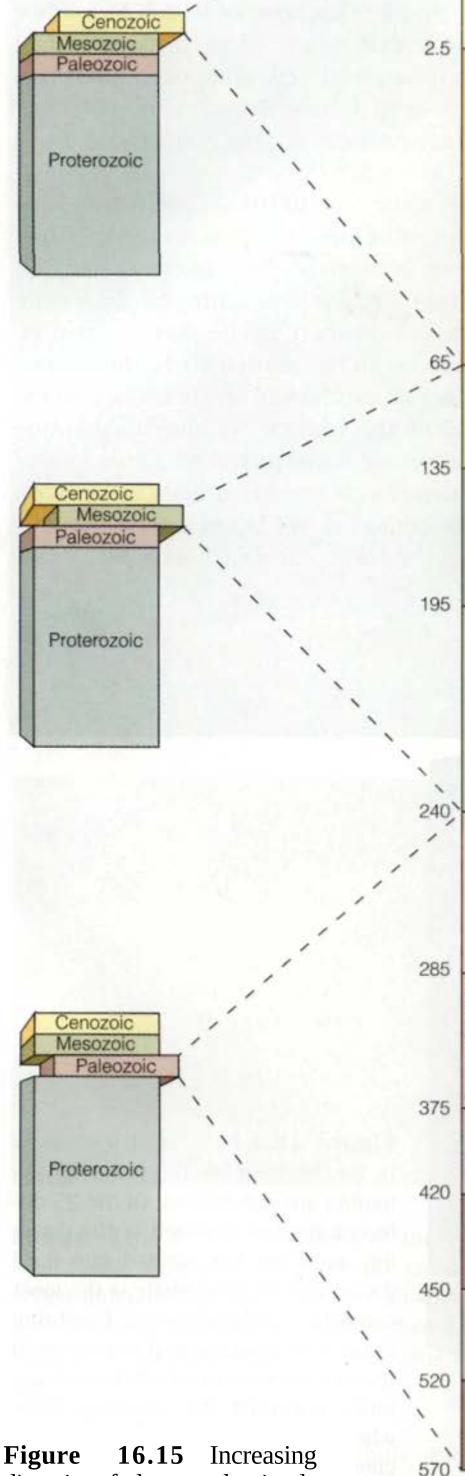


Figure 16.14 Most of the fossils in the Burgess Shale of British Columbia are soft-bodied. Of the 25 different species depicted in the drawing, only the five circled had hard parts. The Burgess Shale is the most complete assemblage of Cambrian fauna ever found, and it is presumed that the abundance of soft-bodied animals reflected the situation elsewhere in the ocean. The most ancient chordate, *Pikaia*, is seen as a small fish. (See Fig. 16.20 for a photo of a *Pikaia* fossil.)



Millions of years ago	PERIOD	PLANT EVOLUTION	ANIMAL EVOLUTION
2.5	**Quaternary**	Repeated glaciation	Appearance of *Homo sapiens* First *Homo*
65	**Tertiary**	Decline of forests, spread of grasslands Explosive radiation of flowering plants	Appearance of hominids Appearance of first apes All modern genera of mammals present In seas, bony fish abound Rise of mammals First placental mammals
135	**Cretaceous**	First flowering plants	Dinosaurs extinct Modern birds
195	**Jurassic**	Forests of gymnosperms and ferns over most of the earth	First birds Age of dinosaurs
240	**Triassic**	Gymnosperms dominant	Explosive radiation of dinosaurs First dinosaurs First mammals Complex arthropods dominant in seas First beetles
285	**Permian**	Widespread extinction Decline of nonseed plants	Widespread extinction Appearance of mammal-like reptiles Increase of reptiles and insects Decline of amphibians
375	**Carboniferous**: Pennsylvanian, Mississippian	Gymnosperms appear Widespread forests of giant club moss trees, horsetails and tree fern — create vast coal deposits	Early reptiles First winged insects Increase of amphibians
420	**Devonian**	First seed plants Development of vascular plants: club mosses and ferns	Amphibians diversify into many forms First land vertebrates — amphibians
450	**Silurian**	First vascular plants First land plants	Golden Age of fishes First land invertebrates — land scorpions
520	**Ordovician**		First vertebrates — fishes Increase of marine invertebrates
570	**Cambrian**	Algae dominant	Trilobites dominant Explosive evolution of marine life

Figure 16.15 Increasing diversity of plants and animals with time.

LIFE ON LAND

The Paleozoic Era lasted about 330 million years and is divided into six periods starting with the Cambrian, followed in succession by the Ordovician, Silurian, Devonian, Carboniferous, and Permian. The five periods of the Mesozoic and Cenozoic Eras follow. Each period is characterized by the appearance of major groups of plants and animals, as summarized in Figure 16.15.

The great proliferation of life in the Cambrian was entirely confined to the sea. Successful phyla diversified and filled niches; unsuccessful phyla have disappeared into the abyss of time. By 500 million years ago, the main plans for animal life had been settled. The one big step that remained was to leave the sea and occupy the land. Eventually, plants, insects, and animals all took the step.

Here are some of the requirements for life on land (they are the same for all organisms.)

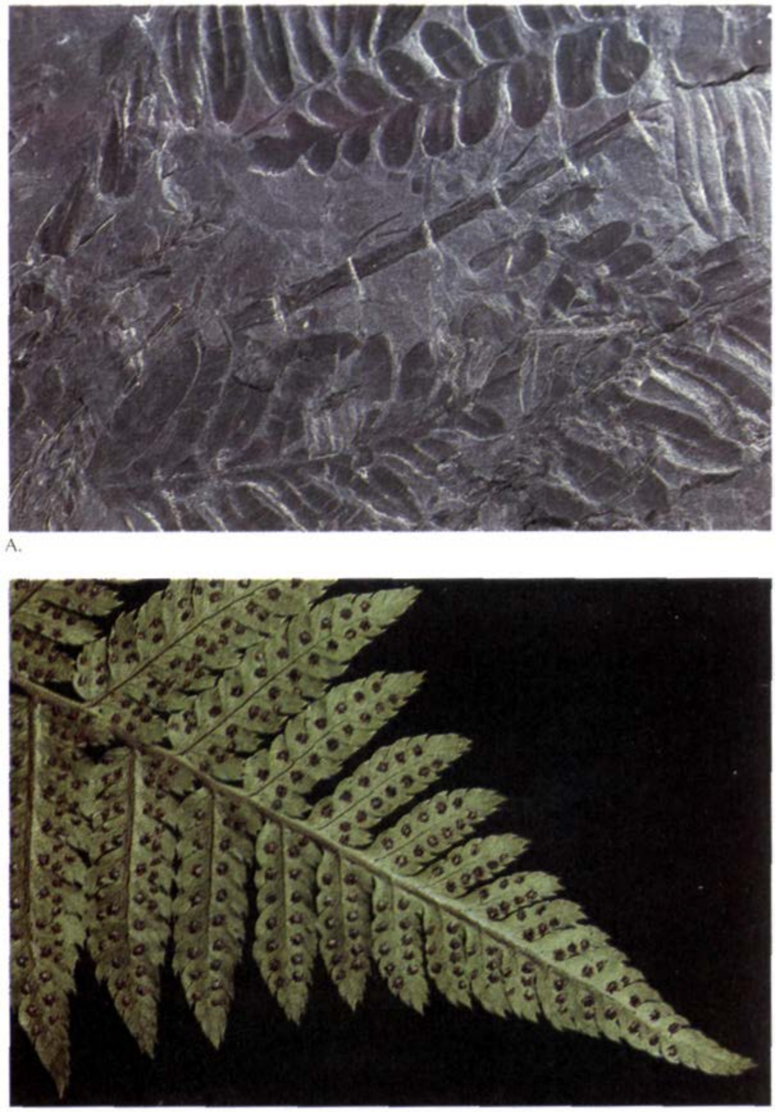
1. *Structural support,* needed because, while aquatic organisms are buoyed up by the water, on land gravity becomes a real force with which to contend.
2. *An internal aquatic environment,* with a plumbing system giving it access to all parts of the organism, and devices for conserving the water against losses to the surrounding atmosphere.
3. *Means for exchanging gases with air* instead of with water.
4. *A moist environment for the reproductive system,* essential for all sexually reproducing organisms.

Selection for these necessities is largely what has shaped terrestrial organisms into the familiar forms we know today.

Plants

It is assumed (so far without proof) that land plants evolved from the earliest known eucaryotes, green algae. What evolved were vascular plants that had structural support from stems and limbs (requirement 1) and a *vascular system* (requirement 2), which is a system of channelways, by which sap is transferred from the roots to the leaves. Requirement 3 (gas exchange) is by diffusion and is controlled by adjustable openings in the leaves called *stomata* (singular *stoma*, Greek for mouth). When carbon dioxide pressure inside the leaf is high, the stomata open; when low, they close. (The stomata also close when the plant is short of water, thereby protecting it against desiccation.) Gas exchange was managed by Devonian plants much as it is in today's plants.

The earliest plants, of which mosses and ferns are modern examples, were *seedless plants* (Fig. 16.16). Many of the seedless plants can tolerate some



A.

B.

Figure 16.16 Ferns and club mosses are modern representatives of the seedless plants that first established themselves on the land in the Silurian. A. Fossil fern about 350 million years old. B. *Thelypteris phegopteris,* a modern fern that is also known as the long beech fern, showing spores on the undersides of the frond.

A.

B.

Figure 16.17 Naked-seed plants developed from the seedless late in the Devonian Period. A. A leaf of *Glossopteris*, a family of seed-fern plants that spread through Southern Gondwana. B. Leaves of modern fossil gingkos. The fossil is from North Dakota. Gingkos are long-lived relics of the ancient family of naked-seed plants.

drought; mosses and ferns survive dry spells, releasing spores that lie dormant until moisture returns. But all of these plants rely on moisture for the sexual phase of the reproductive cycles; without it the sex cells have no medium in which to reach each other and fuse; therefore, fertilization does not occur. Consequently, the seedless plants have never been able to colonize habitats where moisture is not unfailingly present for at least part of the growing season.

The seedless plants reached their peak in the Carboniferous Period, when they dominated the vast forests that were then growing on the tropical floodplains and deltas of North America, Europe, and Asia, producing a fossil fuel (coal) that started the Industrial Revolution two centuries ago.

By the Middle Devonian, a few plants were already on the way to meeting requirement 4—that is, providing their own moist environment in order to facilitate sexual reproduction. The plants that evolved were the **gymnosperms** or naked-seed plants such as *Glossopteris* of Gondwana fame (Fig. 16.17A). The female cell is attached to the vascular system and thus has a supply of moisture. The male cell is carried in a pollen grain coated with a waxy coating. When the two fuse, a seed results. The seed is simply a supply of moisture and nutrients to sustain the early growth of the young plant until it becomes self-supporting by photosynthesis. Although the change was not a radical one, it freed the vascular plants once and for all from the swampy lowlands that were their first habitat. Naked-seed plants survive today; examples are the gingkos (Fig.l6.17B) and the conifers.

Gymnosperms had a huge success. Freed from the swampy habitat, they did not have to compete with the great seedless trees of the coal forests, but instead struck out on their own in virgin territory, the drier uplands of the newly forming supercontinent, Pangaea. By the end of the Carboniferous Period, they had spread over most of the world, and by the Triassic they were rivaling in size their former cousins of the swamps.

Life has one drawback for gymnosperms. The male cell-carrier, the pollen, is spread through the air. What chance has a pollen grain loose in the air of finding a female cell? To ensure success, gymnosperms have to make huge amounts of pollen.

Flowering plants (**angiosperms,** or enclosed-seed plants) solved the problem of the random distribution of pollen. For a small incentive (nectar or a share of the pollen), insects will deliver the pollen (Fig. 16.18). It took longer for the angiosperms to evolve than it did for the gymnosperms, but by the end of the Cretaceous Period angiosperms had become the dominant land plants. Their life cycle is not significantly different from that of gymnosperms, but they have specialized in symbiosis with animals: insects for pollination, birds and quadrupeds for seed dispersal.

The last frontier for plants—so far, at least—has been the dry steppes, savannas, and prairies. These were not colonized until the Tertiary Period, when

Figure 16.18 Pollen from a hollyhock (an angiosperm) coats a bumble bee collecting nectar. As the bumble bee moves from plant to plant, the pollen is efficiently distributed.

grasses evolved. In arriving at them, we have also reached the culmination of animal life on land, the great grazing faunas of the high plains of all continents except Antarctica. It is now time to go back to the Paleozoic Era and consider the first, tentative steps of the insects and quadrupeds.

Insects

Among the many creatures in the Cambrian seas, there were many that belong to the phylum *Arthropoda,* so-called from the presence of jointed legs. They include crabs, spiders, centipedes, and insects and are the most diverse phylum on the Earth. They, not the fish from which we humans are descended, "were the first creatures to make the change from sea to land.

Arthropods, with a few exceptions, were quite small, had lightweight structures, and were covered with a shell of chitin to provide structure support. They were thus admirably preadapted for life on land in regard to structural support and water conservation. The earliest to go on land were probably Silurian centipedes and millipedes; by the Carboniferous, insects were abundant and included dragonflies with a wing span of up to 60 cm (24 in). For all their success as land creatures, the arthropods have always had very primitive respiratory and vascular systems. They breathe through tiny tubes that penetrate the outer coating. Because the respiring mass of an aerobic organism increases as the cube of its length, while the area available for gas exchange increases only as the square, it is clear that this mode of respiration must severely limit the size of an organism. This is why most insects are small.

The arthropods have an open vascular system. That is, their "blood" does not circulate in closed vessels, but is simply body fluid bathing the internal organs and generally kept in sluggish motion by a "heart" that is little more than a contractile tube. At first it seems odd that the arthropods, with such a primitive arrangement, should have diversified into more than a million terrestrial species. The arthropods' vascular system isn't great, but it obviously works. And it's close to indestructible; whoever heard of a cockroach having a heart attack?

Animals

Inconspicuous among the fossils of the Burgess Shale is a small fossil called *Pikaia* (Fig. 16.19). Pikaia is a *chordate,* a member of the phylum to which we humans belong, by virtue of possessing a *notochord,* a cartilaginous rod running along the back of the body. (We and other vertebrates have one as embryos, later replaced by the backbone). *Pikaia* (which may well be one of our ancestors) and other Cambrian fish were jawless, probably feeding on organic matter dredged from the seafloor. Jawed fish come next.

With jaws came a great burst of diversification; as can be imagined, their possession must have given access to a whole array of ecological niches that until then had been vacant. The original jawless fish, a few as long as many centimeters, were quickly joined by larger armored fish, including 9 m (30 ft) carnivores, as well as sharks and other boneless (cartilaginous) fish, and the huge Order of ray-finned fish that are familiar as today's game and food fish. (This rapid diversification of the fish was mentioned briefly in Chapter 15.)

The first fish to venture on land, an obscure group called the *crossopterygians,* did so in the Devonian. They gave rise to the amphibians. The crossopterygians had several features that served to make the transition possible. Their lobe-like fins, for example, were preadapted as limbs because the lobes contain (much foreshortened) the elements of a quadruped limb, complete with small bones to form the extremity. They also had internal nostrils characteristic of air-breathing animals. Being fish, the crossopterygians already had a serviceable vascular system that was adequate to make a start on land. Water conservation, however, never became a strong point with amphibians: they retain permeable skins to this day, which is one reason why they have never become independent of the aquatic environment.

Despite their limitations, the amphibians ruled the land for many millions of years during the Devonian Period. They had one difficulty that limited their expansion into many niches: they never met the reproductive requirement for life on land. In most species, the female amphibian lays her eggs in the water, the male fertilizes them there after a courtship ritual, and the young hatch as fish (tadpoles). Like the seedless plants, the amphibians, with one foot on the land, so to speak, have remained tied to the water for breeding. Although some became quite large (2 to 3 meters, or 2 to 3 yds long), they never diversified much. One branch went on to become reptiles; those of the rest that survive are frogs, toads, newts, salamanders, and limbless water "snakes" that seem to have decided that, after all, they prefer a fish's life.

The reptiles freed themselves from the water by evolving an egg that could be incubated outside of it and by getting themselves a watertight skin. These two "inventions" gave them the versatility to occupy terrestrial niches that the amphibians had missed be-

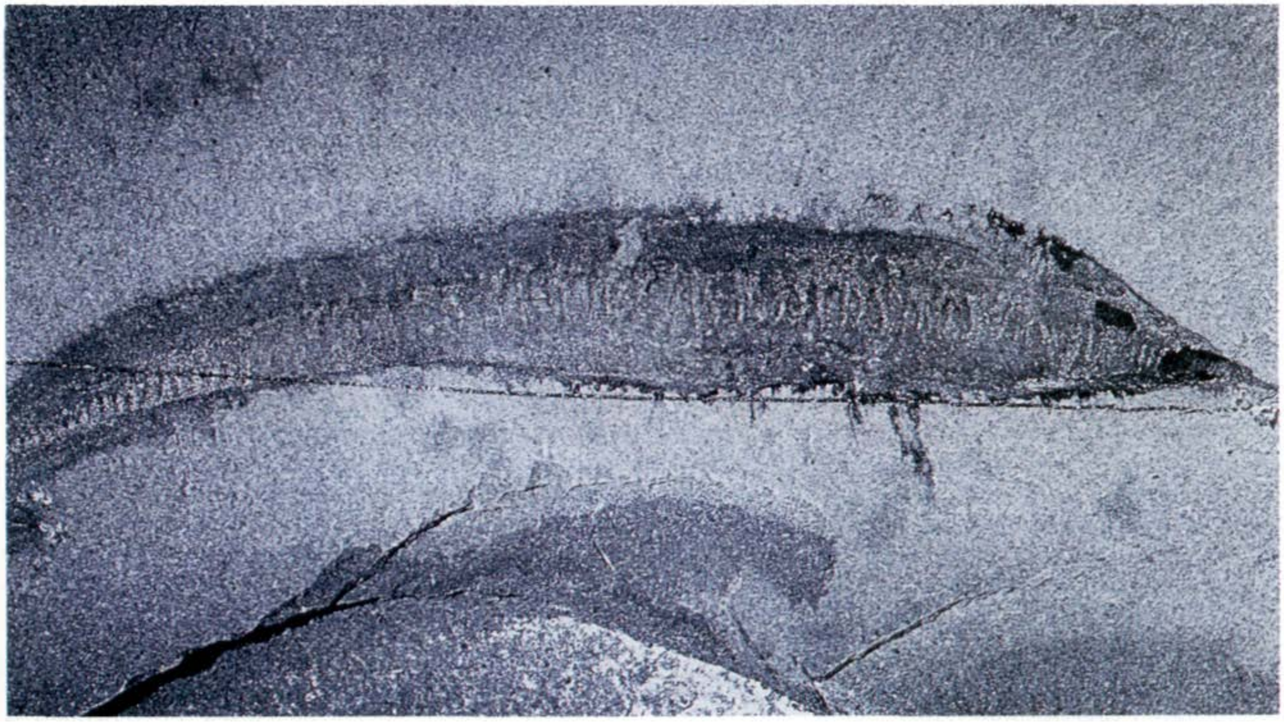


Figure 16.19 *Pikaia*, a soft-bodied animal from the Burgess Shale in British Columbia, is the earliest known chordate. *Pikaia* is the most ancient member of the group that became the vertebrates and to which we humans belong. See Figure 16.15 for the curious creatures with which *Pikaia* coexisted.

cause of their bondage to the water. The amniotic egg did for reptilian diversity what jaws did for diversity in fishes. Originating in the Carboniferous coal swamps, by the Jurassic Period the reptiles had moved over the land, up in the air and back to the water (as veritable sea monsters), in addition to having produced the two Orders of dinosaurs (the largest quadrupeds ever to walk the Earth) and given rise to two new vertebrate Classes, mammals and birds.

As pointed out in Chapter 15, the mammals are in many ways better equipped as quadrupeds to occupy terrestrial niches than were the great reptiles. It is difficult to pick out a single mammalian "invention" comparable to the jaws of fish or the reptilian egg, for the mammal is a fine-tuned quadruped, adapted to a faster and more versatile life than the reptiles could ever have led (notwithstanding the current vogue of popularity they are enjoying). The mammalian "invention" is perhaps just that: a set of interdependent improvements managed by a more capable brain and supported by a faster metabolism. The placental uterus is sometimes regarded as the key to mammalian success; but it's really only a piece of equipment mandated by the delicate intricacy of the fetus that lives in it, especially the brain. By comparing brain-to-body weight ratios in archaic and modern reptiles and mammals, it can be shown that increase in mammalian brain size is a continuing process, whereas in reptiles increase has not occurred: the ratios in modern reptiles do not differ significantly from those in archaic ones. What is perhaps significant is that had something not come along and wiped out the ruling reptiles, we should not even be here to think about them. Fortunately, something did disrupt the reptiles, and that leads to the last and in some ways most intriguing question about the biosphere: Why and how did great groups of animals suddenly die out and new ones arise to take their places?

EXTINCTIONS AND THE BIOSPHERE

Few of the species alive at the beginning of the Cambrian Period had living descendants at the end. Throughout the Cambrian, species died out and became extinct; new species evolved to occupy the vacated niches. In the nineteenth century, when **paleontologists** (scientists who study extinct organisms) started, to study the fossil record in a systematic way, they quickly recognized that evidence of extinction is widespread and that most of the species that have *ever* lived on the Earth are now extinct. Indeed, as discussed in Chapter 7, those nineteenth century scientists divided the geological column into smaller units largely on the basis of the appearance and/or disappearance of certain key fossils.

Nowhere is the evidence of extinction more detailed or more striking than in the fossil record of marine animals. From that record we can estimate that the average time span of an ocean-dwelling species is about 4 million years. The record is equally clear, however, that organisms do not steadily disappear and new organisms do not steadily appear. Rather, the evidence suggests that a series of massive death-events have occurred, followed by the rapid appearance of new species that occupied vacant niches.

Paleontologists David Raup and J. John Sepkoski, Jr., are two of the leading researchers of the extinction record. Using the appearance and disappearance of genera rather than species, Sepkoski has analyzed the fossil record of 34,000 genera of marine fossils from the Cambrian to the present. Figure 16.20 is a diagrammatic representation of 19 extinction events he has identified. In order to estimate the magnitude of an extinction event, Sepkoski has calculated the percentage of the genera that disappeared at each event. Note that the greatest of the extinction events, at the Permian-Triassic boundary, led to the demise of 70 percent of all marine genera. The extinction event that marks the Cretaceous-Tertiary boundary, and that marked the end of the dinosaurs, was small by comparison, with only 45 percent of marine genera disappearing.

Figure 16.20 is not necessarily an accurate representation of the extent of extinction because some genera contained many more species than others. Raup and Sepkoski have attempted to extrapolate from genera to species and in so doing have estimated that the Permian-Triassic boundary event probably saw the extinction of 95 percent of all living species, while the Cretaceous-Tertiary boundary event may have seen the elimination of as many as 77 percent of all species.

What caused the extinctions? At this stage of research, we have no complete answers. The most widely held hypothesis for the Cretaceous-Tertiary event, for example, is the impact of a great meteorite and the consequent disruption of the climate. Despite the fact that a large-impact crater of just the right age has been identified beneath younger sedimentary rocks of the Yucatan Peninsula, extinction due to impact is still a hypothesis (though a very appealing one because of the favorable evidence). It is still a hypothesis because it is not yet clear just *bow* and *why* an impact would kill so many genera of organisms but leave an even larger number alive.

The most probable reason for the greatest killing ever in the geologic record, the Permian-Triassic boundary event, is, as discussed in Chapter 15, the plate tectonic assembly of Pangaea and the consequent loss of habitats. This, too, can only be considered a hypothesis until further research proves the concept to be correct. Other extinction events besides impacts and plate tectonics may be due to such various causes as falls in sea level, climatic changes, prolonged volcanic eruptions, or events of ocean anoxia (lack of oxygen).

Despite the lack of certainty about why many of the extinctions occurred, the fossil evidence is clear that they did. Many lessons can be read from the record, but three are of paramount importance. First, clearly, natural perturbations to the Earth system do happen. Second, and equally clearly, the biosphere is remarkably resilient, so that no matter how many species become extinct, new species always arise. The third lesson is a sobering one for us humans to appreciate—in the long run the biosphere is not dependent on any one species or even any one group of species.

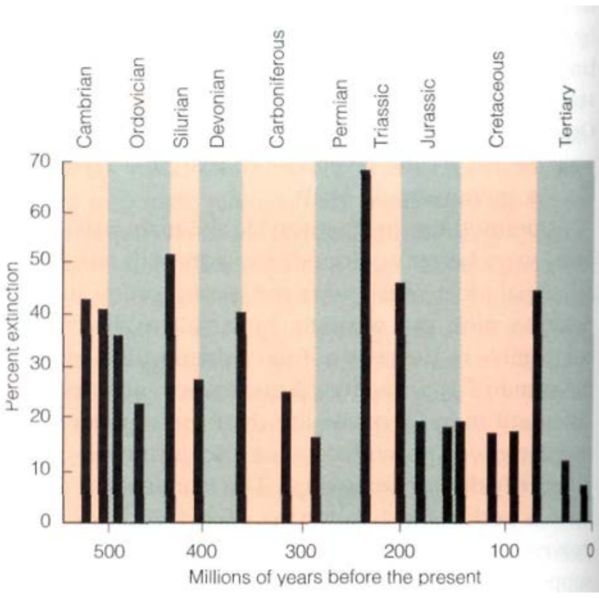


Figure 16.20 The extraordinary frequency of great extinction events that have occurred during the Phanerozoic Eon. The percentage of extinction was determined from the disappearance of genera of well-skeletonized animals.

Guest Essay

The Maturation of Earth and Life

Andrew H Knoll is professor of organismic and evolutionary biology, and also earth and planetary sciences, at Harvard University. His interest in the evolution of life, which was sparked during his undergraduate days at Lehigh University, was fanned to a flame under the tutelage of the late Elso Barghoorn, Knoll's advisor and predecessor at Harvard. An avid field geologist, Knoll has traversed the globe, spending nights in such remote areas as the high Arctic, the Australian Outback, and the Namib Desert of Namibia. In 1991 he was elected to the National Academy of Sciences.

In terms of its cosmological life expectancy, Earth is a middle-aged planet. In maturity, our planet contains large and widely dispersed continents, is glaciated at the poles, and is bathed in an oxygen-rich atmosphere. Perhaps most conspicuous, it supports an exuberance of life ranging from bacteria to redwoods. However, available portraits of the youthful Earth suggest a different countenance—a watery surface punctuated by volcanoes and limited areas of continental crust, an atmosphere rich in carbon dioxide and nearly devoid of oxygen, and no sign of biological activity.

The Earth's surface has changed markedly during the past 4 billion years. Given the close relationship between organisms and environment on the present-day Earth, it should not be surprising to learn that the two have evolved together through time, each influencing the other in a complex system of biogeochemical feedbacks.

The interplay between evolution and environment was particularly strong during our planet's early history. It is no overstatement to suggest that the evolutionary history of metabolism, explicit in evolutionary trees of both bacteria and cellular organisms with nuclei, mirrors the chemical evolution of the oceans and atmosphere. One facet of this coevolution that has particularly interested me concerns the explosive evolution of animals near the Proterozoic-Cambrian boundary (ca. 543 million years before the present). Anyone who has ever walked through a stratigraphic section spanning this boundary knows that the mineralized skeletons of invertebrates increase abruptly as one strides into the Cambrian, as do the abundance and diversity of animal tracks, trails, and burrows. All the extant phyla of "higher" animals—that is, animals characterized by complex organ systems and a shape that is symmetric about a central axis—appeared within the first 20 million years of the Cambrian Period. Exquisitely well-preserved fossils from the Burgess Shale and similar deposits suggest that this remarkable diversification also included types of animals that no longer exist. Simple, mostly unskeletonized fossils of sponges, cnidarians (sea anemones, jellyfish, and their relatives), and ancestral bilaterians appear only slightly earlier, perhaps 580 million years ago. Although this seems like a long time ago, fossils provide unambiguous evidence of microbial life as early as 3500 million years before the present. Why did animals radiate so late in the evolutionary day?

Among the several theories advanced to account for this pattern, one really fascinates me: it holds that prior to 580 million years ago, the Earth's atmosphere contained too little oxygen to support the biology of large animals. If true, this suggests that both the Earth and its biota leapt toward maturity near the end of the Proterozoic Eon. It also provides a most dramatic example of coevolution between life and environment.

A wealth of sedimentological data support the hypothesis of late Proterozoic environmental change. Extensive glaciogenic rocks and sedimentary iron deposits document large-scale changes in the atmosphere and oceans at this time. Knowledge of the present-day Earth leads us to expect that such variations should relate to changes in the systems of biogeochemical cycling that control our environment. My colleagues and I have confirmed this hypothesis by documenting strong variations through late Proterozoic time in the ratio of the two stable isotopes of carbon (^{13}C and ^{12}C) in marine carbonate minerals and organic matter. Because this ratio reflects the relative rates of carbonate and organic carbon burial of the sea floor, we have been able to infer remarkable variations in the late Proterozoic carbon cycle—with high rates of organic carbon burial characterizing the period just prior to the initial appearance of large animals. Other colleagues have documented the intricate tectonic dance of the continents during this period, with a late Proterozoic supercontinent splintering and reamalgamating, only to break apart once more as the Cambrian began. Tectonic events also leave an isotopic calling card—in this case the ratio of two isotopes of strontium (^{87}Sr and ^{86}Sr) in marine limestones. This ratio reflects the relative contributions of hydrothermal activity (low $^{87}Sr/^{86}Sr$) and continental weathering (variable but generally high $^{87}Sr/^{86}Sr$) to seawater. Strong late Proterozoic variation in the strontium isotopic content of limestones provides independent evidence of anomalously strong activity at midocean hydrothermal ridges, followed by the uplift of Himalaya-scale mountains.

These tectonic events provide clues to understanding the dynamic behavior of late Proterozoic climate and biogeochemistry.

As we learn more, the interval first singled out because of its biological importance turns out to be a time of remarkable tectonic, climactic, and biogeochemical change as well. There is as yet no direct means of measuring oxygen levels on the latest Proterozoic and Cambrian Earth, but geochemical models strongly support the hypothesis that atmospheric oxygen levels did indeed rise sharply 590 to 580 million years ago, just prior to the initial radiation of large animals. Thus, our planet may have come upon both biological and environmental maturity rather late in its development. Further biological innovations amplified by ecological interactions may have triggered the subsequent radiation that marks the Proterozoic-Cambrian boundary.

In short, a scientific odyssey that began with paleontology has led me to explore pivotal events in Earth's tectonic and environmental development. This has not been a diversion. On the contrary, it has convinced me that a satisfactory understanding of biological events near the end of the Proterozoic Eon hinges on our ability to place biology in the context of a dynamic planetary surface. Similarly, any understanding of the latest Proterozoic ice ages or atmospheric history requires that we treat Earth's tectonic, biogeochemical, climactic, and biological history as an integrated system.

The end-Proterozoic example is striking, but it is hardly unique. It may even be general. Environmental dynamics have exerted a profound influence on evolution, and wee *versa.* As the Earth sciences become more fully integrated, we gain the prospect of new insights into long-standing problems of Earth history. Equally exciting, we will be able to generate new questions that will occupy our imaginations for decades to come.

Summary

1. Life is the ability of an organism to grow, to reproduce, and to metabolize.
2. Growth in all living matter occurs by the polymerization of small organic molecules. Polymerization uses energy that it gets from the environment.
3. Bacteria, which are the simplest living organisms, and viruses, which are "almost" alive, cannot have been the first life forms because they are too complicated.
4. Life consists of cells. Cells with a nucleus are called eucaryotes, those without a nucleus are procaryotes.
5. No one knows how life began or what the first living cell was like but it must have been an anaerobic procaryote. Fossil procaryotes have been found in rocks 3.5 billion years old, so life appeared early in the Earth's history.
6. Both organisms and populations evolve. This means that the biosphere has evolved through geologic time and is still evolving today.
7. Because the biosphere has evolved, the Earth system has also evolved.
8. Biogeochemical cycles move chemical elements essential to the biosphere between the major Earth reservoirs. The most important biogeochemical elements are carbon, oxygen, nitrogen, and phosphorus.
9. Growth of living matter involves the construction of new cells. Details of cell production are passed on from cell to cell, generation after generation, through information stored in DNA.
10. The central problem of life is that proteins to build and run cells cannot be made without DNA and DNA won't work without proteins that serve as enzymes.
11. The three essential steps to form life are chemosynthesis, biosynthesis, and development of the complex cellular machinery needed for reproduction.
12. Chemosynthesis is the synthesis of small organic molecules such as amino acids from gases in the atmosphere or in space.
13. Biosynthesis is the polymerization of small organic molecules to form biopolymers such as proteins.
14. Some hypotheses suggest that life may have arisen in the sea, possibly near submarine hot springs. Another hypothesis, called panspermia, is that life arose in space and arrived on the Earth ready made.
15. The earliest life on the Earth was anaerobic.
16. A lot of oxygen was produced by photosynthesizing bacteria and neutralized by reducing agents such as dissolved iron in seawater. When the neutralizing capacity was exceeded, oxygen began to accumulate in the atmosphere. This started at least 1.5 billion years ago.

17. Eucaryotes arose from the procaryotes after oxygen started accumulating in the atmosphere. Primitive procaryotes get their energy by fermentation, whereas eucaryotes obtain their energy by using oxygen for respiration.
18. Procaryotes cannot reproduce sexually; nor can they associate to form three-dimensional, multicellular structures. Eucaryotes can both reproduce sexually and make three-dimensional structures.
19. The earliest multicellular organisms known are soft-bodied marine animals called the Ediacaran animals, after the place in Australia where they were first found.
20. The Cambrian Period, which began the Phanerozoic Eon, was a time of great diversity and population explosion among marine life. Some of the Cambrian phyla became extinct, but others survived and became the ancestors from which all modern forms of life descended.
21. The first skeletons, both internal and external, evolved in the Cambrian.
22. For life to leave the sea and live on land, four requirements have to be met: a structural support system; an internal aquatic system; a means of exchanging gases with air instead of water; and the availability of a moist environment for the reproductive system.
23. The first organisms to leave the sea were plants and insects. They did so in the early Silurian. The first land animals, amphibians, left the sea in the early Devonian.
24. The first plants were seedless. They gave rise to the naked-seeded gymnosperms, and the gymnosperms in turn gave rise to the enclosed-seed angiosperms.
25. The amphibians were succeeded by the reptiles. Amphibians have porous skins and require water in which to lay their eggs. Reptiles developed watertight skins and amniotic eggs and thus freed themselves from the water.
26. Mammals arose from and finally supplanted the reptiles by developing a more capable brain and a faster metabolism.
27. The fossil record reveals that the biosphere has been repeatedly disrupted by times of great extinctions. The causes of the extinctions are the subjects of research but appear to include giant meteorite impacts, plate tectonic rearrangements of continents, declines of sea level, prolonged volcanic eruptions, and severe climatic changes.

Important Terms to Remember

amino acid (p. 417)
angiosperm (p. 433)
biogeochemical cycles (p. 417)
biosynthesis (p. 419)
cell (p. 418)
chemosynthesis (p. 419)
cytoplasm (p. 422)
deoxyribonucleic acid (DNA) (p. 418)
ediacaran animals (p. 427)
enzyme (p. 419)
eucaryotic cell (eucaryotes) (p. 422)
evolution (p. 417)
gymnosperms (p. 433)
metabolism (p. 416)
orangelles (p. 422)
paleontologist (p. 436)
panspermia (p. 421)
procaryotic cell (procaryotes) (p. 422)
protein (p. 417)
ribonucleic acid (RNA) (p. 418)

Questions for Review

1. What are the essential characteristics of life?
2. Could life have started as a virus? Could it have started as bacteria? Explain why or why not.
3. It is thought that the Earth system has evolved through time because of the changes in the biosphere. What kind of changes in the biosphere could bring changes in the Earth system?
4. What are the three essential steps that must have occurred in order for life to form?
5. What is the Oparin "soup" hypothesis? Where and how did chemosynthesis occur in the Oparin hypothesis, and why is this part of the hypothesis now in doubt?
6. Describe the ways by which biosynthesis may have happened on the primitive Earth.
7. What characteristic must the earliest cells have had? Which organisms living today are thought to be representative of the most ancient life

forms?

8. What is the panspermia hypothesis, and why do some scientists consider it favorably?
9. Why cannot anaerobic cells have a nucleus?
10. What are the two kinds of cells, and how do they differ?
11. It is hypothesized that eucaryotes arose from procaryotes. Why is that conclusion held, and how might eucaryotes have developed?
12. Contrast the size, feeding properties, and aggregating abilities of procaryotic and eucaryotic cells. Why are eucaryotic cells more efficient?
13. Describe why the eucaryotic cells arose. When did they arise? Did all the procaryotes die when the eucaryotes developed?
14. How long ago did the first eucaryotic cells develop? When do fossils of multicelled animals formed by the aggregation of eucaryotic cells first appear in the geologic record?
15. Life expanded dramatically in the Cambrian Period and occupied innumerable ecological niches. What organisms occupied all the ecological niches during the Proterozoic Eon?
16. Give two hypotheses explaining why biological diversity exploded in the Cambrian Period.
17. What lessons can be learned from the fossil assemblages of the Burgess Shale?
18. When did life start leaving the sea and inhabiting the land? In what order did animals, insects, and plants become established on the land?
19. What four requirements had to be met in order for organisms to leave the sea and live on land? Are the requirements the same for plants, insects, and animals?
20. In what ways have plants modified their reproductive cycle since moving on to the land?
21. Describe how symbiosis has helped the angiosperms to develop.
22. Judged by their diversity, insects are a very successful group. Can you offer any possible reason for their success?
23. Describe two great inventions that dramatically changed the ability of animals to occupy new ecological niches.
24. Which animals first made their way from the sea to the land, and what group of animals evolved from them? Name two kinds of animals still living today that are representatives of the first group of land dwellers.
25. Amphibians are tied to the water. What characteristics did the reptiles develop that allowed them to free themselves from a dependence on water?
26. What evidence in the fossil record suggests that rapid extinctions have occurred many times in the past? Describe four hypotheses for the extinctions.

Questions for A Closer Look

1. Why is nitrogen an essential chemical element for life?
2. Describe the three forms in which nitrogen occurs in nature. In which form is it used by organisms?
3. How is nitrogen removed from the atmosphere and made accessible to organisms?
4. How is nitrogen returned to the atmosphere in order to keep the main reservoir of nitrogen in balance?
5. None of the higher plants can fix nitrogen gas from the atmosphere. How do they get their nitrogen?

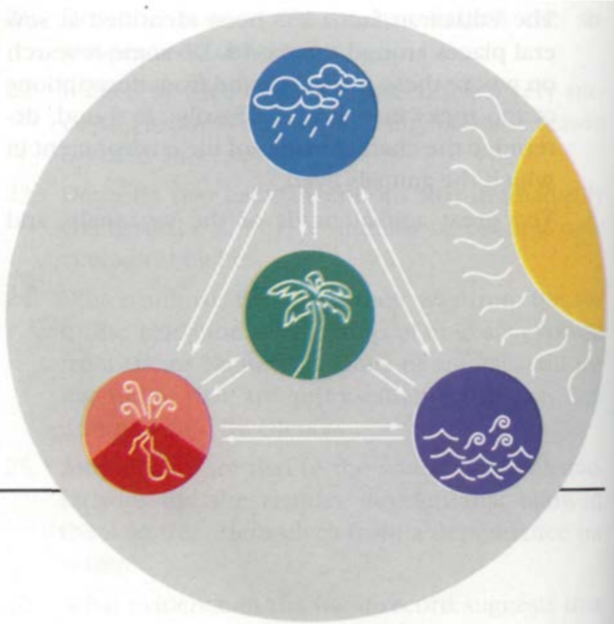
Questions for Discussion

1. Could some form of life as we know it on Earth have developed on any of the other bodies of the solar system? If you had an unmanned space craft available to visit other planets and moons, what tests would you carry out to see if life now exists or ever existed there?
2. There are many intriguing but still unanswered questions concerning the history of the biosphere. Discuss the following questions:
   a. In the absence of fossils, what kinds of evidence might be used to determine when the biosphere came into being?
   b. What kind of research would you carry out in order to try and answer why the great Cambrian explosion of biodiversity occurred and why animals started to develop skeletons?
3. Choose one of the great extinctions in the fossil record, research it, and discuss evidence bearing on the cause of the extinction.

4. The Ediacaran fauna has been identified at several places around the world. Do some research on where these places are and from descriptions of the rocks in which the fossils are found, determine the characteristics of the environment in which the animals lived.
5. The great animal herds of the savannahs and prairies such as buffalo and antelope appeared during the Tertiary Period and seem to be coincident with the appearance of grasses. Do some research on where and when the herds appeared. Discuss what effect climate change may have played in the rise of grasses and grass-grazing animals.

PART SIX

Living on the Earth



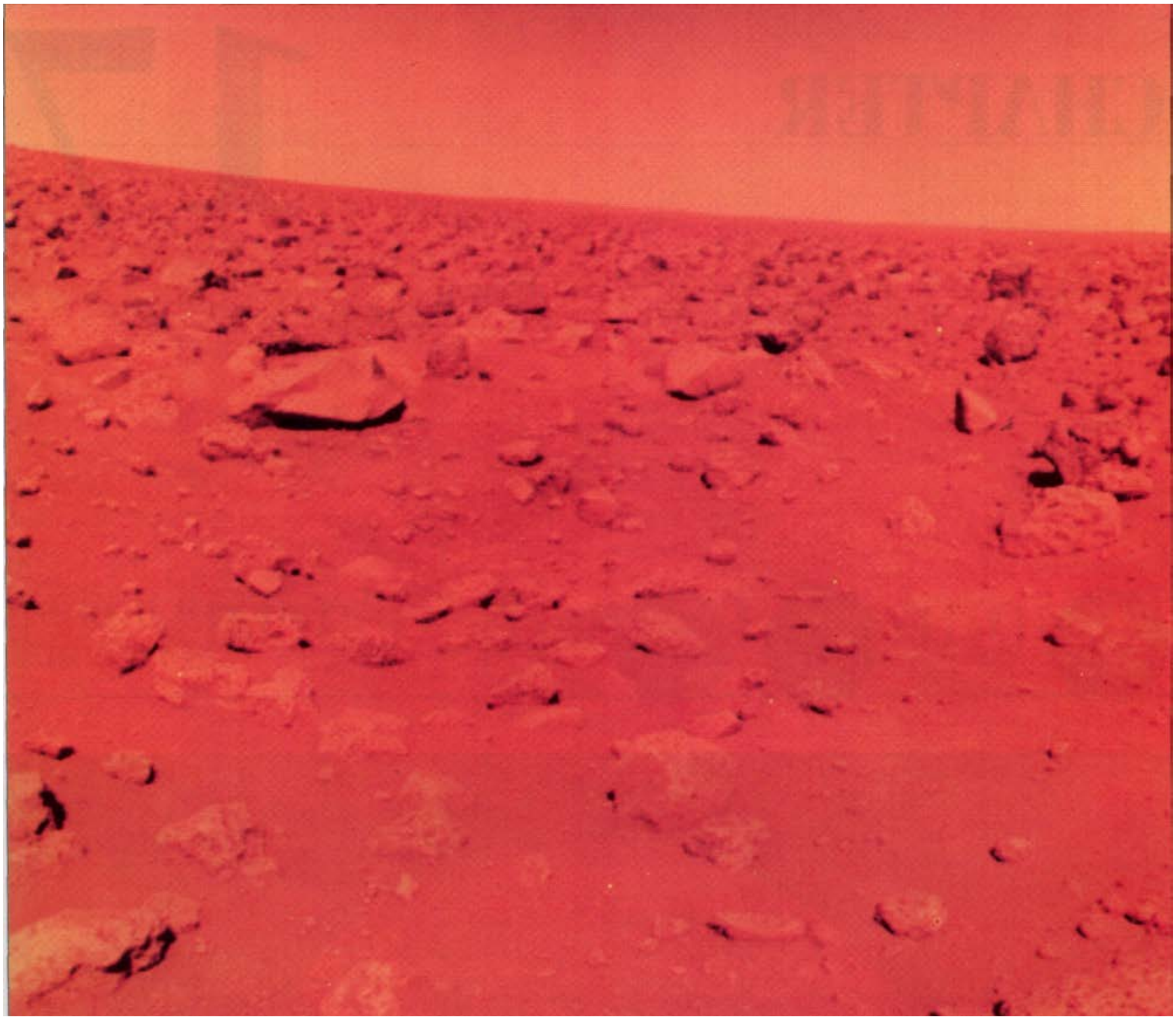
Terraforming

When astronauts Armstrong, Collins, and Aldrin returned from the Moon with the first lunar samples, a new scientific horizon had been reached. The year was 1969, and for the first time it was possible to entertain seriously the possibility that humans might live and work beyond planet Earth. Many of the early hopes have been found impractical, but two continue to be considered: one is to make one of the other planets habitable, and the other is to mine other bodies in the solar system.

Some scientists regard the concept of changing a planet to make it habitable an environmental monstrosity, but others look at the notion with favor and have coined the term *terraforming* for changing a hostile planet into something like the Earth.

The candidate planet is Mars, and the time needed to effect a transformation is estimated to be up to 100,000 years. The first phase of the transformation would be the construction of factories on the surface of Mars. The factories would produce greenhouse gases and release them into the thin Martian atmosphere. As the gases accumulated, they would raise the atmospheric pressure and also trap more of the Sun's heat, thereby heating the surface. This would cause the carbon dioxide snow on the Martian poles to melt and increase the CO_2 content of the atmosphere. Eventually, the surface temperature on Mars, presently a chilly -60°C, would warm to above 0° and water would be able to flow on the surface. Plants would be able to grow in a warm CO_2-rich atmosphere, but, of course, humans could not breathe the air until photosynthesis built up the oxygen content of the atmosphere. Initially, humans would have to live in inflatable, domed cities in which a breathable atmosphere could be maintained and carry an oxygen supply with them whenever they went outside to tend to growing plants. Finally, after 100,000 years, Mars would develop an atmosphere like that of the Earth.

Terraforming verges on the edge of science fiction, but getting resources from outside the Earth is not so far-fetched. Among the ideas that are being consid-



The bleak surface of Mars; could it ever be made habitable? This is the first color picture taken by the spacecraft Viking 2 after it landed on the Martian surface in 1976. The photo confirms what astronomers had suspected for a long time—that Mars is reddish colored because the Martian regolith is red. Experiments carried out by Viking 2 seeking evidence of past or present forms of microscopic life in the Martian regolith were negative.

ered is the notion of mining nickel and iron from one of the asteroids. Because the asteroids are small, their gravitational pull is weak, and so it would be easy to lift material from the surface. Mining ores on larger bodies such as the Moon and Mars is an appealing but a difficult notion. Large bodies have strong gravitational pulls that make it expensive to lift things off the surface. Thus, on large bodies it is practical only to use resources there, not to return them to the Earth. There is an even more difficult problem—it is not even clear whether rich ores of the kind found on the Earth occur on the Moon or Mars.

On the Earth most of the rich ores formed as a result of the atmosphere, hydrosphere, or biosphere, or all three, interacting with the lithosphere. Lacking such interactions—because the Earth is the only body in the solar system with a biosphere and a hydrosphere—there is no reason to think that other bodies house mineral treasures like those on the Earth.

Fascinating though the idea of extending our civilization into space may be, the likelihood of it coming to fruition seems very remote. If anything does happen, it will surely be far in the future. In the meantime, we have to learn to live with the resources available on the Earth and to use those resources in ways that keep planet Earth habitable.

CHAPTER 17

Resources from the Earth

Many cultures around the world discovered the distinctive properties of gold and developed the skills to work with it. This mask from Columbia, now in the Bogota Gold Museum, was made about 2,000 years ago by indigenous peoples living in the region.

Natural Resources and Civilization

Civilization and natural resources—the former would not have been possible without the latter. Scholars even mark the stages of civilization by the natural resources our ancestors learned to use; the Stone Age, Bronze Age, and Iron Age are examples. The first resources our ancestors used were the fruits, grain, animals, and fish they ate. These first-used materials were *renewable resources* because new supplies grew each season.

Millions of years ago our ancestors also started to use a different class of natural resources. When they picked up suitably shaped stones and used them as hunting aids, they were using *nonrenewable resources*, so-called because they are not replenished by seasonal growth. Because the most desired stones were found in only a few restricted places, trading started. Next our ancestors started gathering and trading another nonrenewable resource, salt. Originally, dietary needs for salt were satisfied by eating the meat brought home by hunters. When farming started, however, diets became cereal based and extra salt was needed. We don't know when or where the mining of salt started but long before recorded history salt routes criss-crossed the globe.

Metals were first used more than 17,000 years ago. Both copper and gold are found as native metals, and these were the earliest metals to be used. But native copper is rare, and so eventually other sources of copper were needed. By 6000 years ago, our ancestors had learned how to extract copper from certain minerals by smelting. Before another thousand years had passed, they had discovered how to smelt minerals of lead, tin, zinc, silver, and other metals. The technique of mixing metals to make alloys came next; bronze (copper and tin) and pewter (tin, lead, and copper) came into use. Because the smelting of iron is more difficult than the smelting of copper, development of an iron industry came much later—about 3300 years ago.

The first people to use oil (a nonrenewable resource) instead of wood (a renewable resource) for fuel were the Babylonians, about 4500 years ago. The Babylonians lived in what is now Iraq, and they used oil from natural seeps in the valleys of the Tigris and Euphrates rivers. The first people to mine and use coal were the Chinese, about 3100 years ago. At about the same time, the Chinese drilled the first wells for natural gas; some were nearly 100 m (330 ft) deep.

By the time first the Greek and then the Roman empires came into existence about 2500 years ago, our ancestors had come to depend on a very wide range of nonrenewable resources—not just metals and fuels, but also processed materials such as cements, plasters, glasses, and porcelains. The list of materials we mine, process, and use has grown steadily larger ever since. Today we have industrial uses for almost all of the naturally occurring chemical elements, and more than 200 kinds of minerals and fuels are mined and used. Of course, we still harvest renewable resources, but society is now totally dependent on supplies of nonrenewable resources too.

MINERAL RESOURCES

Can you imagine a world without machines? Our modern world with its 5.5 billion people couldn't operate without them. Nor could it operate without bricks and cement, fertilizers, or plastics. All of these things and many more are made from mineral resources. Each of us now relies directly or indirectly (meaning through industry and public works) on a very large annual input of mineral resources (Fig. 17.1).

Machines produce our food, make our clothes, transport us, and help us communicate. Bricks and cement are used to build houses and roadways, salts are used for chemicals and fertilizers, and plastics are made from coal and oil. Everything—metals, fuels, fertilizers, chemicals, and building materials—are dug or pumped from deposits in the Earth. Deposits of minerals and fuels are formed by geological processes. The well-being of everyone is therefore controlled by geological processes.

In some of the previous chapters, we pointed out how geological processes such as weathering, sedimentation, and volcanism can, under suitable conditions, form concentrations of valuable minerals and rocks. The "suitable conditions" are not common, however, and for this reason mineral and energy deposits are few and hard to find. No geological challenge is more difficult than the search for, and discovery of, new resources of minerals and energy. No societal problem is more pressing than the recovery of natural resources with a minimal disruption to the environment. Finding resources and managing the consequences of using them are two of the greatest problems facing human beings today.

We turn first to the minerals we mine. Mineral resources are nonrenewable and can be divided into two groups by the way they are used.

1. *Metallic minerals* are those from which metals such as copper, iron, gold, and zinc can be recovered by smelting.
2. *Nonmetallic minerals* are those used for their physical or chemical properties rather than for the chemical elements they contain; examples are salt, gypsum, sodium carbonate, calcium fluoride, and clay for bricks.

Following a discussion of mineral resources we turn to energy resources. Some energy resources, such as coal and uranium, are nonrenewable resources because they are rocks and minerals formed by the same geological processes that form metallic and nonmetallic minerals. Other energy sources, such as sunlight, wind, and running water, are renewable and have quite different properties than the nonrenewable resources.

Supplies of Minerals

Many industrialized nations contain rich **mineral deposits** (any volume of rock containing an enrichment of one or more minerals) that they are exploiting vig-

Figure 17.1 The average amount of material consumed per person per year (called the per capita consumption) is greater in an industrially advanced country such as the United States than it is for the world as a whole.

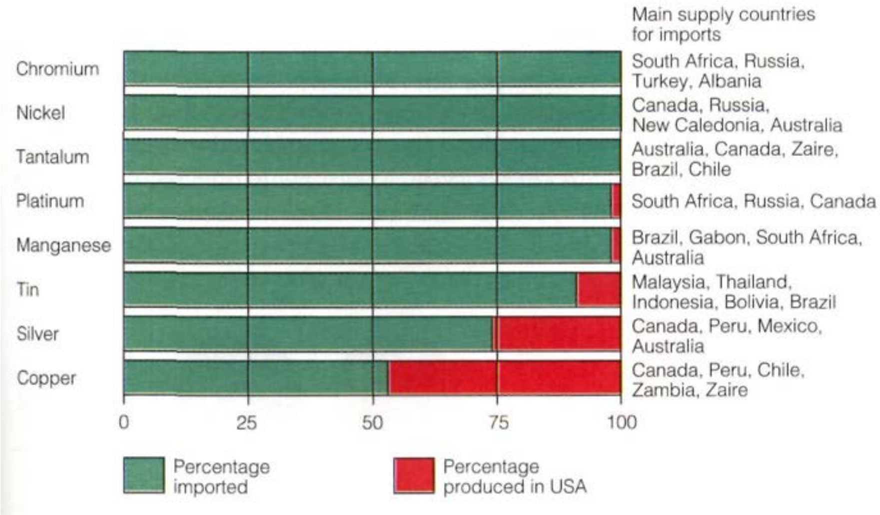


Figure 17.2 Selected mineral substances for which the United States' consumption exceeds production. The difference must be supplied by imports. Data are plotted for 1990, but the percentage changes little from year to year.

orously. Yet no nation is entirely self-sufficient in mineral supplies, and so each must trade with other nations to fulfill its needs (Fig. 17.2).

All mineral resources have three peculiarities that influence their use:

1. Usable minerals are limited in abundance and are localized within the Earth's crust. This is the main reason why no nation is self-sufficient where mineral supplies are concerned.
2. The quantity of a resource available in any one country is never known with accuracy because the likelihood that new deposits will be discovered is difficult to assess. A country that today can supply its needs for a given mineral substance may face a future in which it will become an importing nation. A little more than a century ago, for example, Britain was a great mining nation, producing and exporting tin, copper, tungsten, lead, and iron. Today, the known deposits have been worked out.
3. Unlike fruits and grains, which can be seasonally cropped and thus replenished, deposits of minerals are depleted by mining and are eventually exhausted. This disadvantage can be offset only by finding new occurrences or by using the same material repeatedly—that is, by recycling and making use of scrap.

Much ingenuity has been expended in bringing the production of minerals needed by society to its present state. Because known deposits are being rapidly exploited while demand for minerals continues to increase as the world's population grows ever larger, we can be sure that even more ingenuity will be needed in the future.

Ore

Minerals are sought in deposits from which the desired substances can be recovered least expensively. The more concentrated the desired minerals, the more valuable the deposit. In some deposits the desired minerals are so highly concentrated that even very rare substances such as gold and platinum can be seen with the naked eye. For every desired mineral substance, a *grade* (level of concentration) exists below which the deposit cannot be worked economically (Fig. 17.3). To distinguish between profitable and unprofitable mineral deposits, we use the word **ore,** meaning an aggregate of minerals from which one or more minerals can be extracted profitably. It is not always possible to say exactly what the grade must be, nor how much of a given mineral must be present, in order to constitute an ore. Two deposits may have the same grade and be the same size, but one is ore and the other is not. There could be many reasons for the difference. For example, the uneconomic deposit could be too deeply buried or located in so remote an area that the costs of mining and transport would be so high that the final product would not be competitive with the same product from other deposits. Furthermore, as costs and market prices fluctuate, a particular aggregate of minerals may be an ore at one time but not at another.

Ore minerals such as sphalerite, galena, and chalcopyrite from which zinc, lead, and copper, respec-

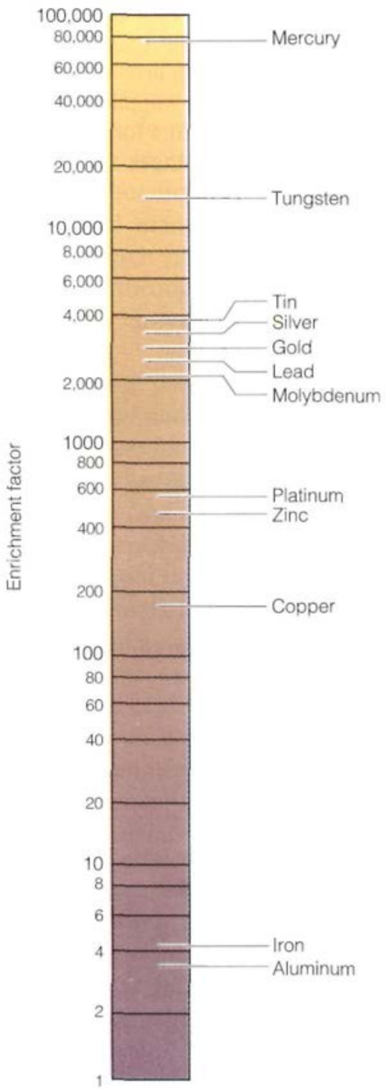


Figure 17.3 Before a mineral deposit can be worked profitably, the percentage of valuable metal in the deposit must be greatly enriched above its average percentage in the Earth's crust. The enrichment is greatest for metals that are least abundant in the crust, such as gold and mercury. As mining and mineral processing have become more efficient and less expensive, it has been possible to work leaner ore, and so there has been a historic decline in enrichment factors. Declines have ceased over the past 20 years, and for some metals enrichment factors have increased slightly. Note that the scale is a magnitude (logarithmic) scale, in which the major divisions increase by multiples of ten.

tively, are extracted are usually mixed with other minerals, collectively termed **gangue** (pronounced gang). Familiar minerals that commonly occur as gangue are quartz, feldspar, mica, calcite, and dolomite.

The ore challenge is twofold: (1) to find the ores (which altogether underlie an infinitesimally small proportion of the Earth's land area); and (2) to mine the ore and get rid of the gangue as cheaply and cleanly as possible. Both steps are technical problems; engineers have been so successful in solving them that some deposits now considered ore are only one-sixth as rich as the lowest grade ores were 100 years ago.

ORIGIN OF MINERAL DEPOSITS

All ores are mineral deposits because each of them is a local enrichment of one or more minerals or mineraloids. The reverse is not true, however. Not all mineral deposits are ores. *Ore* is an economic term, whereas *mineral deposit* is a scientific term. How, where, and why a mineral deposit forms is the result of one or more geological processes. Whether or not a given mineral deposit is an ore is determined by how much we human beings are prepared to pay for its content. Fascinating though the economics of ores and mining is, this topic cannot be explored in this volume. Instead, discussion is limited to the origin of mineral deposits without necessary regard to questions of economics.

In order for a deposit to form, some process or combination of processes must bring about a localized enrichment of one or more minerals. A convenient way to classify mineral deposits is through the principal concentrating process. Minerals become concentrated in five ways:

1. Concentration by hot, aqueous solutions flowing through fractures and pore spaces in crustal rock to form **hydrothermal mineral deposits.**
2. Concentration by magmatic processes within a body of igneous rock to form **magmatic** mineral **deposits.**
3. Concentration by precipitation from lake water or seawater to form **sedimentary mineral deposits.**
4. Concentration by flowing surface water in streams or along the shore to form **placers.**
5. Concentration by weathering processes to form **residual mineral deposits.**

Hydrothermal Mineral Deposits

Many of the most famous mines in the world contain ores that were formed when their ore minerals were deposited from hydrothermal solutions. It is probable that more mineral deposits have been formed by deposition from hydrothermal solutions than by any other mechanism. However, the origins of hydrothermal solutions are often difficult to decipher. Some solutions originate when water dissolved in a magma is released as the magma rises and cools. Other solutions are formed from rainwater or seawater that circulates deep in the crust. (For a discussion of what is known about modern deposit-forming solutions, see "A Closer Look: Modern Hydrothermal Mineral Deposits.")

The heat source for seawater hydrothermal solutions of the kind illustrated in Figure 17.4A and B is spreading center volcanism. Because the ore minerals deposited are always sulfides, mineral deposits formed from such solutions are called *volcanogenic massive sulfide deposits.*

The ore-mineral constituents in volcanogenic massive sulfide deposits originate from the igneous rocks of the oceanic crust. Heated seawater reacts with the rocks it is in contact with, causing changes in both mineral composition and solution composition. For example, feldspars are changed to clays and epidote, and pyroxenes are changed to chlorites. As the minerals are transformed, trace metals such as copper and zinc, present by atomic substitution, are released and become concentrated in the slowly evolving hydrothermal solution.

Causes of Precipitation

When a hydrothermal solution moves slowly upward, as with groundwater percolating through an aquifer, the solution cools very slowly. If dissolved minerals were precipitated from such a slow-moving solution, they would be spread over great distances and would not be sufficiently concentrated to form an ore. But when a solution flows rapidly, as in an open fracture through a mass of shattered rock, or through a layer of porous tephra where flow is less restricted, cooling can be sudden and happen over short distances. Rapid precipitation and a concentrated mineral deposit are

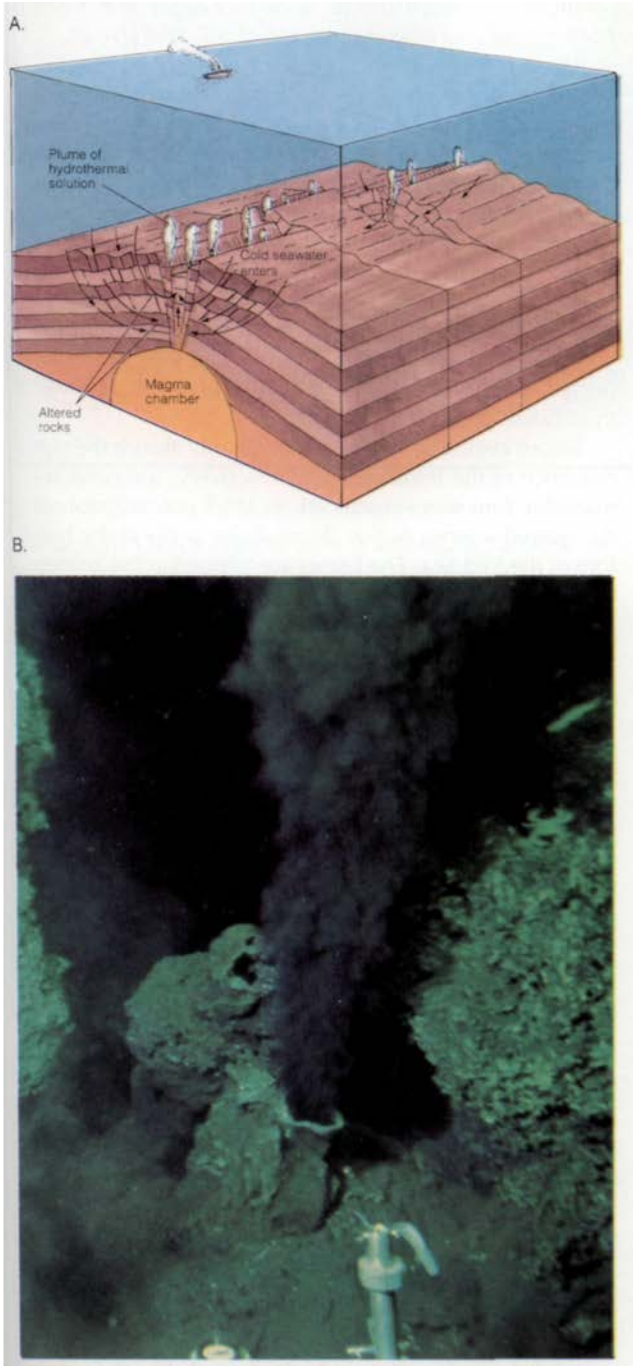


Figure 17.4 Hydrothermal solutions form mineral deposits on the seafloor. A. Seawater penetrates volcanic rocks on the seafloor at a spreading center. Heated by a magma chamber, seawater becomes a hydrothermal solution, alters rocks it passes through extracting metals in the process, and rises at a midocean ridge as a hydrothermal plume. B. A so-called black smoker photographed at a depth of 2500 m below sea level on the East Pacific Rise at 21°N latitude. The "smoker" has a temperature of 320°C. The rising hydrothermal solution is actually clear; the black color is due to fine particles of iron sulfide and other minerals precipitated from solution as the plume is cooled through contact with cold seawater. The chimneylike structure is composed of pyrite, chalcopyrite, and other ore minerals deposited by the hydrothermal solution.

A Closer Look

Modern Hydrothermal Mineral Deposits

Three extraordinary discoveries over a 15-year period changed our thinking about hydrothermal mineral deposits. The first discovery, in 1962, was accidental. Until that year, no one was sure where to look for modern hydrothermal solutions or even how to recognize one when it was found. Drillers seeking oil and gas in the Imperial Valley of southern California were astonished when they struck a 320°C (608°F) brine at a depth of 1.5 km (0.9 mi). As the brine flowed upward, it cooled and precipitated minerals it had been carrying in solution. Over three months, the well deposited 8 tons of siliceous scale containing 20 percent copper and 8 percent silver by weight. The drillers found a hydrothermal solution that could, under suitable flow conditions, form a rich mineral deposit.

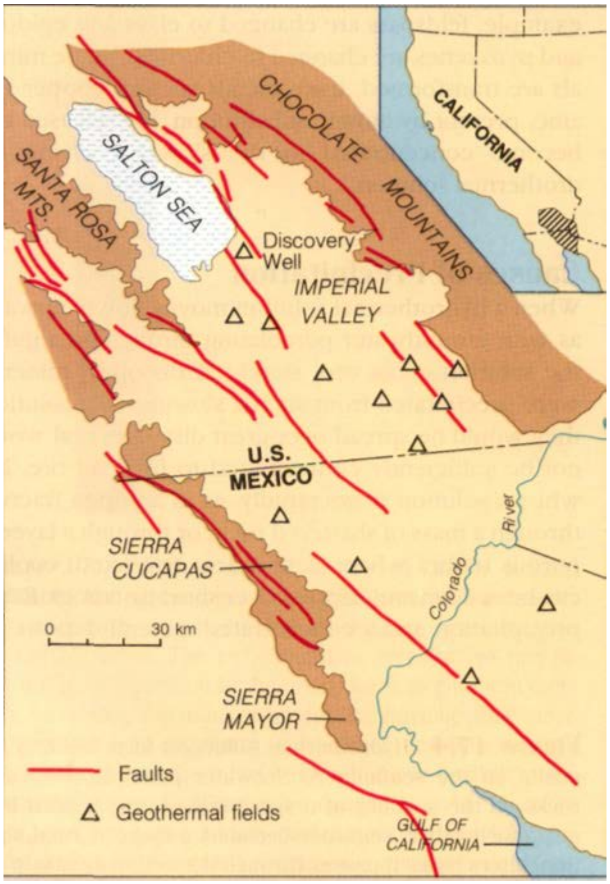


Figure CI7.1 The Imperial Valley graben (also known as the Salton Trough). The graben is bounded by the Chocolate Mountains on the east and the Santa Rosa Mountains on the west. Hydrothermal solutions were discovered in a well drilled on the southern end of the Salton Sea. Places where geothermal activity is known, and where other hydrothermal solutions may be present at depth, are marked with triangles.

The Imperial Valley is a sediment-filled graben covering the join between the Pacific and North American plates, where the East Pacific Rise passes under North America (Fig. C17.1). Volcanism is the source of heat for the brine solution discovered in 1962. These brines provided the first unambiguous evidence that hydrothermal solutions can leach metals such as copper and silver from ordinary sediments.

The Imperial Valley brines also answered some important questions about the composition of hydrothermal solutions. The solubilities in water of ore minerals such as sphalerite (ZnS) and galena (PbS) are so incredibly low that geologists have long puzzled as to how a solution could transport both metals and sulfur in the same solution. The answer is that in brines of appropriate composition, Cl^- ions form anionic complexes such as $(ZnCl_4)^{2-}$ which shield the metal ions from S^{2-} ions in solution and thus effectively raise the solubility. Chloride complexing, as the process is called, had long been suspected, but it was the Imperial Valley brines that proved their existence.

Before geologists had a chance to fully absorb the significance of the Imperial Valley discovery, a second remarkable find was announced. In 1964 oceanographers discovered a series of hot, dense, brine pools at the bottom of the Red Sea. The brines are trapped in the graben formed by the spreading center between the Arabian and African plates (Fig. C17.2), and they are so much more saltier, and therefore denser, than seawater that they remain ponded in the graben even though they are as hot as 60°C (140°F). Many such brine pools have now been discovered.

The Red Sea brines rise up the normal faults associated with the central rift of a spreading center and, like the Imperial Valley brines, have evolved to their present compositions through reactions with the enclosing rocks. The Red Sea brine discovery was surprising, but even more surprising was the discovery that sediments at the bottom of the pools contained ore minerals such as chalcopyrite, galena, and sphalerite. In other words, the oceanographers had discovered modern stratabound mineral deposits in the process of formation.

The third remarkable discovery was really a series of discoveries that commenced in 1978. Scientists using deep-diving submarines made a series of dives on the East Pacific Rise at 21°N latitude. To their amazement, they found 300°C (572°F) hot springs emerging from the seafloor 2500 m (2700 yd) below sea level. Around the hot springs lay a blanket of sulfide minerals. The submariners watched a modern volcanogenic massive sulfide deposit forming before their eyes.

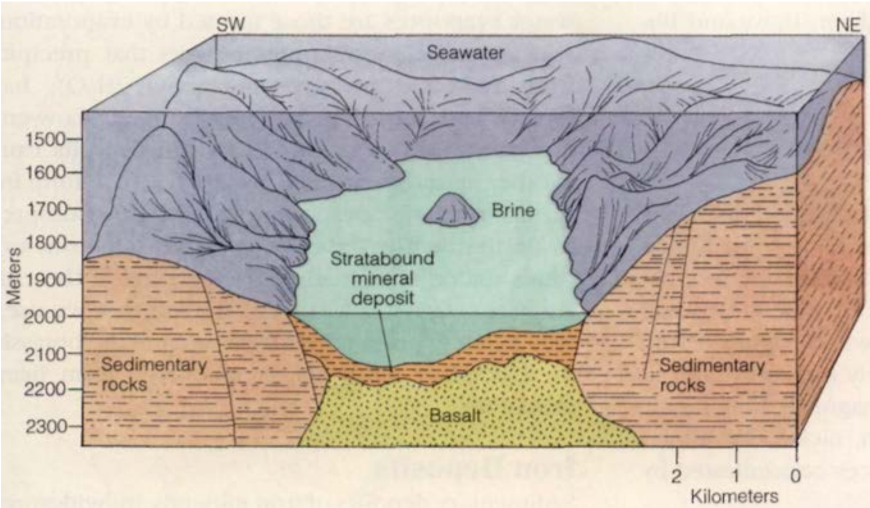


Figure C17.2 Topography of the Red Sea graben near the Atlantis II brine pool. Hot, dense brines rise up normal faults, pond on the floor of the graben, and form stratabound deposits rich in copper and zinc.

Each of the discovery sites—Imperial Valley, the Red Sea, and 21°N—is on a spreading center, so there is no doubt that the deposits are forming as a result of plate tectonics. Soon the hunt was on to see if seafloor deposits could be found above subduction zones. In 1989 a joint German-Japanese oceanographic expedition to the western Pacific discovered the first modern subduction-related deposits, and two years later Canadian and Australian scientists discovered another modern subduction-related deposit in the Manus basin just north of Papua-New Guinea. No longer are geologists limited to speculating about how certain mineral deposits *might* have formed. Today they can be studied as they grow.

the result. Other effects—such as boiling, a rapid decrease in pressure, composition changes of the solution caused by reactions with adjacent rock, and cooling as a result of mixing with seawater—can also cause rapid precipitation and form concentrated deposits. When valuable minerals are present, an ore can be the result.

Examples of Precipitation

Veins form when hydrothermal solutions deposit minerals in open fractures, and many such veins are found in regions of volcanic activity (Fig. 17.5). The famous gold deposits at Cripple Creek, Colorado, were formed in fractures associated with a small caldera, and the huge tin and silver deposits in Bolivia are in fractures that are localized in and around stratovolcanoes like those shown in Figure 6.17. In each case the fractures formed as a result of volcanic activity, and the magma chambers that fed the volcanoes served as the sources of the hydrothermal solutions that rose up and formed the mineralized veins.

A cooling granitic stock or batholith is a source of heat just as the magma chamber beneath a volcano is—and it can also be a source of hydrothermal solutions. Such solutions move outward from a cooling stock and will flow through any fracture or channel, altering the surrounding rock in the process and commonly depositing valuable minerals. Many famous ore bodies are associated with intrusive igneous rocks.

Figure 17.5 A rich vein in Potosi, Bolivia, containing chalcopyrite, sphalerite, and galena cutting andesite. The andesite has been altered by the hydrothermal solution that deposited the ore minerals.

The tin deposits of Cornwall, England, and the copper deposits at Butte, Montana; Bingham, Utah; and Bisbee, Arizona, are examples.

Magmatic Mineral Deposits

The processes of melting and crystallization discussed in Chapter 5 are two ways of separating some minerals from others. Fractional crystallization in particular can lead to the creation of valuable mineral deposits such as the chromite layers shown in Figure 5.16B. The processes involved are entirely magmatic, and so such deposits are referred to as magmatic mineral deposits. Chromium, iron, platinum, nickel, vanadium, and titanium are the main resources concentrated by fractional crystallization.

Pegmatites, which are very coarse-grained igneous rocks formed by fractional crystallization of granitic magma, commonly contain rich concentrations of elements such as lithium, beryllium, cesium, and niobium. Much of the world's lithium is mined from pegmatites such as those at King's Mountain, North Carolina, and Bikita in Zimbabwe. The great Tanco pegmatite in Manitoba, Canada, produces much of the world's cesium, and pegmatites in many countries yield beryl, one of the main ore minerals of beryllium.

Sedimentary Mineral Deposits

The term *sedimentary mineral deposit* is applied to any local concentration of minerals formed through processes of sedimentation. Any process of sedimentation can form localized concentrations of minerals, but it has become common practice to restrict use of the term *sedimentary* to those mineral deposits formed through precipitation of substances carried in solution.

Evaporite Deposits

The most direct way in which sedimentary mineral deposits form is by evaporation of lake water or seawater. The layers of salts that precipitate as a consequence of evaporation are called evaporite deposits.

Examples of salts that precipitate from lake waters of suitable composition are sodium carbonate (Na_2CO_3), sodium sulfate (Na_2SO_4), and borax ($Na_2B_4O_7{\cdot}10H_2O$). Huge lake-water evaporite deposits of sodium carbonate were laid down in the Green River basin of Wyoming during the Eocene Epoch. Borax and other boron-containing minerals are mined from evaporite lake deposits in Death Valley and Searles and Borax lakes, all in California, and in Argentina, Bolivia, Turkey, and China.

Much more common and important than lake-water evaporites are those formed by evaporation of seawater. The most important salts that precipitate from seawater are gypsum ($CaSO_4{\cdot}2H_2O$), halite ($NaCl$), and carnallite ($KCl{\cdot}MgCl_2{\cdot}6H_2O$). Low-grade metamorphism of marine evaporite deposits causes another important mineral, sylvite (KCl), to form from carnallite. Marine evaporite deposits are widespread; in North America, for example, strata of marine evaporites underlie as much as 30 percent of the entire land area (Fig. 17.6). Most of the salt that we use, as well as the gypsum used for plaster and the potassium used in plant fertilizers, is recovered from marine evaporites.

Iron Deposits

Sedimentary deposits of iron minerals are widespread, but the amount of iron in average seawater is so small that such deposits cannot have formed from seawater that is the same as today's seawater.

All sedimentary iron deposits are tiny by comparison with the class of deposits characterized by the *Lake Superior-type iron deposits.* These remarkable deposits, mined principally in Michigan and Minnesota, were long the mainstay of the United States steel industry but are declining in importance today as imported ores replace them. The deposits are of early Proterozoic age (about 2 billion years or older) and are found in sedimentary basins on every craton, particularly in Labrador, Venezuela, Brazil, Russia, India, South Africa, and Australia. Every aspect of the Lake Superior deposits indicates chemical precipitation. The deposits are interbedded layers of chert and iron minerals. Because the deposits are so large, it is inferred that the iron and silica must have been transported in surface water, but the cause of precipitation remains unknown. Many experts suspect that Lake Superior-type deposits may be ancient marine evaporites that formed from seawater of a different composition from today's seawater.

Lake Superior-type iron deposits are not ores. The grades of the deposits range from 15 to 30 percent Fe by weight, and the deposits are so fine-grained that the iron minerals cannot be easily separated from the gangue. Two additional processes can form ore. First, leaching of silica during weathering can lead to a local enrichment of iron and produce ores containing as much as 66 percent Fe. Compare Figure 17.7A, which is a Lake Superior-type iron deposit in the Hamersley Range, Western Australia, with Figure 17.7B, a sample of ore developed by secondary enrichment in the Hamersley Range. The rocks in Figure 17.6A contain about 25 percent Fe, whereas those in Figure 17.6B have had most of the silica leached out by weathering and contain about 60 percent Fe.

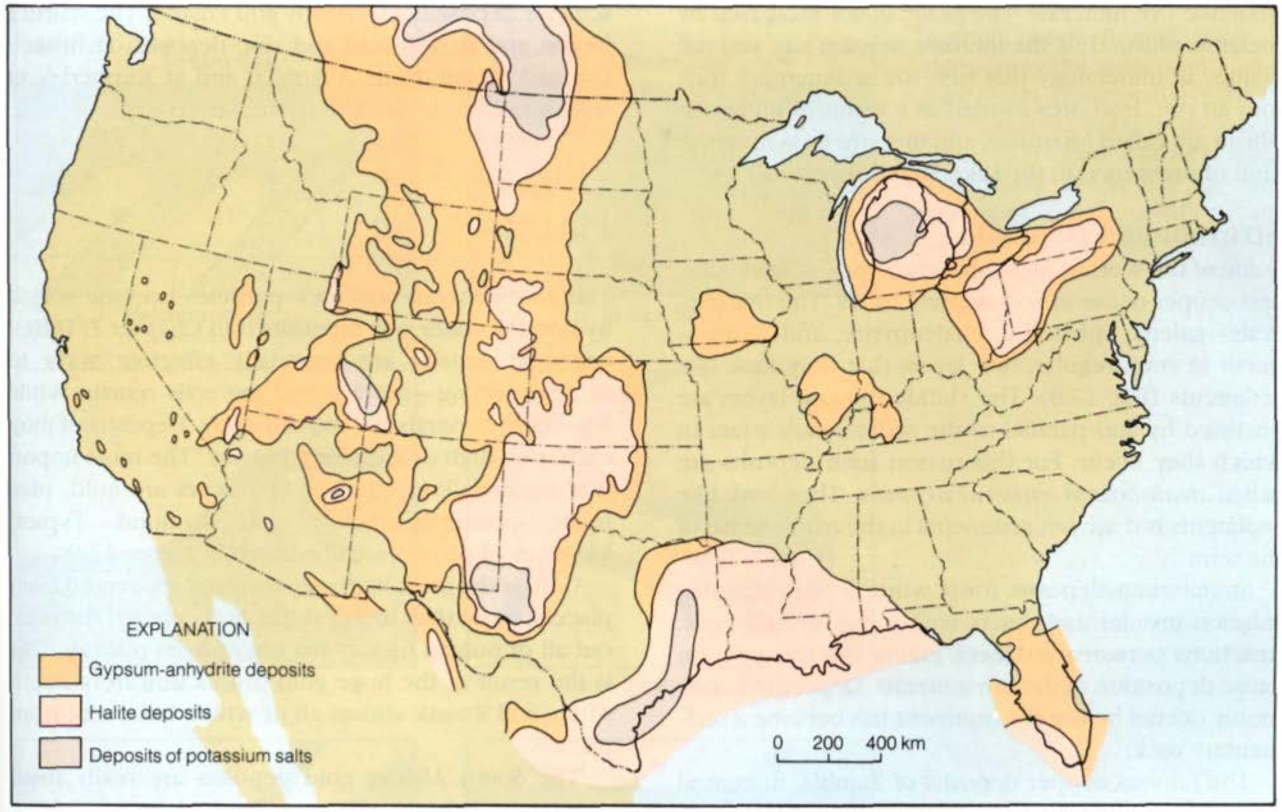


Figure 17.6 Portions of the United States known to be underlain by marine evaporite deposits. The areas underlain by gypsum and anhydrite do not contain halite. The areas underlain by potassium salts are also underlain by halite and by gypsum and anhydrite.

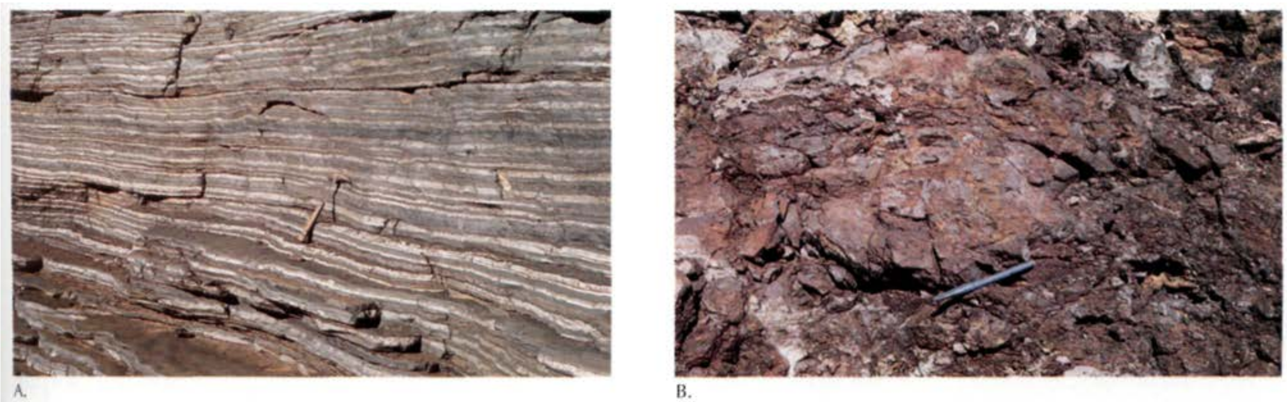


Figure 17.7 Sedimentary iron deposit of the Lake Superior type. A. Unaltered iron-rich sediments of the Brockman Iron Formation in Hamersley Range of Western Australia. The white layers are largely chert, whereas the darker bluish and reddish layers consist mainly of iron-rich silicate, oxide, and carbonate minerals. The grade is about 25 percent iron. B. Altered iron-rich sediment from the same formation shown in A. Leaching of silica during weathering has formed a secondarily enriched mass of iron minerals that is rich enough to be an ore. The grade is about 60 percent iron.

The second way a Lake Superior-type iron deposit can become an ore is through metamorphism. Two changes occur as a result of metamorphism. First, grain sizes increase so that separating ore minerals from the gangue becomes easier and cheaper. Second, new mineral assemblages form, and iron silicate and iron carbonate minerals originally present can be replaced by magnetite or hematite, both of which are desirable ore minerals. The grade is not increased by metamorphism. It is the increase in grain size and the change in mineralogy that turn the sedimentary rock into an ore. Iron ores formed as a result of metamorphism are called *taconites,* and they are now the main kind of ore mined in the Lake Superior region.

Stratabound Deposits

Some of the world's most important ores of lead, zinc, and copper occur in sedimentary rocks. The ore minerals—galena, sphalerite, chalcopyrite, and pyrite—occur in such regular, fine layers that they look like sediments (Fig. 17.8). The sulfide mineral layers are enclosed by and parallel to the sedimentary strata in which they occur. For this reason such deposits are called *stratabound mineral deposits.* They look like sediments but are not sediments in the truest sense of the term.

Stratabound deposits form when a hydrothermal solution invades and reacts with a muddy sediment. Reactions between sediment grains and the solution cause deposition of the ore minerals. Deposition commonly occurs before the sediment has become a sedimentary rock.

The famous copper deposits of Zambia, in central Africa, are stratabound ores, as are the great Kupferschiefer deposits of Germany and Poland. The world's largest and richest lead and zinc deposits, at Broken Hill and Mount Isa in Australia, and at Kimberley in British Columbia, are also stratabound ores.

Figure 17.8 Stratabound ore of lead and zinc from Kimberley, British Columbia. The layers of pyrite (yellow), sphalerite (brown), and galena (grey) are parallel to the layering of the sedimentary rock in which they occur. The specimen is 4 cm across.

Placers

The way minerals and rock particles become sorted by flowing water was mentioned in Chapter 7. Differences in density are especially effective ways to achieve sorting—more dense minerals remain while less dense minerals are washed away. Deposits of minerals with high densities *are placers.* The most important minerals concentrated in placers are gold, platinum, cassiterite (SnO_2), and diamond. Typical locations of placers are illustrated in Figure 17.9-

Gold is the most important mineral recovered from placers; more than half of the gold recovered throughout all of human history has come from placers. This is the result of the huge gold production from South Africa and Russia, almost all of which has come from placers.

The South African gold deposits are really fossil placers, and they have many unusual features. Most placers are found in steam gravels that are geologically young. The South African fossil placers are a series of gold-bearing conglomerates (Fig.17.10) that were laid down 2.7 billion years ago as gravels in the shallow marginal waters of a marine basin. Associated with the gold are grains of pyrite and uranium minerals. As far as size and richness are concerned, nothing like the deposits in the Witwatersrand basin has been discovered anywhere else. Nor has the original source of all the placer gold been discovered, and so it is not possible to say why so much of the world's minable gold should be concentrated in this one sedimentary basin.

Mining in the Witwatersrand basin has reached a depth of 3600 m (11,800 ft). This is the deepest mining in the world, and there are plans to continue mining to depths as great as 4500 m (14,800 ft). Despite such ambitious plans, the heyday of gold mining in South Africa has probably passed because the deposits are running out of ore.

Through the middle years of the 1980s, the price of gold fluctuated between about $14 and $16 a gram. This led to a boom in gold prospecting and to the discovery of a large number of new ore deposits in the United States, Canada, Australia, the Pacific islands, and elsewhere. Most of the new discoveries are hydrothermal deposits. Despite all the new discoveries, South Africa with its huge fossil placers continues to dominate the world's gold production. In 1993 South

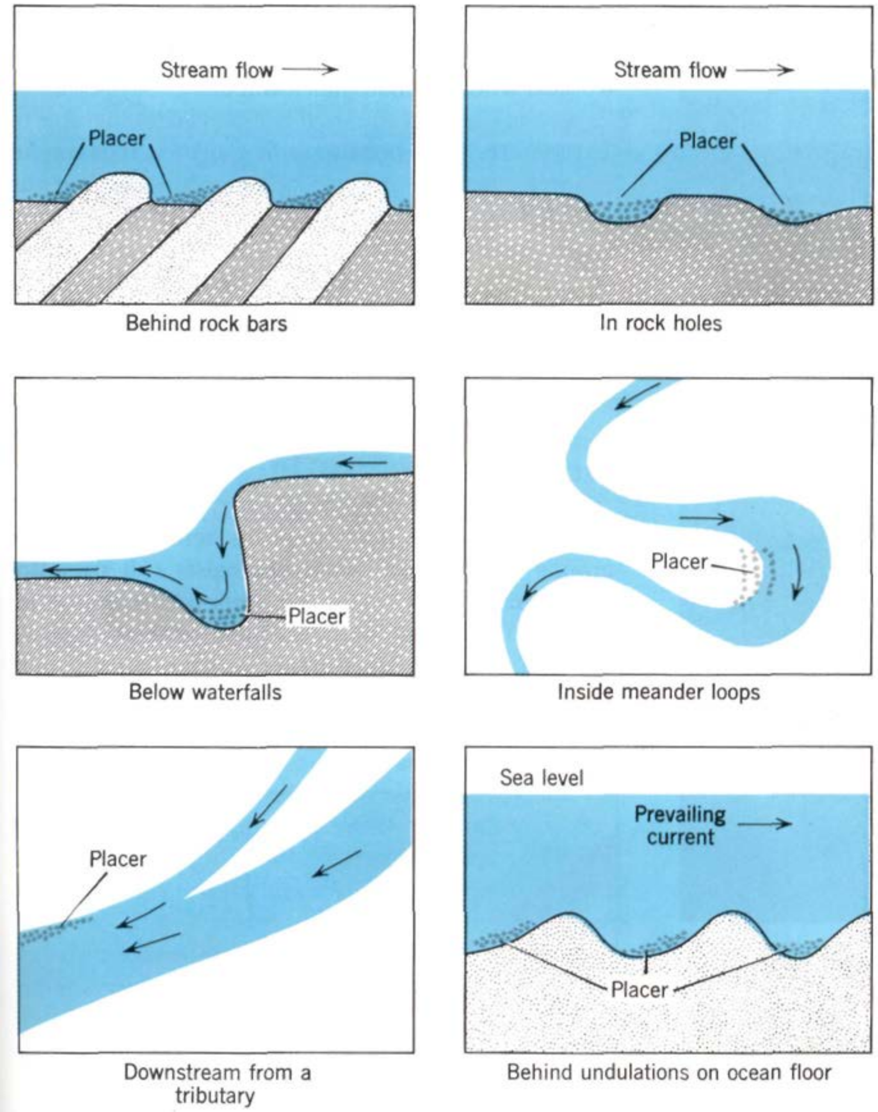


Figure 17.9 Placers occur where barriers allow flowing water to carry away the suspended load of lightweight (low-density) particles while trapping heavy (high-density) particles. Placers can form whenever water moves but are most commonly associated with streams of longshore currents.

Figure 17.10 Gold is recovered from fossil placers in the Witwatersrand basin, South Africa. The gold is found at the base of conglomerate layers interbedded with finer-grained sandstone, here seen in weathered outcrop at the site where gold was first discovered in 1786.

Africa supplied about 35 percent of all the gold produced in the world, but the production rate is dropping steadily.

Residual Mineral Deposits

Weathering occurs because newly exposed rock is not chemically stable when it is in contact with rainwater and the atmosphere. Chemical weathering, in particular, leads to mineral concentration through the removal of soluble materials in solution and the contraction of a less soluble residue.

Limonite is among the least soluble of the many minerals formed during chemical weathering. Under conditions of high rainfall in a warm, tropical climate, other minerals are slowly leached out of a soil, leaving an iron-rich limonitic crust called *laterite* at the sur-

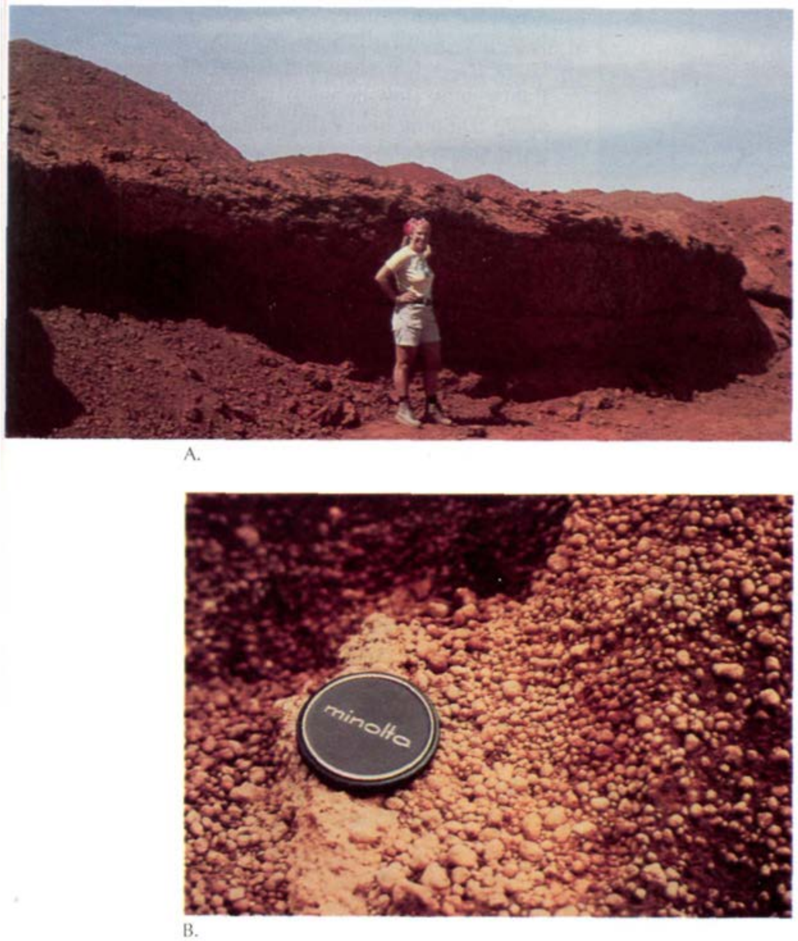


Figure 17.11 Residual mineral deposits rich in iron and aluminum are typically formed under tropical or scmilropical conditions. A. Red laterite enriched in iron, near Djenne, Mali. Laterites can sometimes be rich enough to be residual Iron ores. Such ores have been mined in the past, but no large mining activity of residual iron ore is occurring today. B. Bauxite from Weipa in Queensland, Australia. Long-continued leaching of clastic sedimentary rocks under tropical conditions has removed most of the original constituents, such as silica, calcium, and magnesium, leaving a rich bauxite consisting largely of the mineral gibbsite ($Al(OH)_3$). Nodules of gibbsite form by repeated solution and redeposition. The Weipa bauxite deposits are among the largest and richest in the world.

face (Fig. 17.11). In a few places, laterites can even be mined for iron.

Although iron-rich laterite is by far the most common kind of residual mineral deposit, the most important deposits as far as human exploitation are concerned are the aluminous laterites called *bauxites.* Bauxites are the source of the world's aluminum.

Bauxites are widespread, but they are concentrated in the tropics because that is where lateritic weathering occurs. Where bauxites are found in present-day temperate conditions, such as France, China, Hungary, and Arkansas, it is clear that the climate was tropical when the bauxites formed.

All bauxites and iron-rich laterites are vulnerable to erosion. They are not found in glaciated regions, for example, because overriding glaciers scrape off the soft surface materials. The vulnerability of bauxites and laterites means that most deposits are geologically young. More than 90 percent of all known deposits of bauxites, for example, formed during the last 60 million years, and all of the very large deposits formed less than 25 million years ago.

Metallogenic Provinces

Many kinds of mineral deposits tend to occur in groups and to form what geologists call **metallogenic provinces.** These are defined as limited regions of the crust within which mineral deposits occur in unusually large numbers. A striking example is the metallogenic province shown in Figure 17.12

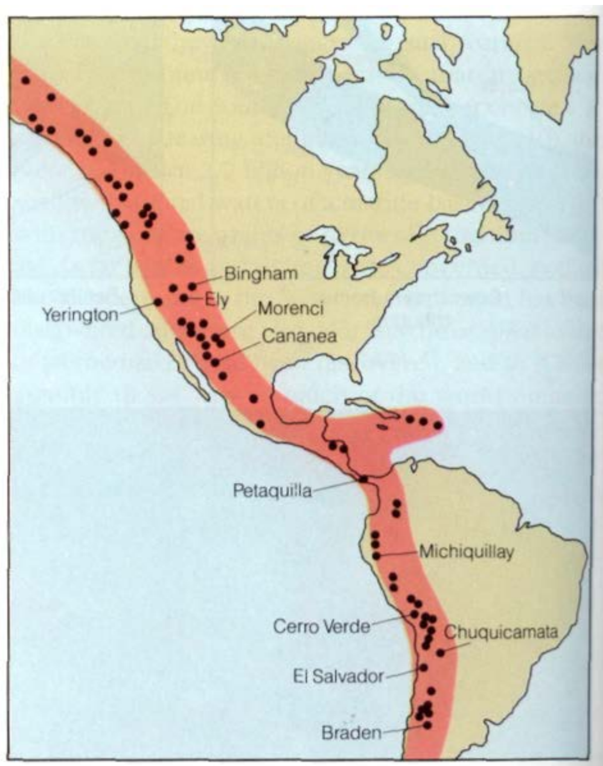


Figure 17.12 A metallogenic province of rich porphyry-copper deposits occurs along the western edge of the Americas. These chalcopyrite-rich deposits were formed by hydrothermal solutions generated by stratovolcanoes; the volcanoes formed above the subduction edges of the South and North American plates.

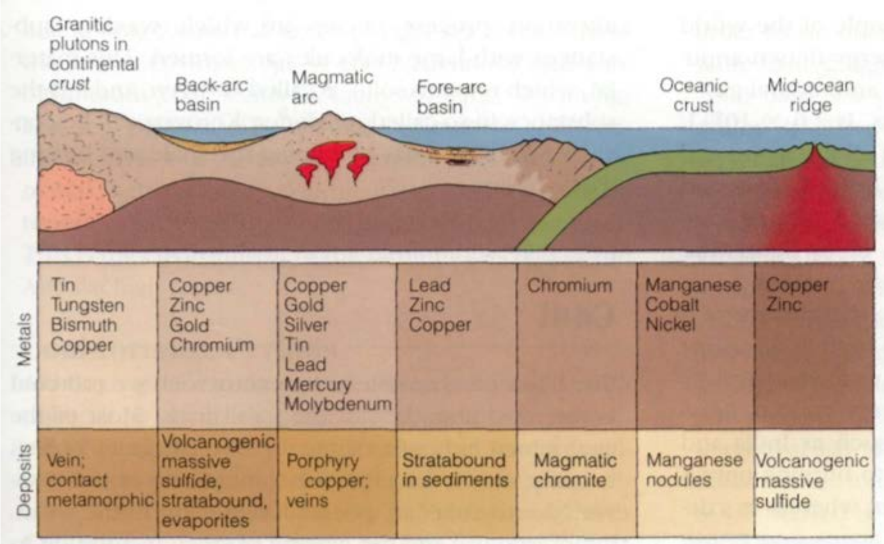


Figure 17.13 Locations of certain kinds of mineral deposits in terms of plate structures.

which runs along the western side of the Americas. Within the province is the world's greatest concentration of large hydrothermal copper deposits. These deposits are associated with intrusive igneous rocks that are invariably porphyritic (Chapter 5); they are therefore *called porphyry copper deposits.* The intrusive igneous rocks, and therefore the deposits themselves, were formed as a consequence of subduction because they are in, or adjacent to, old stratovolcanoes.

Metallogenic provinces form as a result of either climatic control (as in the formation of bauxite deposits in the tropics) or plate tectonics. Magmatic, hydrothermal, and stratabound deposits all form near present or past plate boundaries (Fig. 17.13). This is hardly surprising, for the deposits are related directly or indirectly to igneous activity, and most igneous activity we now know is related to plate tectonics.

ENERGY RESOURCES

A healthy, hard-working person can produce just enough muscle energy to keep a single 75-watt light bulb burning for 8 hours a day. It costs about 10 cents to purchase the same amount of energy from the local electrical utility. Viewed strictly as machines, humans aren't worth much. By comparison, the amount of mechanical and electrical energy used each 8-hour working day in North America could keep four hundred 75-watt bulbs burning for every person living there.

To see where all the energy is used, it is necessary to sum up all the energy employed to grow and transport food, make clothes, cut lumber for new homes, light streets, heat and cool office buildings, and do myriad other things. The uses can be grouped into three categories: transportation, home and commerce, and industry (meaning all manufacturing and raw material processing as well as the growing of foodstuffs). The present-day uses of energy in the United States are summarized in Figure 17.14.

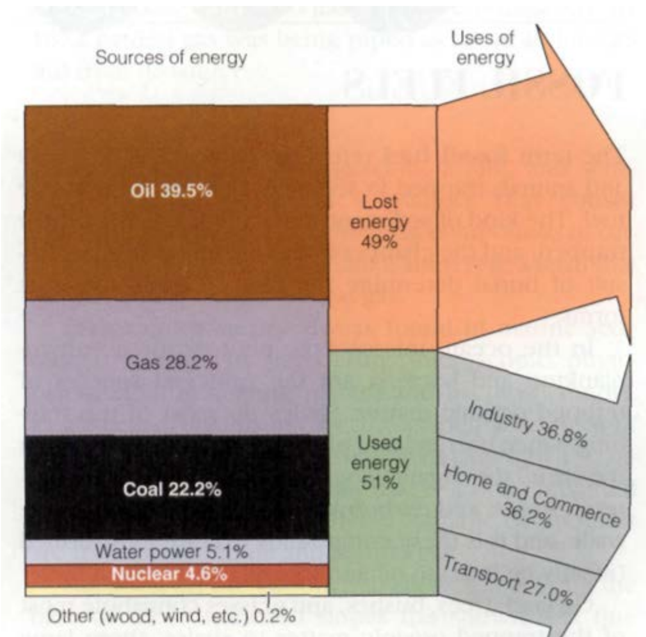


Figure 17.14 Uses and sources of energy in the United States. Lost energy arises both from inefficiencies of use and from the fact that the laws of thermodynamics impose a limit to the efficiency of any engine and therefore a limit on the fraction of available energy that can be usefully employed.

How much energy do all the people of the world use? The total is enormous. The energy drawn annually from the major fuels—coal, oil, and natural gas—plus that from nuclear power plants, is 2.6×10^{20} J. Nobody keeps accurate accounts of all the wood and animal dung burned in the cooking fires of Africa and Asia, but the amount has been estimated to be so large that when it is added to the 2.6×10^{20} J figure, the world's total energy consumption rises to about 3.0×10^{20} J annually. This is equivalent to the burning of 2 metric tons (440 lb) of coal or 10 barrels of oil for every living man, woman, and child each year! Energy consumption around the world is very uneven, however. In less developed countries such as India and Tanzania, energy use is equivalent to burning only 3 or 4 barrels of oil per person per year, whereas in a developed country such as the United States, energy use is equivalent to burning more than 50 barrels of oil per person per year.

FOSSIL FUELS

The term **fossil fuel** refers to the remains of plants and animals trapped in sediment that can be used for fuel. The kind of sediment, the kind of organic matter trapped, and the changes in the organic matter as a result of burial determine the kind of fossil fuel that forms.

In the ocean, microscopic photosynthetic phytoplankton and bacteria are the principal sources of trapped organic matter. Shales do most of the trapping. Once bacteria and phytoplankton are trapped in the shale, the organic compounds they contain—*proteins, lipids,* and *carbohydrates*—become part of the shale, and it is these compounds that are transformed (mainly by heat) to oil and natural gas.

On land, trees, bushes, and grasses contribute most of the trapped organic matter to shales; these large land plants are rich in resins, waxes, and lignins, which tend to remain solid and form coals rather than oil of natural gas.

In many marine and lake shales, burial temperatures never reach the levels at which the original organic molecules are converted to the organic molecules found in oil and natural gas. Instead, an alteration process occurs in which waxlike substances with large molecules are formed. This material, which remains solid, is called *kerogen,* and it is the substance in so-called *oil shales.* Kerogen can be converted to oil and gas by mining the shale and heating it in a retort.

Coal

The black combustible sedimentary rock we call **coal** is the most abundant of the fossil fuels. Most of the coal mined either is eventually burned under boilers to make steam for electrical generators, or it is converted into coke, an essential ingredient in the smelting of iron ore and the making of steel. In addition to its use as a fuel, coal is a raw material for nylon and many other plastics, as well as a multitude of other organic chemicals. The conditions under which organic matter accumulates in swamps as *peat,* then during burial and diagenesis is converted to coal, is discussed in Chapter 7. Coalification involves the loss of volatile materials such as H_2O, CO_2, and CH_4 (methane). As the volatiles escape, the remaining coal is increasingly enriched in carbon. Through coalification, peat is converted successively into *lignite* (one type of coal), *subbituminous coal,* and *bituminous coal* (Fig. 7.9). These coals are sedimentary rocks. However, *anthracite,* a still later phase in the coalification process, is a metamorphic rock.

Because of its low volatile content, anthracite is hard to ignite, but once alight it burns with almost no smoke. In contrast, lignite is rich in volatiles, burns smokily, and ignites so easily that it is dangerously subject to spontaneous ignition.

In regions where metamorphism has been intense, coal has been changed so thoroughly that it has been converted to graphite, in which all volatiles have been lost. Graphite will not burn in an ordinary fire.

Occurrence of Coal

A coal seam is a flat, lens-shaped body having the same surface area as the swamp in which it originally accumulated. Most coal seams tend to occur in groups. In western Pennsylvania, for example, 60 seams of bituminous coal are found. This clustering indicates that the coal must have formed in a slowly subsiding site of sedimentation.

Coal swamps seem to have formed in many sedimentary environments, of which two types predominate. One consists of slowly subsiding basins in continental interiors and the swampy margins of shallow inland seas formed at times of high sea level. This is the home environment of the bituminous and subbituminous coal seams in Utah, Montana, Wyoming, and the Dakotas. The second sedimentary environment consists of continental margins with wide continental shelves (that is, continental margins in plate interiors), that were flooded at times of high sea level. This is the environment of the bituminous coals of the Appalachian region.

Coal-forming Periods

Although peat can form under even subarctic conditions, it is clear that the luxuriant plant growth needed to form thick and extensive coal seams developed most readily under a tropical or semitropical climate. The Great Dismal Swamp in Virginia and North Carolina is one of the largest modern peat swamps. It contains an average thickness of 2 m (8 ft) of peat. However, unless the swamp lasts millions of years, even that dense growth is insufficient to produce a coal seam as thick as some of the seams in Pennsylvania.

Peat formation has been widespread and more or less continuous from the time land plants first appeared about 450 million years ago, during the Silurian Period. The size of peat swamps has varied greatly, however, and so, too, as a consequence, has the amount of coal formed. By far the greatest period of coal swamp formation occurred during the Carboniferous (hence its name) and Permian periods, when Pangaea existed. The great coal beds of Europe and the eastern United States formed at this time, when the plants of coal swamps were giant ferns and scale trees (gymnosperms). The second great period of coal deposition peaked during the Cretaceous period but commenced in the early Jurassic and continued until the mid-Tertiary. The plants of the coal swamps during this period were flowering plants (angiosperms), much like flowering plants today.

Petroleum: Oil and Natural Gas

Rock oil is one of the earliest resources our ancestors learned to use. However, the major use of oil really started in about 1847 when a merchant in Pittsburgh, Pennsylvania, started bottling and selling rock oil from natural seeps to be used as a lubricant. Five years later, in 1852, a Canadian chemist discovered that heating and distillation of rock oil yielded kerosene, a liquid that could be used in lamps. This discovery spelled doom for candles and whale-oil lamps. Wells were soon being dug by hand near Oil Springs, Ontario, in order to produce oil. In Romania in 1856, using the same hand-digging process, workers were producing 2000 barrels a year.[1] In 1859 the first oil well was drilled in Titusville, Pennsylvania. On August 27, 1859, at a depth of 21.2 m (70 ft), oil-bearing strata were encountered and up to 35 barrels of oil a day were pumped out. Oil was soon discovered in West Virginia (1860), Colorado (1862), Texas (1866), California (1875), and many other places.

The earliest known use of natural gas was about 3000 years ago in China, where gas seeping out of the ground was collected and transmitted through bamboo pipes to be used to evaporate saltwater in order to recover salt. It wasn't long before the Chinese were drilling wells to increase the flow of gas. Modern uses of gas started in the early seventeenth century in Europe, where gas made from wood and coal was used for illumination. Commercial gas companies were founded as early as 1812 in London and 1816 in Baltimore. The stage was set for the exploitation of an accidental discovery at Fredonia, New York, in 1821. A water well drilled in that year produced not only water, but also bubbles of a mysterious gas. The gas was accidentally ignited and produced such a spectacular flame that a new well was drilled on the same site; wooden pipes were installed to carry the gas to a nearby hotel, where 66 gas lights were installed. By 1872 natural gas was being piped as far as 40 km (25 mi) from its source.

Origin of Petroleum

Petroleum is defined as gaseous, liquid, and semisolid naturally occurring substances that consist chiefly of *hydrocarbons* (chemical compounds of carbon and hydrogen). Petroleum is therefore a term that includes both oil and natural gas.

Petroleum is nearly always found in marine sedimentary rocks. In the ocean, microscopic phytoplankton (tiny floating plants) and bacteria (simple, single-celled organisms) are the principal sources of organic matter trapped in sediment. Most of the organic matter is trapped in clay that is slowly converted to shale. During this conversion, organic compounds are transformed to oil and natural gas.

Sampling on the continental shelves and along the base of the continental slopes has shown that fine muds beneath the seafloor contain up to 8 percent organic matter. From such observations, geologists conclude that oil and gas originate primarily as organic matter trapped in sediment. Two additional kinds of

[1] A barrel is equal to 42 U.S. gallons and is the volume generally used when commercial production of oil is discussed.

evidence support the hypothesis that petroleum is a product of the decomposition of organic matter:

1. Oil possesses optical properties known only in hydrocarbons derived from organic matter.
2. Oil contains nitrogen and certain compounds believed to originate only in living matter.

A long and complex chain of reactions apparently is involved in the conversion of organic matter to petroleum. In addition, chemical changes may occur in oil and gas even after they have accumulated. This explains why chemical differences exist between the oil in one body of petroleum and another.

Once petroleum has formed in a shale, it is free to move. It is now well established that petroleum migrates through aquifers and can become trapped in reservoirs.

The migration of petroleum deserves further discussion. The sediment in which organic matter is accumulating today is rich in clay minerals, whereas most of the strata that constitute oil pools are sandstones (consisting of quartz grains), limestones and dolostones (consisting of carbonate minerals), and much-fractured rock of other kinds. Long ago, geologists realized that oil and gas form in one kind of material (shale) and at some later time migrate to another (sandstone or limestone).

Petroleum migration is analogous to groundwater migration. When oil and gas are squeezed out of the shale in which they originated and enter a body of sandstone or limestone somewhere above, they can migrate more easily than before because most sandstones and limestones are porous and therefore more permeable than any shale. The force of molecular attraction between oil and quartz or carbonate minerals is weaker than that between water and quartz or carbonate minerals. Hence, because oil and water do not mix, water remains fastened to the quartz or carbonate grains, while oil occupies the central parts of the larger openings in the porous sandstone or limestone. Because it is lighter than water, the oil tends to glide upward past the carbonate- and quartz-held water. In this way, it becomes segregated from the water; when it encounters a trap, it can form a pool.

Most of the petroleum that forms in sediments does not find a suitable trap, and eventually it makes its way, along with groundwater, to the surface. It is estimated that no more than 0.1 percent of all the organic matter originally buried in a sediment is eventually trapped in an oil pool. It is not surprising, therefore, that the highest ratio of oil and gas pools to volume of sediment is found in rock no older than 2.5 million years and that nearly 60 percent of all the oil and gas discovered so far has been found in strata of Cenozoic age (Fig. 17.15). This does not mean that older rocks produced less petroleum. It simply means that oil in older rocks has had a longer time in which to escape.

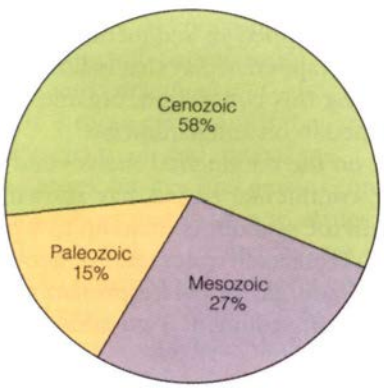


Figure 17.15 Percentage of world's total oil production from strata of different ages.

Distribution of Petroleum

Petroleum deposits, like coal, are frequent but are distributed unevenly. The reasons for the uneven distrib ution are not as obvious as they are with coal. Suitable source sediments for petroleum are very widespread and seem as likely to form in subarctic waters as in tropical regions. The critical controls seem to be a supply of heat to effect the conversion of solid organic matter to oil and gas, and the formation of a suitable trap before the petroleum has leaked away.

Conversion of solid organic matter to oil and gas happens within a specific range of depth and temperature defined by the geothermal gradients shown in Figure 17.16. If a thermal gradient is too low (less than 1.8°C/100 m, or 1°F/100 ft), conversion does not occur to either oil or gas. If the gradient is above 5.5°C/100 m, or 3°F/100 ft, conversion to gas starts at such shallow depths that very little trapping occurs. The depth-temperature window within which oil and gas form and are trapped lies between the two thermal gradients. Once oil and gas have been formed, they will accumulate in pools only if suitable traps are present. Most oil and gas pools are found beneath anticlines; the timing of the folding event that forms an anticline is therefore a critical part of the trapping process. If folding occurs after petroleum has formed and migrated, pools cannot form. The great oil pools in the Middle East arose through the fortunate coincidence of the right thermal gradient and the development of anticlinal traps during the collision of Europe and Asia with Africa.

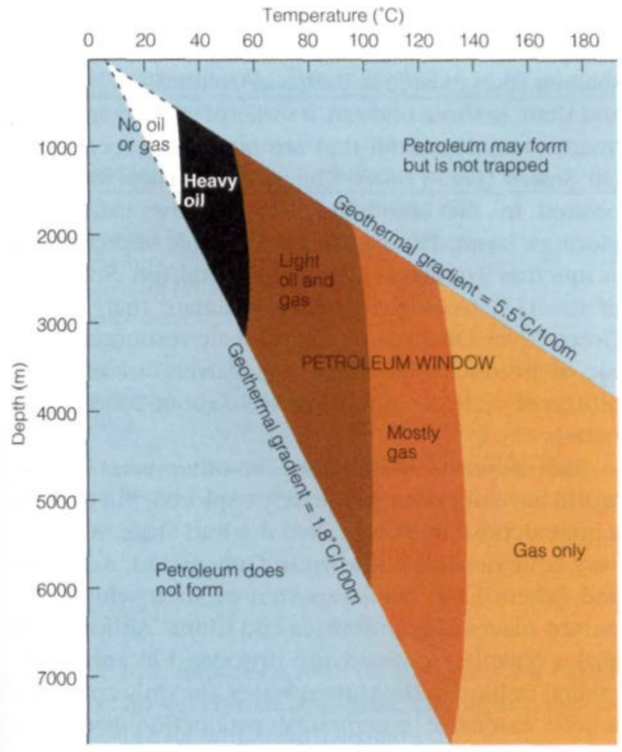


Figure 17.16 The petroleum window is that combination of depth and temperature within which oil and gas are generated and trapped.

How much oil is there in the world? This is an extremely controversial question. Approximately 600 billion barrels of oil have already been pumped out of the ground. A lot of additional oil has been located by drilling but is still waiting to be pumped out. Probably a great deal more oil remains to be found by drilling. The volume of strata in a basin of sediment can be accurately estimated for coal, but we can only guess at the volume of undiscovered oil. Guesses are based on the accumulated experience of a century of drilling. Knowing how much oil has been found in an intensively drilled area, such as eastern Texas, for example, experts make estimates of probable volumes in other regions where rock types and structures are similar to what we find in eastern Texas. Using this approach, and considering all the sedimentary basins of the world (Fig. 17.17), experts estimate that somewhere between 1500 and 3000 billion barrels of oil will eventually be discovered. (See the "Guest Essay" at the end of this chapter for more discussion about the search for oil.)

Tars

Oil that is exceedingly viscous and thick will not flow easily and cannot be pumped. Colloquially called **tar,** heavy, viscous oil acts as a cementing agent between mineral grains in an oil pool. The tar can be recovered

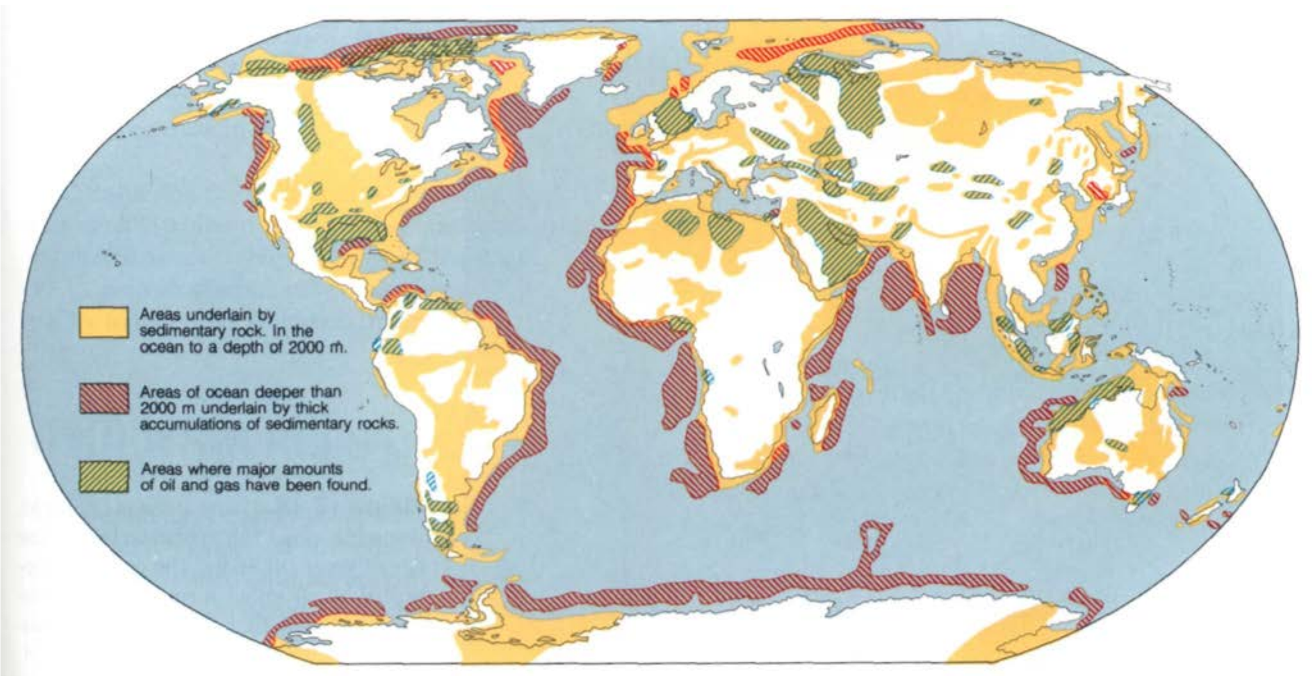


Figure 17.17 Areas underlain by sedimentary rock and regions where large accumulations of oil and gas have been located. Where the ocean is deeper than 2000 m, sedimentary rock has yet to be tested for its oil and gas potential.

only if the sandstone is mined and heated enough to make the tar flow. The resulting tar must then be processed to recover the valuable gasoline fraction. The cost of mining and treating "tar sands," as heavy, viscous oil deposits are called, is high, but it is technically possible and someday tar sands may be an important source of fuel. The largest known occurrence of tar sands is in Alberta, Canada, where the *Athabasca Tar Sand* covers an area of 5000 km^2 (1900 mi^2) and reaches a thickness of 60 m (200 ft). Similar deposits almost as large are known in Venezuela and in Russia.

Oil Shale

Another potential source of petroleum is kerogen in shale. If the kerogen is heated, it breaks down and forms liquid and gaseous hydrocarbons similar to those in oil and gas. All shales contain some kerogen, but to be considered an energy resource the kerogen must yield more energy than is required to mine and heat it. Only those shales that yield 40 or more liters of distillate per ton can be considered because the energy needed to mine and process a ton of shale is equivalent to that created by burning 40 liters (0.25 barrels) of oil.

The world's largest deposit of rich oil shale is in the United States. During the Eocene Epoch, many large, shallow lakes existed in basins in Colorado, Wyoming, and Utah; in three of them, a series of rich organic sediments was deposited that are now the Green River Oil Shales (Fig. 17.18). The richest shales were deposited in the lake in Colorado, now called the Piceance basin. These shales are capable of producing as much as 240 liters (1.5 barrels) of oil/ton. Scientists of the U.S. Geological Survey estimate that, in the Green River Oil Shale alone, oil-shale resources capable of producing 50 liters (0.3 barrels) or more of oil/ton of shale can ultimately yield about 2000 billion barrels of oil.

Rich deposits of oil shale in other parts of the world have not been adequately explored, but there is a huge deposit in Brazil called the Irati Shale. Another very large deposit is known in Queensland, Australia, and others have been reported in such widely dispersed places as South Africa and China. Although oil shales have been mined and processed in an experimental fashion in the United States, the only countries where extensive commercial production has been tried are Russia and China. Production expenses today make exploitation of oil shales in all countries unat-

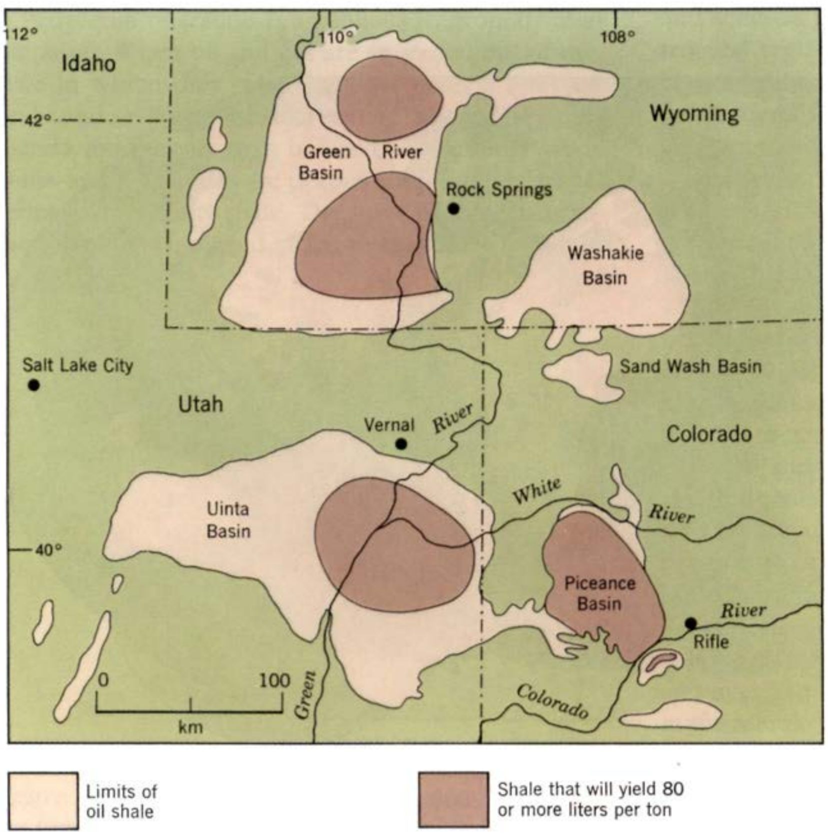


Figure 17.18 Vast areas of Colorado, Wyoming, and Utah are underlain by the Green River Oil Shale. The extensive deposits of oil shale formed as organic-rich sediment that accumulated in ancient freshwater lakes and was buried, compacted, and cemented. If heated, the solid organic matter in the shale is converted to hydrocarbons similar to those in petroleum.

Table 17.1
Amounts of Fossil Fuels Possibly Recoverable Worldwide
(Unit of Comparison is a Barrel of Oil)

Fossil Fuel	Total Amount in Ground (billions of barrels)	Amount Possibly Recoverable (billions of barrels)
Coal	About 100,000	62,730[a]
Oil and gas (flowing)	1500-3000	1500-3000
Trapped oil in pumped-out pools	1500-3000	0-?
Viscous oil (tar sands)	3000-6000	500-?
Oil shale	Total unknown; much greater than coal	1000-?

[a]0.22 ton of coal = 1 barrel of oil.

tractive by comparison with oil and gas. Most experts believe, however, that large-scale mining and processing of oil shale will eventually happen.

How Much Fossil Fuel?

Are supplies of fossil fuels adequate to meet future demand? If we use a barrel of oil as our unit of measurement, we can compare quantities of all fossil fuels directly. Approximately 0.22 ton of coal produces the same amount of heat energy as one barrel of oil. Thus, the world's recoverable coal reserves of $13{,}800 \times 10^9$ tons are equivalent to about 63,000 billion barrels of oil.

Considering the approximate world-use rate of barrels of oil (30 billion barrels a year) and comparing the estimated recoverable amounts of fossil fuels (Table 17.1), we can see that only coal seems to have the capacity to meet our long-term demands.

OTHER SOURCES OF ENERGY

Three sources of energy other than fossil fuels have already been developed to some extent: the Earth's plant life (so-called biomass energy), hydroelectric energy, and nuclear energy. Five others—the Sun's heat, winds, waves, tides, and the Earth's internal heat—have been tested and developed on a limited basis, but none has yet been developed on a large scale. The day may not be far off, however, when one or more of the five could become locally important.

Biomass Energy

Scientists working for the United Nations estimate that wood and animal dung used for cooking and heating fires now amounts to energy production of 4×10^{19} J annually. This is approximately 14 percent of the world's total energy use. The greatest use of wood as a fuel occurs in developing countries, where the cost of fossil fuel is very high in relation to income.

Measurements made on living plant matter indicate that new plant growth on land equals 1.5×10^{11} metric tons of dry plant matter each year. If all of this were burned, or used in some other way as a biomass energy source, it would produce almost nine times more energy than the world uses each year. Obviously, this is a ridiculous suggestion because in order to do so all the forests would have to be destroyed, plants could not be eaten, and agricultural soils would be devastated. Nevertheless, controlled harvesting of fuel plants could probably increase the fraction of the biomass now used for fuel without serious disruption to forests or to food supplies. In several parts of the world, such as Brazil, China, and the United States, experiments are already under way to develop this obvious energy source.

Hydroelectric Power

Hydroelectric power is recovered from the potential energy of stream water as it flows to the sea. As discussed in Chapter 9, in order to convert the power of flowing water into electricity efficiently, it is necessary to dam streams. Unfortunately, reservoirs behind dams fill with silt, and so even though water power is continuous, dams and reservoirs have limited lifetimes.

Water power has been used in small ways for thousands of years, but only in the twentieth century has it been used to any significant extent for generating electricity. All the water flowing in the streams of the world has a total recoverable energy estimated as 9.2×10^{19} J/yr, an amount equivalent to burning 15 billion barrels of oil per year. Thus, even if all the possible hydropower in the world were developed, we could satisfy only about one-third of the present world energy needs. We have to conclude that, for those fortunate countries with large rivers and suitable dam sites, hydropower is very important, but for most countries hydropower holds limited potential for development.

Nuclear Energy

Nuclear energy is the heat energy produced during controlled transformation of suitable radioactive isotopes (a process called **fission**). Three of the radioactive atoms that keep the Earth hot by spontaneous radioactive decay—^{238}U, ^{235}U, and ^{232}Th—can be mined and used in this way. Fission is accomplished by bombarding the radioactive atoms with neutrons, thus accelerating the rate of decay and the release of heat energy. The device in which this operation is carried out is called a **pile.**

When ^{235}U fissions, it not only releases heat and forms new elements but also ejects some neutrons from its nucleus. These neutrons can then be used to induce more ^{235}U atoms to fission, and a continuous chain reaction occurs. The function of a pile is to regulate the flux of neutrons so that the rate of fission can be controlled. When a chain reaction proceeds without control, an atomic explosion occurs. Controlled fission, therefore, is the method used by nuclear power plants, and a tremendous amount of energy can be obtained in the process. The fissioning of one gram of ^{235}U produces as much heat as the burning of 13.7 barrels of oil. Unfortunately, however, ^{235}U is the only natural radioactive isotope that will maintain a chain reaction, and it is the least abundant of the three radioactive isotopes that are mined for nuclear energy. Only one atom of each 138.8 atoms of uranium in nature is ^{235}U. The remaining atoms are ^{238}U, which will not sustain a chain reaction. However, if ^{238}U is placed in a pile with ^{235}U that is undergoing a chain reaction, some of the neutrons will bombard the ^{238}U and convert it to plutonium-239 (^{239}Pu). This new isotope can, under suitable conditions, sustain a chain reaction of its own. The pile in which the conversion of ^{238}U takes place is called a **breeder reactor.** The same kind of device can be used to convert ^{232}Th into ^{233}U, which also will sustain a chain reaction. Unfortunately, breeder reactors and nuclear power plants based on them are more complex and less safe than ^{235}U plants, so all the present nuclear power plants use ^{235}U.

Already there are several piles in nuclear power plants operating around the world. They utilize the heat energy from fission to produce steam that drives turbines and generates electricity. Approximately 8 percent of the world's electrical power is derived from nuclear power plants. In France, more than half of all the electrical power comes from nuclear plants; the fraction is rising sharply in some other European countries and Japan, too. The reason for the increase is obvious. Japan and most European countries do not have adequate supplies of fossil fuels in order to be self-sufficient.

Many problems are associated with nuclear energy. The isotopes used in power plants are the same isotopes used in atomic weapons, so a security problem exists. The possibility of a power plant failing in some unexpected way creates a safety problem. The dreadful Chernobyl disaster in 1986 in the Ukraine is an example of such an event. Finally, the problem of safe burial of dangerous radioactive waste matter must be faced. Some of the waste matter will retain dangerous levels of radioactivity for thousands of years.

Geothermal Power

Geothermal power, as the Earth's internal heat flux is called, has been used for more than 50 years in New Zealand, Italy, and Iceland, and more recently in other parts of the world, including the United States. The source of heat in most instances is a body of magma. The magma heats groundwater, and the steam that is formed can be used to drive turbines and make electricity. How this is done is illustrated in Figure 17.19.

Most of the world's geothermal stream reservoirs are close to plate margins because plate margins are where most recent volcanic activity has occurred. A depth of 3 km (1.9 mi) seems to be the rough lower limit for big geothermal steam and hot-water pools. It is estimated that the world's geothermal reservoirs can yield about 8×10^{19} J—equivalent to burning 13 billion barrels of oil. This estimate incorporates the observation that, in New Zealand and Italy, only about 1 percent of the energy in a geothermal reservoir is recoverable. If the recovery efficiency were to rise, the estimate of recoverable geothermal resources would also rise. But even if the efficiency rose to 50 percent, geothermal power, like hydropower, could only satisfy a small part of human energy needs. For this reason, a good deal of attention is being given to creating artificial geothermal steam fields. So far experiments have only been partially successful.

Energy from Winds, Waves, Tides, and Sunlight

The most obvious source of energy is the Sun. The amount of energy reaching the Earth each year from the Sun is approximately 4×10^{24}J—that is, ten thousand times more than we humans use. We already put some of the Sun's energy to work in greenhouses and in solar homes, but the amount so used is tiny. The major challenge is to convert solar energy directly to electricity. Devices that effect such a conversion, called photovoltaic devices, have been invented. So far, their costs are too high and their efficiencies too low for most uses, although they are already widely used in small calculators, radios, and other devices that use very little power.

Winds and waves are both secondary expressions of solar energy. As discussed in Chapter 13, winds, in particular, have been used as an energy source for thousands of years through sails on ships and windmills. Today, huge farms of windmills are constructed in suitably windy places. Although problems and high costs are associated with windmills, it seems very likely that by the year 2000 or sooner, they will be cost-competitive with coal-burning electrical power plants. Unfortunately, much of the wind energy is in high-altitude winds. Steady surface winds only have about 10 percent of the energy the human race now uses. As with hydro- and geothermal power, therefore, wind power may become locally significant but will probably not be globally important.

Waves, which arise from winds blowing over the ocean, contain an enormous amount of energy. We can see how powerful waves are along any coastline during a storm. Wave power has been used to ring bells and blow whistles as navigational aids for centuries, but so far no one has discovered how to tap wave energy on a large scale. Devices that have been designed to do this tend to fail because of corrosion or storm damage.

Tides arise from the gravitational forces exerted on the Earth by the Moon and the Sun. As discussed in Chapter 8, if a dam is put across the mouth of a bay so that water can be trapped at high tide, the outward flowing water at low tide can drive a turbine. Unfortunately, the efficiency of the process is low, and few places around the world have tides high enough to make tidal energy feasible.

It is clear that there are numerous sources of energy and that far more energy is available than we can use. What is not yet clear is when, or even whether, we will be clever enough to learn how to tap the different energy sources in nonhazardous ways that don't disrupt the environment.

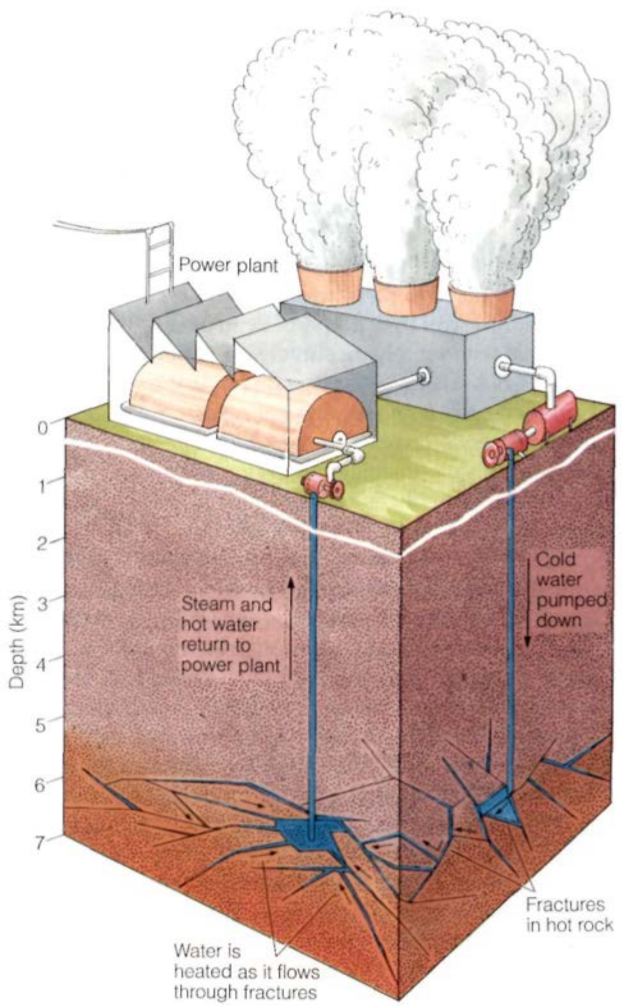


Figure 17.19 Geothermal energy. Water in fractures in hot rock forms steam that is brought to the surface and used to run a power plant. After use, waste water Irom condensed steam is pumped back underground again.

Guest Essay

The Pursuit of Undiscovered Fuel

Pinar Oya Yilmaz holds a Ph.D. in structural geology from the University of Texas at Austin. At present she is a Research Specialist and Technical Area Network Coordinator at the Exxon Production Research Company in Houston, Texas. She has served on a number of professional committees, and currently is Vice-President of the Geological Society of America's International Division.

As fossil fuel exploration opportunities diminish, scientists must put more emphasis on developing existing oil and gas resources. Such development, in turn, relies on improving techniques to locate these resources. While technical advancement has progressed using traditional geologic approaches, computer-aided analysis and modeling also have become increasingly important.

Several advances based on traditional geological studies help in the search for oil and gas. Plate tectonic reconstructions of regions are one of the advances: they are used to create a framework of paleogeography and paleoclimatology at the time sediments were being deposited in basins. This reconstruction tells earth scientists whether it is likely that petroleum has formed in a particular area. In addition, advances in organic geochemistry can now be used to evaluate the quality of source rock, its oil or gas yield, and its state of maturation. Some of the most important advances have come in our understanding of how sediments are deposited and how properties such as grain size and shape, cement, and porosity vary from place to place in a stratum. Using such information, reservoir rock properties can be evaluated for oil potential without relying entirely on drilling.

One of the areas where technological breakthroughs have occurred is in geophysical research. Geophysical research encompasses three different areas: petrophysics, seismic imaging, and borehole geophysics. All of these areas concentrate on directly recognizing the presence of gas and oil in the subsurface.

Petrophysics seeks to improve scientists' ability to evaluate sediments that could hold oil. This estimation involves analysis of sediment geometry, continuity, lithology, and fluid content through the combined use of surface seismic data and well logs. These logs involve lowering electronic instruments into a well borehole to determine resistivity, density, and porosity of sediments. Such electrical tools allow earth scientists to distinguish different sediments and types of fluids around the well bore. Advances in well log technologies emphasize new interpretation methods for rocks, such as shaly sands. In addition, they concentrate on methods which allow determination of type of fluid in reservoirs after a well has been drilled and has produced oil for some time. The well bore is also used for gathering seismic data closer to sediments of interest to locate structural and stratigraphic complexities.

Advances in seismic imaging provide explorationists with new tools to image the earth's subsurface. The goal of seismic imaging techniques is to create an accurate computer-generated "picture" of the subsurface in three dimensions to enable earth scientists to make the right decision on where to drill for oil. A scientist determines the area of interest and plans horizons which will be targeted for three dimensional (3-D) seismic imaging surveys. After the data is acquired and processed, the interpretation phase begins. High-powered computer workstations running GIS software (the same type of software used for remote sensing analysis discussed in the Chapter 15 Guest Essay) are used for complete geological and geophysical integration, including data from the surface and data provided from subsurface wells. Since seismic data lie in the time domain but geology must be interpreted in depth, well control is extremely important in understanding seismic data results. The interpreter has to be certain that proper depth conversions are used. Many dry holes have been drilled over seismic "highs" that were not checked for absolute depths.

Borehole geophysics integrates petrophysics and seismic analysis to better image the subsurface in structurally or stratigraphically complex areas, where there are abrupt vertical or lateral changes in sedimentary units. Borehole geophysics technologies place seismic tools downhole, at depths closer to the sediments of interest, in order to enhance resolution and to make it easier to image vertical features. Subsurface features are imaged around the well bore in the same manner as other seismic surveys. The images collected can be used to identify faults, locate salt-sediment interfaces, and characterize reservoirs according to rock type, porosity, and fluid content.

As the pursuit of undiscovered fuel intensifies, earth scientists must search for innovative and advanced methods to locate oil and gas. New tools and techniques enhance data-interpreters' predictive ability when searching for undiscovered oil. All of these advanced tools and techniques help earth scientists decide which parts of sedimentary basins housing oil fields show the greatest promise for new discoveries and the highest return on investment.

Summary

1. When a mineral deposit can be worked profitably, it is called an ore. The waste material mixed with ore minerals is gangue.
2. Mineral deposits form when minerals become concentrated in one of five ways: (1) precipitation from hydrothermal solutions to form hydrothermal mineral deposits; (2) concentration through crystallization to form magmatic mineral deposits; (3) concentration from lake water or seawater to form sedimentary mineral deposits; (4) concentration in flowing water to form placers; and (5) concentration through weathering to form residual deposits.
3. Hydrothermal solutions are brines, and they can be given off by cooling magma or else form when either groundwater or seawater penetrates the crust, becomes heated, and reacts with the enclosing rocks.
4. Hydrothermal mineral deposits form when hydrothermal solutions deposit dissolved minerals because of cooling, boiling, pressure drop, mixing with cold groundwater or seawater, or through chemical reactions with enclosing rocks.
5. Chromite, the main ore mineral of chromium, is the most important mineral concentrated by fractional crystallization.
6. Sedimentary mineral deposits are varied. The largest and most important are evaporites. Marine evaporite deposits supply most of the world's gypsum, halite, and potassium minerals.
7. Gold, platinum, cassiterite, diamonds, and other minerals are commonly found mechanically concentrated in placers.
8. Bauxite, the main ore of aluminum, is the most important kind of residual mineral deposit. Bauxite forms as a result of tropical weathering.
9. The distribution of many kinds of mineral deposits is controlled by plate tectonics because most magmas and most sedimentary basins are where they are because of plate tectonics.
10. Nonmetallic substances are used mainly as chemicals, fertilizers, building materials, and ceramics and abrasives.
11. Coal originated as plant matter in ancient swamps and is both abundant and widely distributed.
12. Oil and gas originated as organic matter trapped in shales and decomposed chemically owing to heat and pressure following burial. Later, these fluids moved through reservoir rocks and were caught in geologic traps to form pools.
13. When heated, part of the solid organic matter found in shale—called kerogen—will convert to oil and gas. Oil from shales is the world's largest resource of fossil fuel. Unfortunately, most shale contains so little kerogen that more oil must be burned to heat the shale than is produced by the conversion process.
14. Nuclear energy is derived from atomic nuclei of radioactive isotopes, chiefly uranium. The nuclear energy available from naturally occurring radioactive elements is the single largest energy resource now available.
15. Other sources of energy currently used to some extent are geothermal heat, energy from flowing streams, winds, waves, tides, and the Sun's heat.

Important Terms to Remember

breeder reactor (p. 464)
coal (p. 458)
fission (p. 464)
fossil fuel (p. 458)
gangue (p. 448)
hydrothermal mineral deposit (p. 448)
magmatic mineral deposit (p. 448)
metallogenic province (p. 456)
mineral deposit (p. 446)
ore (p. 447)
petroleum (p. 459)
pile (p. *464)*
placer (p. 448)
residual mineral deposit (p. 448)
sedimentary mineral deposit (p. 448)
tar (p. 461)

Questions for Review

1. Describe the difference between renewable and nonrenewable resources. Name three things that you use daily that rely on renewable resources and three that rely on nonrenewable resources.
2. What are mineral deposits? Describe five ways by which a mineral deposit can form.
3. If there are any mineral deposits in the area where you live or study, what kind of deposits are they and how do such deposits form?
4. What factors determine whether or not a mineral deposit is ore?
5. How do hydrothermal solutions form and how do they form mineral deposits?
6. Briefly describe the formation of three different kinds of sedimentary mineral deposits.
7. What factors control the concentration of minerals in placers? Name four minerals mined from placers.
8. How do residual mineral deposits form? What are the principal resources concentrated in residual deposits? If you were prospecting for such deposits, in what parts of the world would you concentrate your search?
9. What is a fossil fuel? Name four different kinds of fossil fuel.
10. Explain the steps that occur as organic matter becomes coal.
11. During what two periods in the Earth's history was most coal formed? Explain why coal formed when and where it did, and why coal is now found where it is.
12. What kind of rocks serve as source rocks for petroleum? In what kinds of rocks does petroleum tend to be trapped? Why?
13. What is the source of the organic matter that forms petroleum? What observations lead geologists to conclude that organic matter really is the source of petroleum?
14. Oil drillers find more petroleum per unit volume of rock in Cenozoic rocks than in Paleozoic rocks of the same kind. Explain.
15. Oil shales are rich in organic matter. Explain why such shales have not served as source rock for petroleum. How can oil shales be used as an energy resource?
16. Discuss the relative amounts of energy available from the different fossil fuels. What is your opinion about how fossil fuels will be used in the future?
17. What is nuclear energy? How is it used to make electricity and what possible dangers are there in developing nuclear energy?
18. What limitations are there to the development of hydroelectric power? Of wave and wind power?
19. What is geothermal power and where is it found? Compare the magnitude of available geothermal power with the magnitude of petroleum resources.
20. Would it be possible to increase our use of biomass as an energy resource? What are the limitations to use of biomass energy?

Questions for A Closer Look

1. Under what geologic circumstances is it possible to sample and test modern ore-forming solutions?
2. If you were in command of a research program, where in the world would you send it to find presently undiscovered hydrothermal solutions?
3. If mineral deposits of the kind now forming in the Red Sea also formed in the Atlantic when it was a long, narrow sea, where would those ore deposits be today?

Questions for Discussion

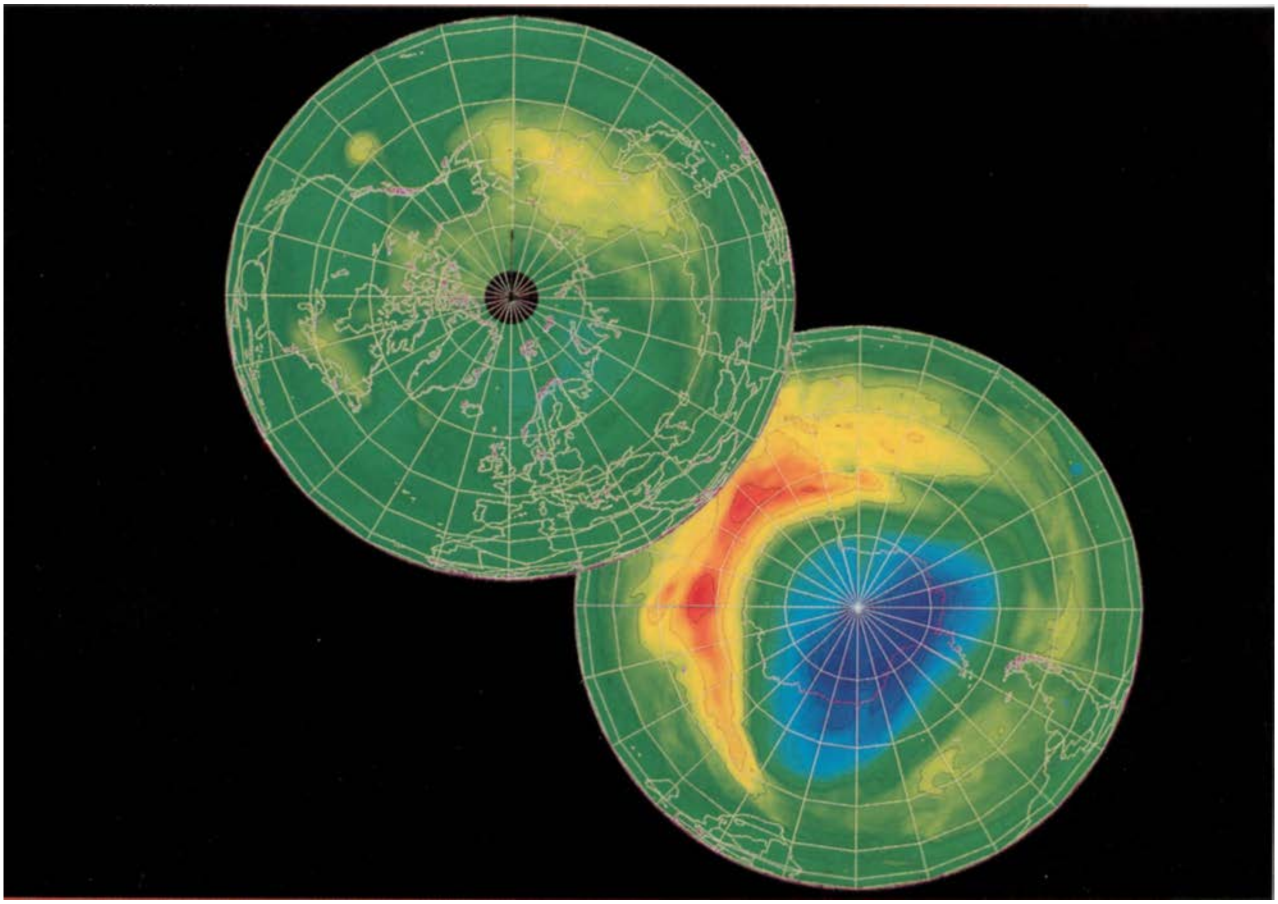
1. Taking into consideration the world's population (5.5 billion in 1994) and the fact that it is growing by approximately 100 million a year, how do you think the world's energy needs will be met over the next century? If you decide that fossil fuel is to be the energy source of choice, what consequence might that have for the atmosphere?
2. Discuss the role that recycling might play in making available metallic resources in the future.

Can you imagine a scenario in which the mining of new metallic minerals might cease? Would your recycling arguments apply equally to metallic and nonmetallic minerals?

3. Considering how mineral deposits are known to form on the Earth, which planets or moons in the solar system would you recommend as places to prospect for mineral resources? Be sure to specify what kind of resources you would expect to be found.

CHAPTER 18

Global Change: A Planet Under Stress



Total atmospheric ozone over the southern and northern hemispheres, on September 30, 1992, Reddish and yellowish colors indicate high ozone concentrations, whereas bluish colors denote abnormally low concentrations that define the "ozone hole."

Destroying the Ozone Shield

Ozone, a pale blue gas with a pungent odor, is present in the atmosphere in very small amounts, only 20 to 40 parts per billion by volume near the land surface. Without this gas, however, life on the Earth would be very different, for ozone provides living organisms with a protective shield against harmful ultraviolet radiation from the Sun. For humans, direct exposure to ultraviolet light damages the immune system, produces cataracts, substantially increases the frequency of skin cancer, and causes genetic mutations.

Maximum concentrations of ozone are found in a layer between 25 and 35 km (15.5 and 22 mi) above the Earth's surface, a region where ultraviolet radiation breaks down molecules of oxygen (O_2) into two oxygen atoms, which are then able to combine with other O_2 molecules to form molecules of ozone (O_3). The ozone is in turn broken down by ultraviolet radiation, thereby creating a balance among O, O_2, and O_3.

In 1985 British scientists working in Antarctica reported a startling discovery: a vast hole, about the size of Canada, had developed in the ozone layer above the Antarctic region. By 1987 measurements showed that, over Antarctica, concentrations of this life-protecting gas had dropped more than 50 percent since 1979 and, between altitudes of 15 and 20 km (9 and 12 mi), the depletion had reached 95 percent. Record low values were subsequently measured over Australia and New Zealand, and continuing surveys showed that ozone values at all latitudes south of 60° had decreased by 5 percent more since 1979-

What had happened to upset the natural atmospheric balance among the three gaseous forms of oxygen? A decade before the ozone hole was discovered, it was recognized that a group of synthetic industrial gases, the *chlorofluorocarbons (CFCs),* were entering the lower atmosphere and spreading rapidly around the world. As the CFCs ultimately rise into the upper atmosphere, ultraviolet radiation breaks them down, releasing chlorine. It is chlorine, in the form of chlorine monoxide (CIO), that does the damage: the chlorine atoms destroy the ozone, with each chlorine being capable of destroying as many as 100,000 ozone molecules before other chemical reactions remove the chlorine from the atmosphere. The sunlight and very cold springtime temperatures [—80°C (—176°F) or lower] in the upper atmosphere that are critical to ozone destruction are present in the south polar region, which is why the ozone hole is especially pronounced over Antarctica. In the Arctic, the period of the critical spring conditions is much shorter. With the documentation of ozone depletion in the upper atmosphere, scientists for the first time could show that human activity was having a detrimental global effect on one of the Earth's natural systems.

Continuing measurements have shown that atmospheric ozone concentration is decreasing at all latitudes outside the tropics and that in the lower troposphere the rate of decrease is about 10 percent per decade. From ground-based and satellite observations, we know that ozone concentration during February 1993 was as much as 20 percent below normal over much of the northern hemisphere. Although the rate of increase of atmospheric CFC gas concentration has fallen markedly during the last five years with a cut in worldwide production, even if all production stops, recovery to natural conditions is likely to take a century or more.

THE CHANGING EARTH

Since earliest times, people have probably asked important questions about the Earth: How did the major features of our planet originate? Has the land always been the way we see it now? How can we explain the diversity of life on the Earth? What is the place of *Homo sapiens* in the multitude of living things that inhabit our planet? Have the Earth's climates always been as we find them today?

Study of the Earth has made it abundantly clear that our planet is a complex, dynamic system that is in a constant state of change. Furthermore, the solid, liquid, gaseous, and organic realms of the Earth are closely interlinked. A change in one part of the system is likely to affect other parts. For example, a massive earthquake might raise an extensive zone of coastal land, exposing and destroying nearshore marine habitats. A volcano might erupt lava that dams a river, thereby affecting other streams in the drainage system, while tephra and gases ejected into the atmosphere could lead to a hemisphere-wide drop in air temperatures. We can observe and measure such natural changes in progress or interpret them from the geologic record. However, only recently have we come to recognize that humans can also have a profound affect on the four major components of the Earth system—air, oceans, land, and biota.

Time Scales of Change

Realization that the Earth is a dynamic planet and is always changing has come slowly. While it is apparent to anyone that the weather can change from year to year, it is less obvious that climate can change over a person's lifetime. It is far more difficult to comprehend from personal experience that the solid Earth is also changing. A person can witness a volcanic eruption or experience a major earthquake, yet not guess that the ground underfoot is slowly and continuously moving, driven by forces deep beneath the surface. We now can measure the rates at which lithospheric plates move and therefore know that plate motion generally ranges of 1 to 12 cm/yr (0.4 to 4.7 in/yr). This means that over the lifetime of a 70-year-old person the plate on which that person has lived traveled less than 10 m (33 ft), a displacement far too small to be recognized without sophisticated scientific measurement.

Recognizable changes that affect the natural equilibrium of the Earth system generally take place very slowly—over intervals far longer than an individual's lifetime. However, there are several exceptions to this generalization. One exception involves the impact of a large comet or meteoroid like the one believed to have struck the Earth about *66* million years ago at the end of the Cretaceous Period (see Introduction). Fortunately, such events are rare, and no major impact having global consequences has occurred in human history. A second exception is a major explosive volcanic eruption, which ejects so much gas and dust into the stratosphere that much of the incoming solar radiation is reflected back into space, thereby cooling the Earth's surface for several years (Chapter 14).

Changes we cause ourselves are a third exception. We next examine how human activities have brought about unprecedented global changes that will present us with major challenges in coming decades.

Our Planet's Growing Population

We have little direct knowledge about the size of the human population until relatively recent times. The first attempts to estimate global population were made near the end of the seventeenth century, when world population was probably no more than about 700 million (Fig. 18.1). National censuses helped improve estimates during the nineteenth century, but only after the Second World War did estimates of global population become relatively reliable. However, the long-term trend is clear: whereas it took countless millennia for the human species to number a billion (toward the beginning of the nineteenth century), that number has doubled twice since then and is now approaching 6 billion. The sheer growth in population has led to an ever-increasing interaction between people and their environment, so that now even the most remote parts of the planet are being affected.

Over many millennia prior to the Industrial Revolution, people slowly changed the Earth's natural landscapes as they built villages and cities, converted forests to agricultural land, and locally dammed and

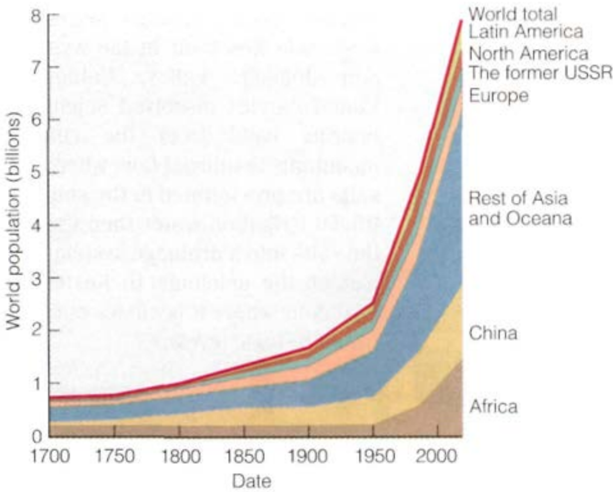


Figure 18.1 Cumulative world population between 1700 and the present obtained by summing values for each of the major inhabited regions of the world. Projected values extend the present estimates to the year 2020.

diverted streams. Then, with the development of industrial technology, mineral and energy resources were needed to fuel an increasingly populous and demanding society. The exploitation of fossil fuels helped raise the standard of living for most people well beyond that of their forebears of only a century or two ago.

In spite of the obvious benefits involved, the exploitation of our planet's rich natural resources has not been without cost. In many parts of the world, environmental deterioration is epidemic. In addition to scarring and poisoning the Earth's land surface, we have also, unwittingly, polluted the oceans and groundwater and changed the composition of the atmosphere. Even in places long considered to be the most remote on the planet—the frigid ice sheets of Antarctica, the vast Amazon rain forest, the trackless desert of Saudi Arabia, the lofty summits of the Himalaya—the impact of human activities is being felt. Today, human and natural geologic activities are inextricably intertwined, and it is increasingly apparent that people have become a major factor—a global factor—in environmental change.

Human-Induced Environmental Changes

In Chapter 9 we saw how construction of the Aswan Dam halted the natural supply of sediment to the Nile delta, depriving the rich agricultural delta lands of their annual input of nutrients and leaving the delta front vulnerable to wave erosion. Three additional examples—from California, Africa, and South America—further illustrate how people can upset the Earth's natural balance, sometimes causing unforseen responses.

Toxic Groundwater in California

Selenium is a naturally occurring element and a necessary trace nutrient in our diet as well as in the diet of livestock and many wild animals. However, if excessive amounts are ingested, it can prove toxic. In 1983, selenium attracted national attention when the U.S. Fish and Wildlife Service reported fish kills and high incidences of mortality, birth defects, and decreased hatching rates in nesting waterfowl at the Kesterson National Wildlife Refuge in California's San Joaquin Valley (Fig. 18.2). Laboratory studies showed high concentrations of selenium in fish from Kesterson Reservoir, located in the refuge, while birds using the reservoir were found to have high concentrations of selenium and obvious symptoms of selenium poisoning.

Geologists subsequently showed that the poisoning of wildlife at Kesterson resulted from a combination of geological, hydrological, and agricultural factors. The western San Joaquin Valley is a prime agricultural area, but because it lies in a zone of arid climate, the land is irrigated. The irrigation artificially raised the water table to such a degree that a system of subsurface drains was established to remove excess water and funnel it northward along the valley to Kesterson Reservoir.

Rainwater falling in the Coast Range immediately west of the San Joaquin Valley dissolves selenium-bearing salts from sedimentary rocks. Surface runoff then carries the dissolved salts to broad alluvial fans in the valley, where the water seeps into the ground and recharges a shallow regional aquifer. High evaporation in this arid region concentrates the salts in the soil. An irrigation system on the fan surfaces flushes salts out of the soil and into the drainage canal that leads to Kesterson Reservoir. Because the reservoir has no outlet, selenium became concentrated there until it reached toxic levels.

Although we are constantly alerted to the environmental impact of pesticides and other manufactured poisons introduced into natural ecosystems, the Kesterson saga illustrates how natural substances that pose no special hazard under normal conditions can reach toxic levels through human intervention. Such problems are increasing in number as an expanding human population places ever greater demands on limited natural resources.

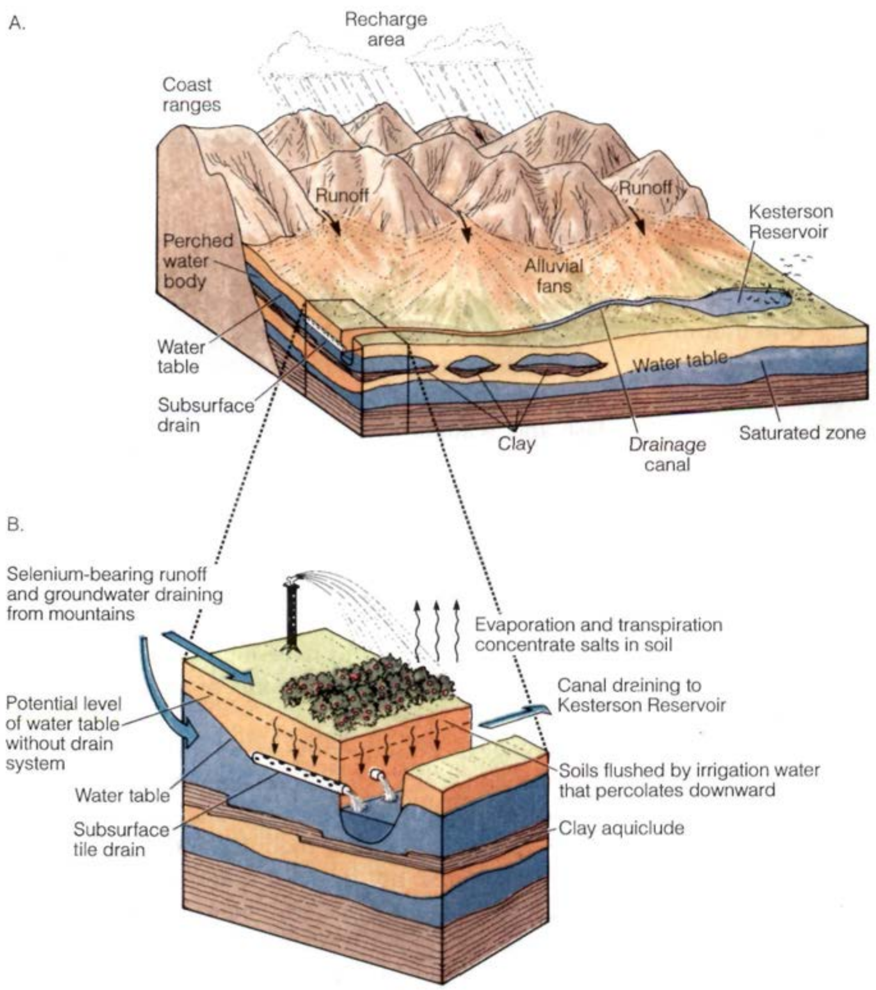


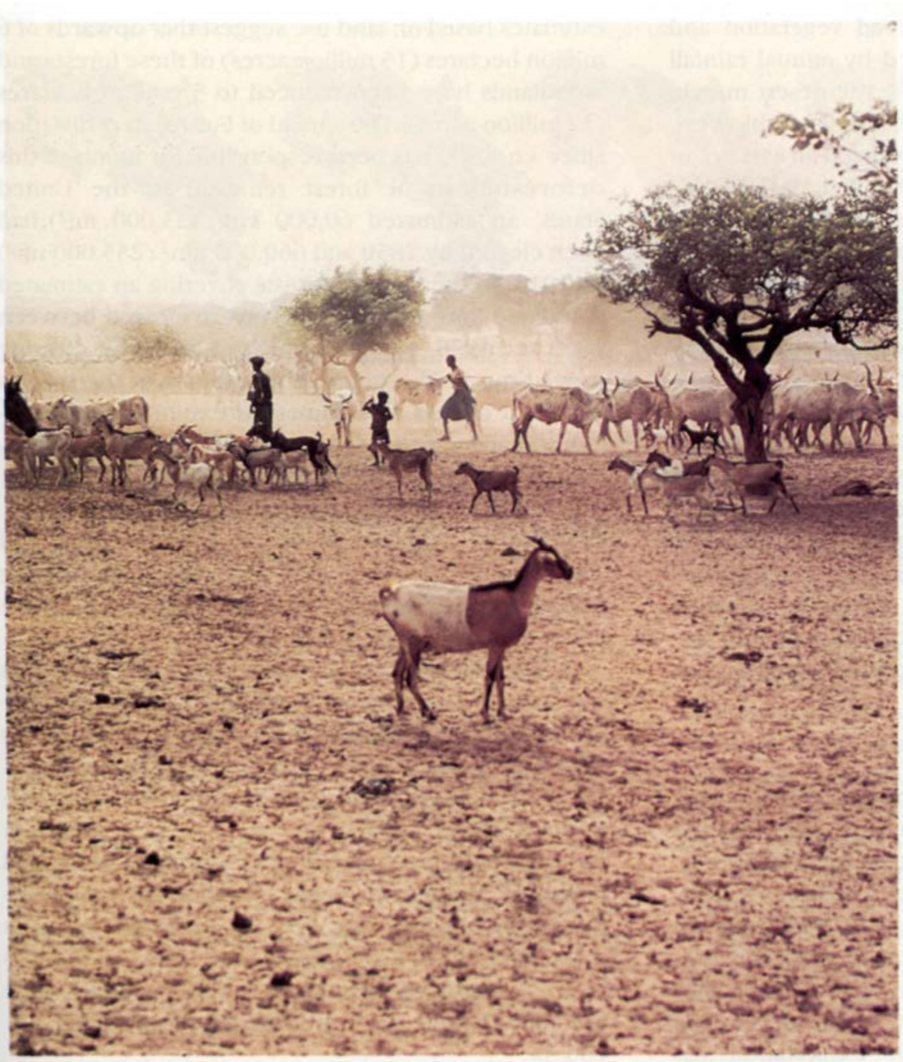
Figure 18.2 Geologic setting of Kesterson Reservoir in the western San Joaquin Valley, California. Runoff carries dissolved selenium-bearing salts from the coastal mountains to alluvial fans where the salts are precipitated in the soil. Artificial irrigation water then flushes the salts into a drainage system that carries the selenium to Kesterson Reservoir where it becomes concentrated to toxic levels.

Desertification in Africa's Sahel

People living in and near desert regions typically lead a precarious life. Food and water are precious commodities, and their availability can mean the difference between life and death. Despite such uncertainties, significant numbers of people occupy the world's dry lands, especially along the desert margins of sub-Saharan Africa.

Recently, significant land degradation and dry lands expansion in the populated Sahel region of Africa, a semi-arid belt bordering the Sahara, has attracted world attention because of widespread famine (Fig. 18.3). The term **desertification** was coined when the United Nations General Assembly convened a conference in 1977 to study the problem of land degradation resulting from human impact. Desertification is the major environmental problem of arid landscapes, which constitute 40 percent of the world's land area. The most obvious symptoms include crop failures or reduced yields, reduction in rangeland biomass available to livestock, reduction in fuelwood supplies, reduction in water supplies resulting from decreased streamflow or a depressed groundwater table, advance of dune sand over agricultural lands (Fig. 18.4), and disruption of life-support systems leading to refugees seeking outside relief.

Although desert expansion can result from natural processes, it is widely thought that excessive human exploitation of dry lands, generally linked to increasing human and livestock populations, can lead to progressive deterioration of the land and ultimately to desertification, sometimes triggered or exacerbated by natural drought. Although a strong scientific case can be made that human activity has caused desertification in some areas, opinions are not always consistent. Studies in the Sahel have led some researchers to conclude that arid land areas have increased by nearly 54 million hectares (133 million acres) since 1931 as a result of a 30 percent decline in rainfall over this interval. Other scientists found no evidence of persistent trends toward desert conditions between 1962 and



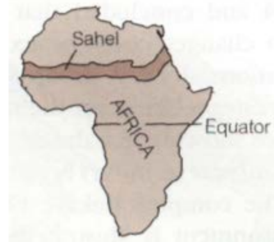
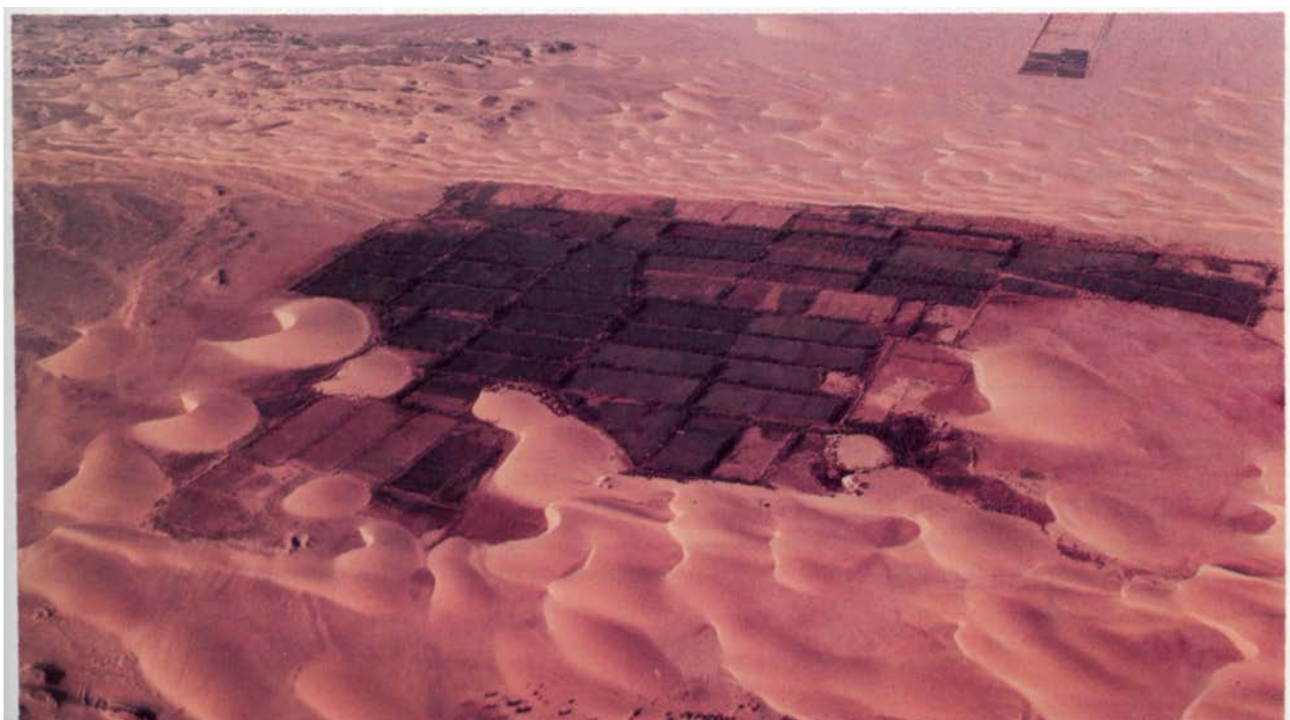


Figure 18.3 Overgrazing during years of drought killed much of the vegetation in this part of the Sahel in Senegal. Without vegetation, topsoil is eroded and the land becomes infertile.

Figure 18.4 Barchan dunes advance from right to left across irrigated fields in the Danakil Depression, Egypt.



1984 and concluded that observed vegetation and crop changes could be explained by annual rainfall variations alone. A group studying the desert margin in western Africa, on the other hand, assembled evidence showing that the desert advanced an average of 10 km/year (6 mi/yr) between 1961 and 1987.

The complex linkage between humans and their environment is illustrated by an hypothesis of progressive desertification that links grazing animals, the vegetation they consume, and the overlying atmosphere. As animals consume vegetation cover in dry lands, the albedo (reflectivity) of the land surface increases. (Sand and bare rock have a higher albedo than does grassland.) This causes more of the incoming solar radiation to be reflected back into space, leading to a cooler ground surface. The cooler ground is associated with descending dry air and therefore with reduced precipitation. In this way, the degraded area becomes more and more desertlike; in other words, a *positive feedback* is set in motion that promotes increasing desertification once the process begins.

Impacts of Deforestation

Over the last 3000 to 4000 years, and most notably within the last century, humans have been the primary agent in changing the world's vegetation cover. As the human population expanded, forests were replaced with agricultural land, and the increasing exploitation of forests for fuel and other economic uses has had a dramatic impact in many regions. Especially hard-hit have been tropical forests in developing countries, where nonsustainable exploitation has been especially damaging. Over major parts of the world, the natural vegetation has all but disappeared, to be replaced by an artificial one dominated by agriculture and introduced species. Although the distribution of pre-agricultural forests is unknown in detail, estimates based on land use suggest that upwards of 6 million hectares (15 million acres) of these forests and woodlands have been reduced to 5 million hectares (12 million acres). The spread of European civilization since AD. 1500 has been responsible for much of this **deforestation,** or forest removal. In the United States, an estimated 60,000 km^2 (23,000 mi^2) had been cleared by 1850 and 660,000 km^2 (255,000 mi^2) by 1910. In the tropics, forests covering an estimated 2,400,000 km^2 (930,000 mi^2) were cleared between I860 and 1978. Although much of the forest clearing carried out in North America and part of the tropics has been related to commercial exploitation of agricultural and forest products, in Africa nearly 60 percent is related to fuelwood production. At present, the tropical forests are receiving the greatest impact. Data for 87 countries in Africa, Latin America, and Asia show that between 1980 and 1990 forest clearing was eliminating an average of about 1 percent of the forest cover per year (Table 18.1).

Recent land clearing has been especially intense in the tropical Americas. In Costa Rica, for example, forest covered 67 percent of the country in 1940, but by 1983 the figure had been reduced to 17 percent (Fig. 18.5). Especially hard-hit has been the Amazon basin of Brazil, where the arrival of waves of new settlers in the late 1970s reached 5000 per month. Massive land clearance adjacent to major access roads rapidly reduced the forest cover, but the unproductive soils made agriculture unprofitable, and the cleared lands have also proved unsuitable for sustainable cattle ranching (largely for 'fast food' hamburger) (see Fig. C11.1).

The environmental implications of deforestation in Amazonia reach well beyond the economic arena, however. Research in this region has shown that when the forest is cleared, the hydrological balance can be severely upset. In one study area streams carry

TABLE 18.1
Food and Agricultural Organization (FA0) 1991 Estimates of Recent Forest Cover and Deforestation for the Tropical Regions

Continent	Forest Area in 1980 (10^6 ha)	Forest Area in 1990 (10^6ha)	Annual Deforestation (1981-1990) (10^6 ha)	Rate of Change (1981-1990) (%/year)
Africa	650	600	5.0	-0.8
Latin America and Caribbean	923	840	8.3	-0.9
Asia	321	275	3.6	-1.2
Total	1894	1715	16.9	

Source: From M. K. Tolba, and D. A. El-Kholy (eds.). The World Environment 1972-1992. (London: Chapman and Hall, 1992), pl69-

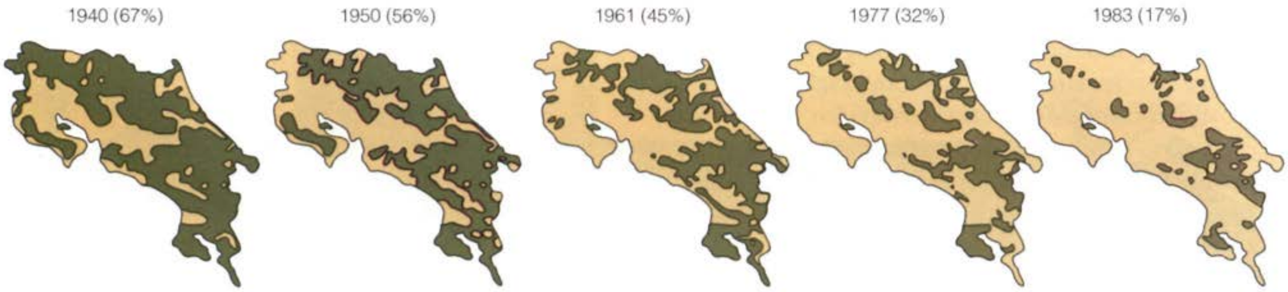


Figure 18.5 Dramatic deforestation in Costa Rica between 1940 and 1983 reduced the percentage of forest as a proportion of the total area of the country from 67 percent to 17 percent.

away 25 percent of the rainfall, while trees and other plants return 50 percent of the precipitation to the atmosphere by transpiration. Moist air masses moving inland from the Atlantic bring about half the water vapor that falls as rain; the other half is supplied by the forest in the form of evapotranspiration. This means that the forest plays a key role in maintaining the precipitation balance, and thus the forest itself, in the Amazon basin.

Removal of forest changes the hydrological equation. Without a forest cover to promote infiltration, more of the rainfall runs off the land and far less is recycled through plants back into the atmosphere. As a result, the potential exists for a *negative feedback,* whereby destruction of forest leads to reduction in the rainfall that the forest requires for its very existence. Although the ultimate effects of continued deforestation are still difficult to predict, an estimated 12 percent of the Amazonian rain forest has already been cleared, and in some regions that rate of land clearance has been proceeding exponentially. If continued unchecked, the repercussions could include replacement of productive rain forest by nonproductive farm and grazing lands, and a significant change in the regional hydrology that might lead to widespread reduction in soil moisture, increased flood discharges, and worsening droughts.

THE CHANGING ATMOSPHERE

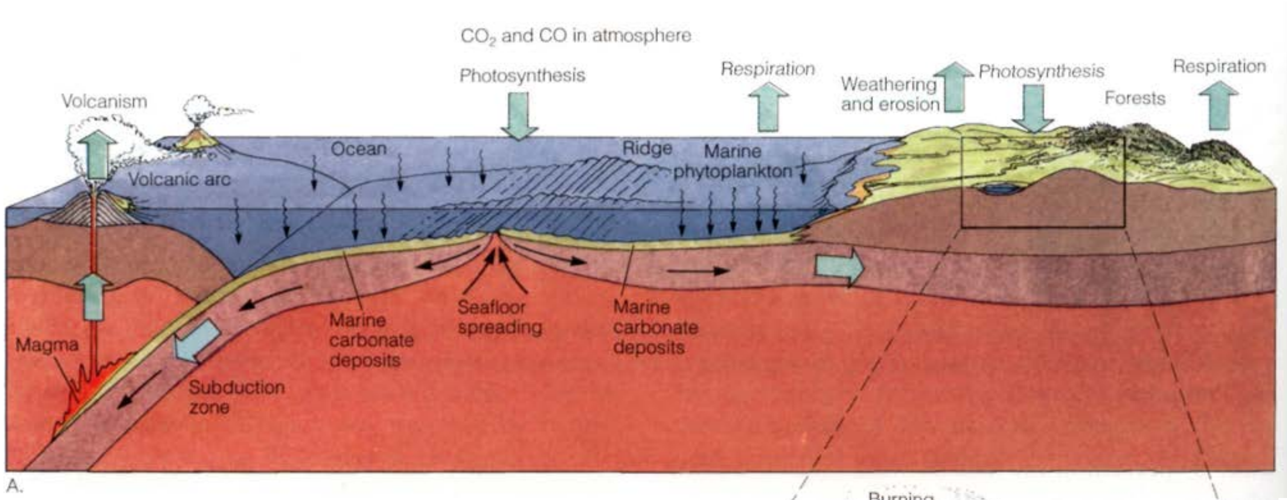
As we learned in Chapter 14, the Earth's climate system consists of a number of interacting subsystems that involve the atmosphere, the hydrosphere, the solid Earth, and the biosphere (Fig. 14.2). The interactions are extremely complex and difficult to analyze. Consequently, only with the advent of supercomputers have we begun to answer some of the basic questions about how the climate system works.

The geologic record plays an important role in this enterprise, for it contains a history of the Earth's changing climates that extends into the remote past. Climatic change can be read from the stratigraphic record in many ways. For example, paleontologists infer past climates from the assemblages of fossil plants and animals they find in ancient strata. Sedimentologists and stratigraphers can infer many things about past climates from the nature of the sediments they study: the present distribution of these sediments, their mineralogy, the varied depositional environments represented, features that indicate the agencies of sediment transport, and the soils that represent former land surfaces. Isotope geologists can determine past surface temperatures from studies of sediments on land, in the oceans, and in polar ice sheets. By reconstructing past climates, the range of climatic variability on different time scales can be determined, and the accuracy of computer models that try to simulate past climatic conditions can be tested.

One important reason for these studies is to learn how the climate system behaves, what controls it, and how it is likely to change in the future. We know it will change, but we lack a clear view of how and at what rate. The answers are important not only scientifically, but sociologically and politically as well, for in extracting and burning the Earth's immense supply of fossil fuels people have unwittingly begun a great geochemical "experiment" that is likely to have a significant impact on our planet and its inhabitants.

The Carbon Cycle

A basic chemical substance involved in the climate "experiment" is carbon, an element that is essential to all forms of life. Carbon occurs in four reservoirs: (1)



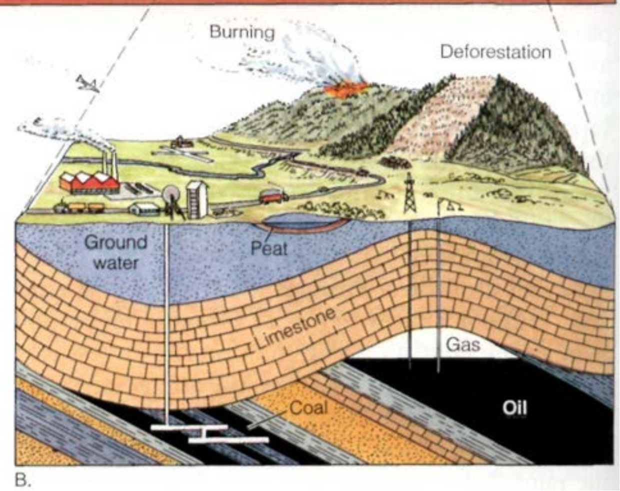


Figure. 18.6 The carbon cycle. A. Natural fluxes of carbon through the atmosphere, hydrosphere, biosphere, and lithosphere. Carbon enters the atmosphere through volcanism, weathering, biological respiration, and decay of organic matter in soils. Photosynthesis incorporates carbon in the biosphere, from which it can become part of the lithosphere if buried with accumulating sediment. B. Human activities release carbon to the atmosphere through burning of forests and fossil fuels.

as carbon dioxide in the atmosphere; (2) in organic compounds in the biosphere; (3) as dissolved carbon dioxide in the hydrosphere; and (4) in the calcium carbonate of limestone and in decaying and buried organic matter (peat, coal, and petroleum) in the lithosphere. Each reservoir is involved in the carbon cycle (Fig. 18.6).

The key to the carbon cycle is the biosphere, where plants continuously extract CO_2 from the atmosphere and then break the CO_2 down by photosynthesis to form organic compounds. Animals consume plants and use these organic compounds in their metabolism. When plants and animals die, the organic compounds decay by combining with oxygen from the atmosphere to form CO_2 again. The passage of material through the biosphere is so rapid that the entire content of CO_2 in the atmosphere cycles every 4.5 years.

Not all dead plant and animal matter in the biosphere decays immediately back to CO_2. A small fraction is transported and redeposited as sediment; some is then buried and incorporated in sedimentary rock where it locally forms deposits of coal and petroleum. The buried organic matter joins the slower moving rock cycle and can reenter the atmosphere naturally only after uplift and erosion have exposed the rock in which it is trapped.

Carbon dioxide from the atmosphere is also dissolved in the waters of the hydrosphere. There it is used by aquatic plants in the same way that land plants use CO_2 from the atmosphere. In addition, aquatic animals extract calcium and carbon dioxide from the water to make shells of $CaCO_3$. When the animals die, the shells accumulate on the seafloor, mixing with any $CaCO_3$ that may have been precipitated as chemical sediment. When compacted and cemented, the $CaCO_3$ forms limestone (Fig. 18.7). In this way, too, some carbon joins the rock cycle. Eventually, the rock cycle will bring the limestone back to the surface where weathering and erosion will break it down; the calcium returns in solution to the ocean, and the carbon escapes as CO_2 to the atmosphere.

Now, let's consider what happens when human activities change these four carbon reservoirs and influ-

Figure 18.7 Vast bodies of carbonate rocks, like the Dolomites of northern Italy, constitute reservoirs of carbon dioxide that has been temporarily removed from the carbon cycle. Once exposed at the surface, carbonate rocks are weathered and eroded, thereby freeing CO_2, which reenters the carbon cycle.

ence the exchanges between them. While any individual action may appear insignificant, the cumulative effects of all human activities are now so great as to be measurable. The burning of fossil fuels and the clearing of forested land cause CO_2 to move from the lithosphere and the biosphere to the atmosphere at rates much faster than they would move naturally. Unless this additional CO_2 is dissolved in the hydrosphere or is buried in sediments as fast as it is generated, the CO_2 content of the atmosphere must inevitably increase. However, the rate at which these natural processes are removing CO_2 from the atmosphere is slower than the rate at which human activities are adding it, leading us to conclude that the CO_2 content of the atmosphere should be increasing.

The Greenhouse Effect

The atmosphere is the engine that drives the Earth's climate system, and the Sun provides the energy that allows the engine to work (Fig. 18.8). Some of the solar radiation that reaches the atmosphere is reflected off clouds and dust and bounces back into space. Of the radiation that reaches the Earth's surface, some is absorbed by the land and oceans, and some is reflected into space by water, snow, ice, and other highly reflective surfaces. This visible reflected solar radiation has a short wavelength.

The Earth also emits long-wave, infrared radiation. However, some of the long-wave radiation does not escape into space, but instead is absorbed by atmospheric water vapor and CO_2. Because this radiant energy is retained in the lower atmosphere, the temperature at the Earth's surface rises. A comparable effect also explains why the air temperature in a glass greenhouse is warmer than the air outside: the glass, acting in much the same way as the atmospheric gases, prevents the escape of radiant energy. Hence, we refer to this phenomenon as the **greenhouse effect.**

It is the greenhouse effect that makes the Earth habitable. Without it, the surface of our planet would be as inhospitable as those of the other planets in the solar system. If the Earth lacked an atmosphere, its surface environment might be like that of the Moon: on the sunlit side the temperature is close to the boiling point of water, while on the dark side it is far below freezing. The nearly airless surface of Mars is a frigid landscape whose closest earthly analogs are the frozen polar deserts. By contrast, Venus is much closer to the Sun, and its atmosphere is so dense that the greenhouse effect generates surface temperatures hot enough to melt lead.

Greenhouse Gases

Dry air consists mainly of three gases: nitrogen (79%), oxygen (20%), and argon (1%). However, water vapor is usually present in the Earth's atmosphere in concentrations of up to several percent and accounts for about 80 percent of the natural greenhouse effect. The remaining 20 percent is due to other gases present in very small amounts. Despite their very low concentrations, measurable in parts per billion by volume (ppbv) of air, these *trace gases* contribute significantly to the greenhouse effect (Table 18.2).

Chief among the trace gases is carbon dioxide, while other significant gases, each a basic part of natural biogeochemical cycles and efficient in absorbing infrared radiation, are methane, nitrous oxide, and ozone. The commercially produced CFCs are an additional important group of greenhouse gases.

Trends in Greenhouse Gas Concentrations

During the past decade, the greenhouse gases have received increasing scientific and public attention as it has become clear that their atmospheric concentrations are rising.

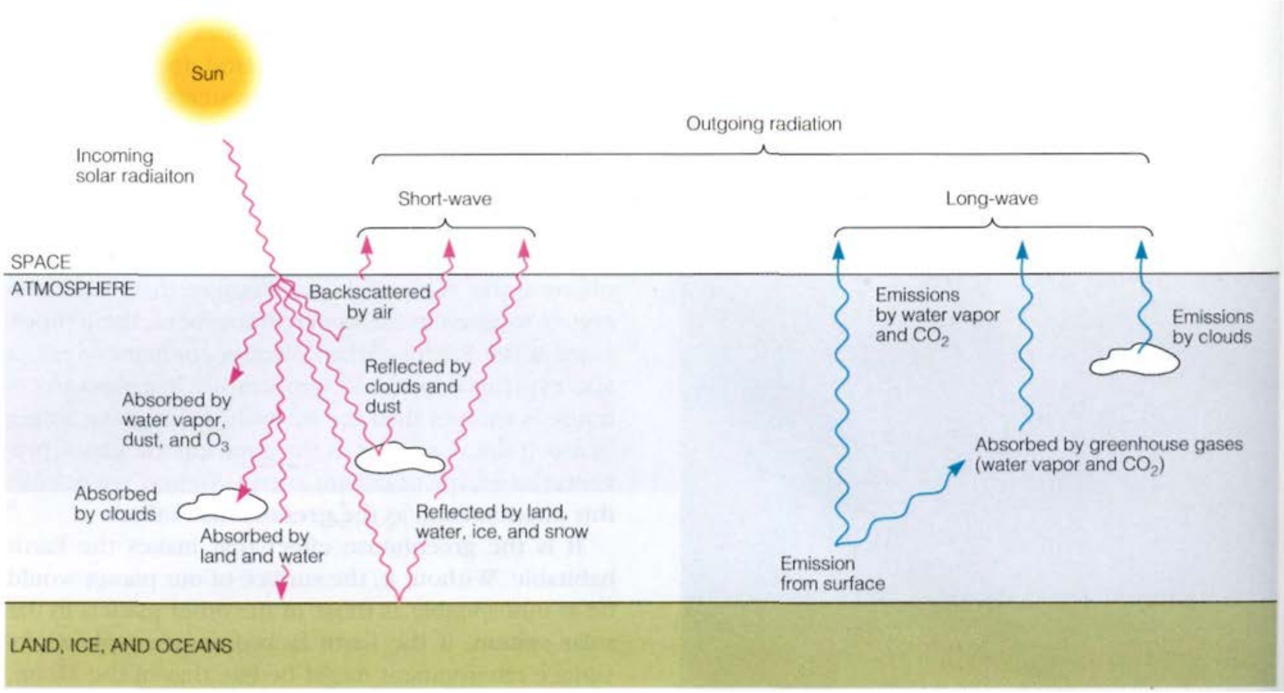


Figure 18.8 Some of the short-wave (visible) solar radiation reaching the Earth is absorbed by land, oceans, clouds, and atmospheric dust and gases, and some is reflected back into space by reflective surfaces that include snow, ice, clouds, and dust. The Earth also radiates long-wave radiation back into space. Greenhouse gases trap some of the outgoing long-wave radiation, causing the air temperature of the lower atmosphere to rise.

Table 18.2
Atmospheric Trace Gases Involved in the Greenhouse Effect

	Carbon Dioxide (CO_2)	Methane (CH_4)	Nitrous Oxide (N_2O)	Chlorofluoro-carbons (CFCs)	Tropospheric Ozone (O_3)	Water Vapor (H_2O)
Greenhouse	Heating	Heating	Heating	Heating	Heating	Heats in air; cools in clouds
Effect on strato-spheric ozone	Can increase or decrease	Can increase or decrease	Can increase or decrease	Decrease	None	Decrease
Principal anthro-pogenic sources	Fossil fuels; deforestation	Rice culture; cattle; fossil fuels, bio-mass burning	Fertilizer; land-use conversion	Refrigerants; aerosols; industrial processes	Hydrocarbons (with NO_x); biomass burning	Land conver-sion; irrigation
Principle natural sources	Balanced in nature	Wetlands	Soils; tropical forests	None	Hydrocarbons	Evapotrans-piration
Atmospheric lifetime	50-200 yr	10 yr	150 yr	60-100 yr	Weeks to months	Days
Present atmosphe-ric concentration at land surface (ppbv)	355,000	1720	310	CFC-11:0.28 CFC-12: 0.48	20-40	3000-6000
Preindustrial (1750-1800) concentration at land surface (ppbv)	280,000	790	288	0	10	Unknown
Present annual rate of increase	0.5%	1.1%	0.3%	5%	0.5-2.0%	Unknown
Relative contribu-tion to the anth-ropenic green-house effect	60%	15%	5%	12%	8%	Unknown

Source: Earthquest 5, No. 1 (1991).

Carbon Dioxide Beginning in 1958, measurements have been made of carbon dioxide concentration in the atmosphere near the top of Mauna Loa volcano, Hawaii. This site was chosen because of its altitude and remote location far from sources of atmospheric pollution. The measurements show two remarkable things. First, the amount of CO_2 fluctuates regularly with an annual rhythm (Fig. 18.9). In effect, the Earth or, more correctly, the biosphere, is breathing! During the growing season, CO_2 is absorbed by vegetation, and the atmospheric concentration falls; then, during the winter dormant period, more CO_2 enters the atmosphere than is removed by vegetation, and the concentration rises. Second, the long-term trend is an unmistakable rise in concentration. Since 1958, the atmospheric CO_2 concentration has risen from 315,000 to 355,000 ppbv, and the rise is not linear but exponential (i.e., the rate is increasing with time).

The rising curve of atmospheric CO_2 immediately raises two questions: (1) Is the observed rise unusual?, and (2) How can it be explained? To answer the first question, we must turn to the geologic record. As discussed in Chapter 14, ice cores from the Antarctic and Greenland ice sheets contain samples of ancient atmosphere. The glacier records show that the preindustrial concentration of CO_2 was close to 280,000 ppbv, the typical value for an interglacial age. The subsequent rapid increase to 355,000 ppbv during the last 100 years is unprecedented in the ice-core record and implies that something very unusual is taking place.

A possible explanation for the extraordinary recent rise in atmospheric CO_2 is immediately suggested if we examine the rate at which this gas has been added to the atmosphere since the beginning of the Industrial Revolution (Fig. 18.10). The curve tracking the increase in CO_2 closely resembles the curve showing the increase in carbon released to the atmosphere by the burning of fossil fuels. Because no known natural mechanism can explain such a rapid increase in CO_2, the inescapable conclusion is that the human burning of fossil fuels must be a primary reason for the observed increase in atmospheric CO_2. Additional contributing factors must be widespread deforestation, with its attendant burning and decay of cleared vegetation, and the use of wood as a primary fuel in many underdeveloped countries that have rapidly growing populations.

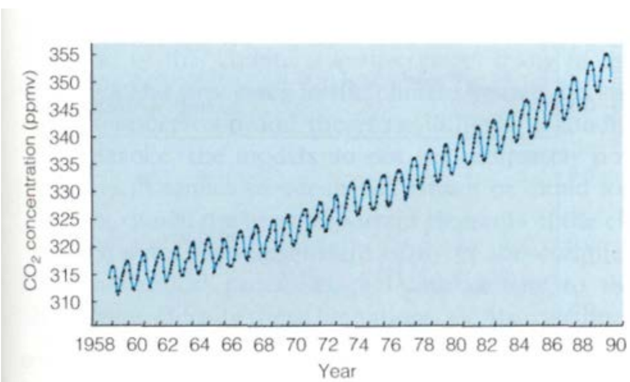


Figure 18.9 Concentration of carbon dioxide in dry air, since 1958, measured at the Mauna Loa Observatory, Hawaii (given in parts per billion by volume = ppbv/1000). Annual fluctuations reflect seasonal changes in biologic uptake of CO_2, while the long-term trend shows a persistent increase in the atmospheric concentration of this greenhouse gas.

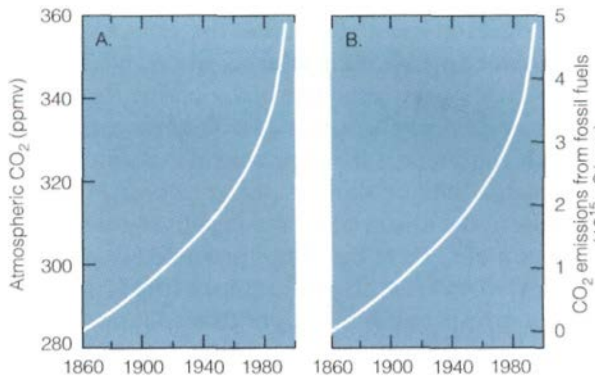


Figure 18.10 Since the beginning of the Industrial Revolution, about 1850, the atmospheric concentration of CO_2 has risen at an increasing rate (A). The increase matches the increasing rate at which CO_2 has been released through the burning of fossil fuels (B).

Methane Methane gas (CH_4) absorbs infrared radiation 25 times more effectively than CO_2, making methane gas an important greenhouse gas despite its relatively low atmospheric concentration (Table 18.2). Since the late 1960s, when measurements of atmospheric methane began, the concentrations have increased at a rate of about 1 percent per year (although since 1984 the rate has decreased slightly, for unknown reasons). Methane levels for earlier times obtained from ice-core studies show an increase that essentially parallels the rise in the human population. This relationship is not surprising, for much of the methane now entering the atmosphere is generated (1) by biological activity related to rice cultivation and (2) as a byproduct of the digestive processes of domestic livestock, especially cattle. The global livestock population increased greatly in the past century, and the total acreage under rice cultivation has increased more than 40 percent since 1950.

In prehistoric times, methane levels, like CO_2 levels, increased and decreased with the glacial/interglacial cycles (Fig. 14.26).

Other Important Trace Gases CFC-12, used mostly as a refrigerant, has 20,000 times the capacity of carbon dioxide to trap ultraviolet radiation, while CFC-11, which is widely used in making plastic foams and as an aerosol propellant, has 17,500 times the capacity. Both compounds are increasing in the atmosphere at an annual rate of about 5 percent. As we have already seen, the observed increasing atmospheric concentration of CFCs in the 1980s produced worldwide concern because the scientific consensus is that these gases destroy ozone in the upper atmosphere, thereby leading to the formation of the Antarctic ozone hole.

Tropospheric ozone and nitrous oxide are increasing annually at rates of 0.5 to 2 percent, and 0.3 percent, respectively, and together account for about 13 percent of the greenhouse effect. Although ozone in the upper atmosphere is beneficial because it traps harmful infrared solar radiation, when this gas builds up in the troposphere it constitutes a greenhouse gas. Tropical forests are important in photosynthetically removing excess tropospheric ozone, which is largely produced by the combustion of fossil fuels. However, the wholesale destruction of these forests could lead to further concentration of ozone in the atmosphere. Nitrous oxide, released by microbial activity in soil, the burning of timber and fossil fuels, and the decay of agricultural residues, has a long lifetime in the atmosphere. Accordingly, atmospheric concentrations are likely to remain well above preindustrial levels even if emission rates stabilize.

Global Warming

If the atmospheric concentration of the greenhouse gases is rising, what does this portend for future climate? Does it mean that the Earth's surface temperature is warming, and, if so, by how much and at what rate? To try and answer these questions, we first look at the historical record and then see how forecasts of the future can be made.

Historical Temperature Trends

Correctly assessing recent global changes in temperature is a very difficult task. The difficulty arises from the fact that very few instrumental measurements were made before 1850, and the majority date to the time since World War II. The earliest records are from Western Europe and eastern North America. Data for oceanic areas, which encompass 70 percent of the globe, are sparse and decrease significantly in number prior to 1945. Therefore, most "global" temperature curves are reconstructed primarily from land stations located mainly in the northern hemisphere.

Numerous curves of average annual temperature variations since the middle or late nineteenth century have been published, and although they differ in detail, they all show one characteristic feature: a long-term rise in temperature during the past century (Fig. 18.11). Although short-term departures from this trend are evident, the total temperature increase since 1860 when reliable hemispheric-wide records begin is about 0.5°C (0.9°F).

Because the interval of rising temperatures coincides with the time of rapidly increasing greenhouse gas emissions, it is tempting to assume that the two phenomena are causally related. However, because the temperature reconstructions prior to 1950 are based on relatively few data that are unequally distributed across the globe, a convincing case is difficult to make. Even if the temperature curves do approximate actual trends, it might be argued that the modest rise in global temperature during the past century falls within the natural variability of the Earth's climate system and would have occurred even if the greenhouse gas concentrations had not increased.

If the historical temperature record is judged to be inconclusive, we can still explore the linkage between greenhouse gas emissions and present and future climate by turning to models of the climate system.

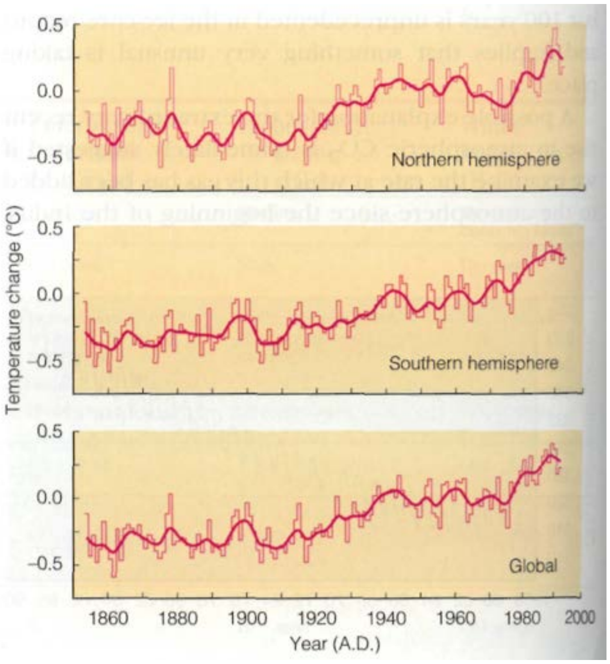


Figure 18.11 Hemispheric and global mean temperature changes since the mid-nineteenth century, incorporating both land and ocean data. Annual averages (blue line) and smoothed averages (red line) are shown. The total global rise for this interval is about 0.5 °C.

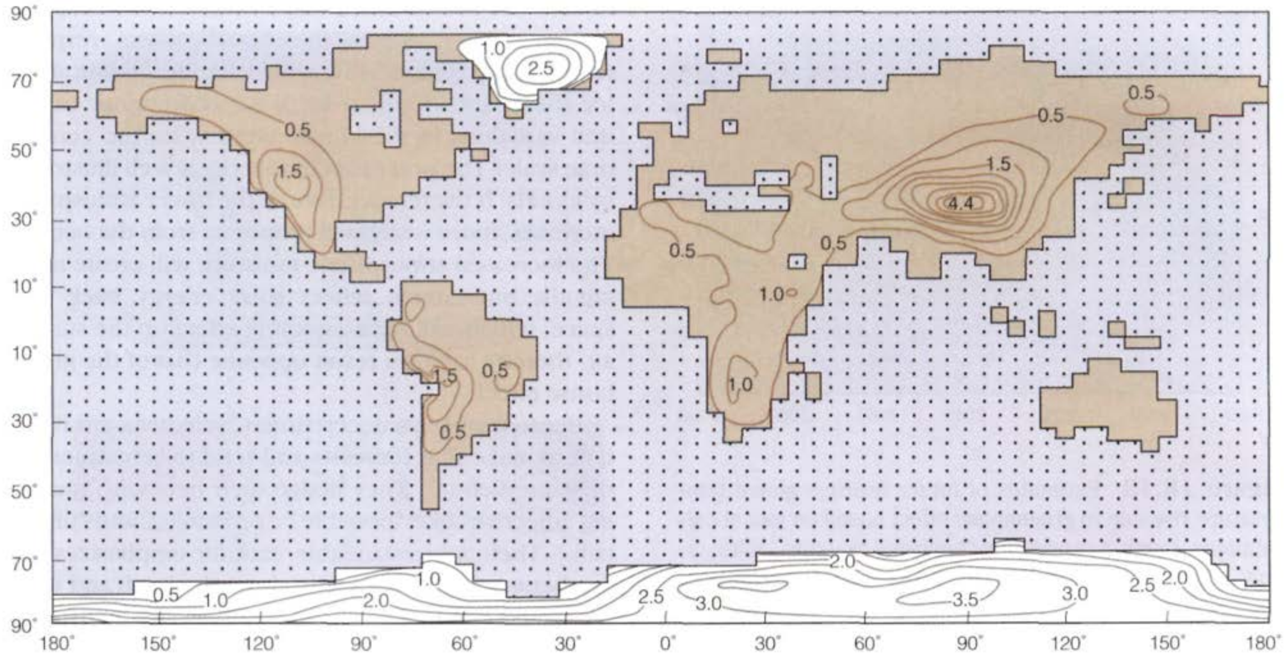


Figure 18.12 View of the Earth as depicted by a global climate model having a grid spacing of 4° of latitude by 5° of longitude. Spacing of gridpoints is shown over the oceans but not over land. The altitudes of mountain ranges and plateaus are generalized and expressed in thousands of meters. Because of the map projection, Antarctica appears at the bottom, unrealistically, as a long, narrow continent.

Climate Models

Three-dimensional mathematical models of the Earth's climate system are an outgrowth of efforts to forecast the weather. The most sophisticated are general circulation models (GCMs) that attempt to link atmospheric, hydrospheric, and biospheric processes. The sheer complexity of these natural systems means that such models, of necessity, are greatly simplified representations of the real world (see "Guest Essay" at the end of this chapter). Furthermore, many of the linkages and processes in the climate system are still poorly understood and therefore difficult to model. For instance, the models do not yet adequately portray the dynamics of ocean circulation or cloud formation, two of the most important elements of the climate system. Also absent are many of the complex biogeochemical processes that link climate to the biosphere. Despite these limitations, GCMs have been very successful in simulating the general character of present-day climates and have greatly improved weather forecasting. This success encourages us to use these models to gain a general global picture of future climates as the Earth's physical and chemical balance changes.

Because the solution of the complex mathematical equations of a GCM requires considerable amounts of computer time, the three-dimensional grid spacing (the distance between points on and above the globe for which the solutions are calculated) in the model experiments is large in order to keep costs manageable. As a result, the resolution of the models is relatively coarse: grid points commonly are separated by 4 or 5 degrees of latitude, or 450 to 550 km (Fig. 18.12). Therefore, although these models can generate a reasonable picture of global and hemispheric climatic conditions, they are poor at resolving conditions at the scale of small countries, states, or counties. Until more powerful computers are built, or the cost of running a model experiment decreases, the spatial resolution of GCMs is likely to remain relatively coarse.

Estimates of Future Greenhouse Warming

Predictions of climatic change related to greenhouse warming are based mainly on the results of several climate models that differ in detail as well as in the assumptions they employ. Nevertheless, the models predict that the anthropogenically generated greenhouse gases already in the atmosphere would cause an average global temperature increase of 0.5 to 1.5°C (0.9 to 2.7°F) (Fig. 18.13). This prediction is consis-

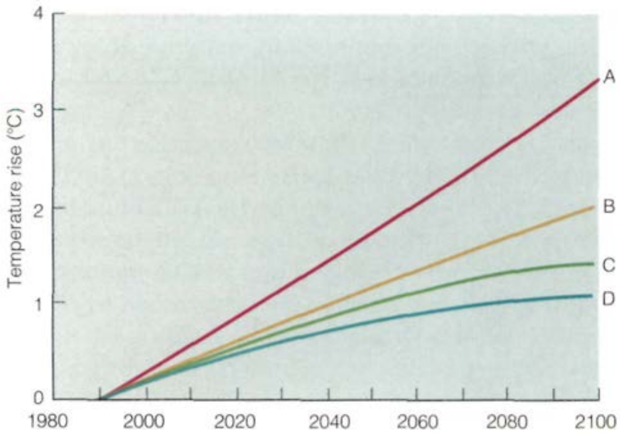


Figure 18.13 Estimates of future average global temperature rise due to greenhouse effect based on model calculations. Curve A represents "business as usual," with energy supply dominated by coal, continuing deforestation, and limited or no control of methane, nitrous oxide, and carbon monoxide emissions. Curve B assumes a shift toward lower-carbon fuels (e.g., natural gas), coupled with large increases in fuel efficiency, reversal of deforestation trends, and stringent carbon monoxide controls. Curve C assumes a shift toward renewable sources (solar, hydro-, and wind power) and nuclear energy in the second half of the twenty-first century, a phaseout of CFC emissions, and limitations on agricultural methane and nitrous oxide emissions. Curve D assumes a shift to renewable and nuclear energy in the first half of the twenty-first century, thereby reducing carbon dioxide emissions to 50 percent of 1985 amounts by 2050, and stabilization of atmospheric concentrations of greenhouse gases through controls in industrialized countries and moderate growth in developing countries. In deriving all these estimates, world population was assumed to reach 10.5 billion in the second half of the next century, and economic growth was assumed to be modest (2-5%) in the next decade but decreasing thereafter.

tent with the 0.5°C rise in temperature inferred from the instrumental record. The models further predict that if the greenhouse gases continue to build up until their combined effect is equivalent to a doubling of the preindustrial CO_2 concentration, then average global temperatures most likely will rise between 1.5 and 4.5°C (2.7 and 8.1°F); a best guess is an increase of 1.8°C (3.2°F) by the year 2030, with an eventual increase of 2.5°C (4.5°F) by the end of the twenty-first century. This does not mean that the temperature will increase uniformly all over the Earth. Rather, the projected temperature change varies geographically, with the greatest change occurring in the polar regions (Fig. 18.14; Table 18.3).

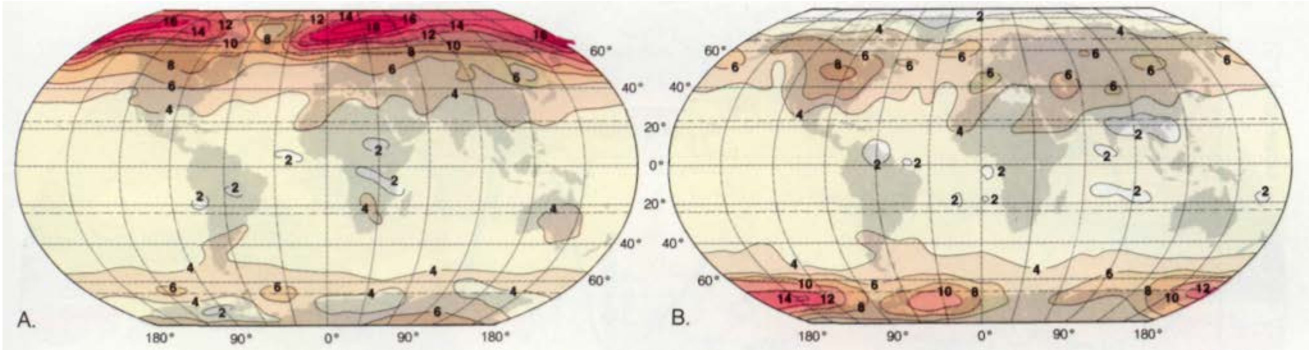
The rate of projected warming depends on a number of basic uncertainties: How rapidly will concentrations of the greenhouse gases increase? How rapidly will the oceans, a major reservoir of heat and a fundamental element in the climate system, respond to changing climate? How will changing climate affect ice sheets and cloud cover? What is the range of natural variations in the climate system on the century time scale? The potential complexity is well illustrated by clouds. If the temperature of the lower atmosphere increases, more water will evaporate from the oceans. The increased atmospheric moisture will create more clouds, but clouds reflect solar energy back into space, which will have a cooling effect on the surface air, thereby having a result opposite that of the greenhouse effect.

Because of such uncertainties, scientists are reluctant to make firm forecasts and tend to be cautious in their predictions. They hedge their bets with qualifying adjectives like "possible," "probable," and "uncertain." Their understandable caution emphasizes the gap between what we know about the Earth and what we would like to know, and points to the many challenges that still face scientists studying global change.

Despite the uncertainties, the general consensus is that (1) human activities have led to increasing atmospheric concentrations of carbon dioxide and other trace gases that have enhanced the greenhouse effect; (2) global mean surface air temperature has increased by0.3 to0.6°C(0.5 to 1.1°F) during the last 100 years, and this increase may be the direct result of the enhanced greenhouse effect; and (3) during the next century global average temperature will likely increase at about 0.3°C (0.5°F) per decade, assuming emission rates do not change. This projected increase may lead to a global average temperature about 1.5 to 1.8°C (2.7 to 3.2°F) warmer than present by the year 2030 and as much as 2.5°C (4.5°F) warmer by the end of the next century (Fig. 18.15). If governmental controls lead to lower emission rates, the per decade rise in temperature may be only 0.1 to 0.2°C (0.2 to 0.4°F). Nevertheless, the temperature increase related to the continued release of greenhouse gases will be larger and more rapid than any experienced in human history. Thus, we may be moving toward a "super interglaciation," warmer than any interglaciation of the past two million years (Fig. 18.15).

Environmental Effects of Global Warming

An increase in global surface air temperature by a few °C does not sound like much. Surely, we can put up with this rather insignificant change. However, if we stop and consider that the difference in average global temperature between the present and the coldest part of the last ice age was only about 5°C (9°F), we can begin to see how a temperature change of even a degree or two could well have global repercussions.



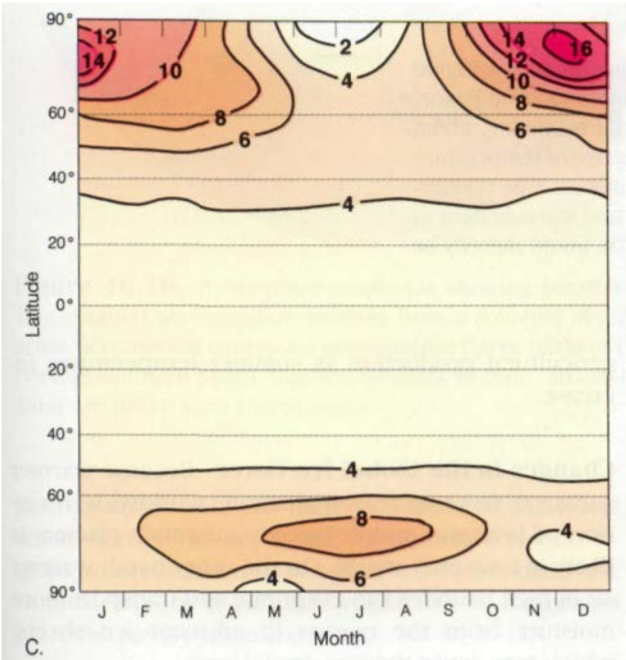


Figure 18.14 A forecast of future changes in surface air temperature (in °C) resulting from an effective doubling of atmospheric CO_2 concentration relative to that of the present. A. Temperature increases for Winter (December, January, February). For example, along the lines labeled 4, the projected temperature increase is everywhere 4°C. B. Temperature increases for Summer (June, July, August).C. A latitudinal cross section showing changes in zonal average air temperature through the year. This graph is a summary of the map patterns shown in A and B, but includes the spring and autumn months as well. Greenhouse warming is greatest at high latitudes, where temperature increases as great as 16 °C are forecast by the model for the northern hemisphere winter.

TABLE 18.3
Estimates of Changes in Average Surface Air Temperature and Precipitation for Selected Regions, Preindustrial Times to 2030 A.D.

Region[a]	Temperature		Precipitation	
	Winter[b]	Summer[b]	Winter	Summer
Central North America	2 to 4°C warmer	2 to 3°C warmer	Oto 15%	5 to 10% drier
Southeast Asia	1 to 3°C warmer	2 to 3°C warmer C°	5% drier to 15% wetter	5 to 15% wetter
Sahel, Africa	1 to 3°C warmer	1 to 2°C warmer	10% drier to no change	0 to 5%
Southern Europe	1 to 3°C warmer	2 to 3°C warmer	0 to 10% wetter	5 to 15% drier
Australia	1 to 3°C warmer	2°C warmer	5 to 15% wetter	no change

[a]Within each region, considerable variation occurs. Confidence in these estimates is low, especially with regard to precipitation values.

[b]Winter = December, January, February; Summer = June, July, August.

Source: J. T., Houghton, G. J, Jenkins, and J. J. Epbraums (eds), Climate Change, the IPCC Scientific Assessment. (Cambridge: Cambridge University Press, 1990), Table 5.1.

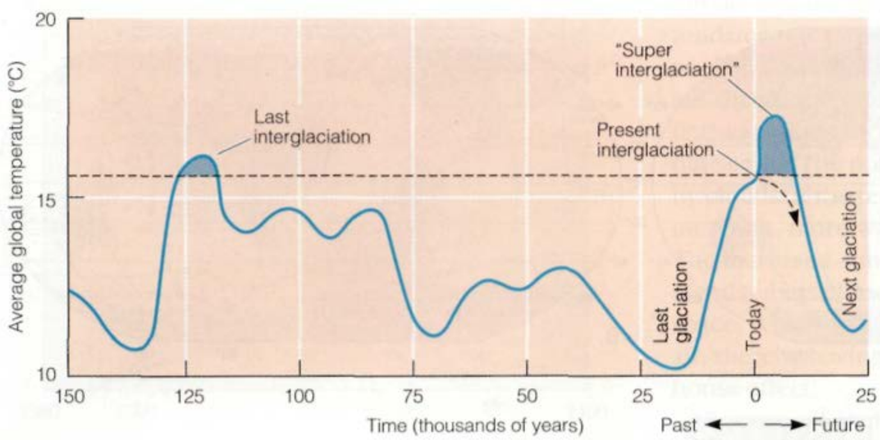


Figure 18.15 The course of average global temperature during the past 150,000 years and 25,000 years into the future. The natural course of climate (dashed [color] line) would be declining temperatures leading to the next glacial maximum, about 23,000 years from now. With greenhouse warming, a continuing rise of temperature may lead to a "super interglaciation" within the next several centuries. The temperature may then be warmer than during the last interglaciation and warmer than at any time in human history. The decline toward the next glaciation would thereby be delayed a millennium or more.

Global warming is just one result of our great geochemical "experiment." There are many physical and biological side effects that are of considerable interest and concern. Among them are the following:

Changes in Precipitation and Vegetation A warmer atmosphere will lead to increased evaporation from oceans, lakes, and streams and to greater precipitation. However, the distribution of precipitation will be uneven (Fig. 18.16; Table 18.3). Climate models suggest that precipitation rates in the equatorial regions will increase, in part because warmer temperatures will increase rates of evaporation over the tropical oceans and promote the formation of rainclouds. By contrast, the interior portions of large continents, which are distant from precipitation sources, will become both warmer and drier. Shifting patterns of precipitation and warmer temperatures will likely lead to significant local and regional changes in stream runoff and groundwater levels.

Shifting precipitation patterns are likely to upset ecosystems, causing vegetation communities and the animals dependent on them to adjust to new conditions. Forest boundaries may shift during coming centuries in response to altered temperature and precipitation patterns. Some prime midcontinental agricultural regions are likely to face increased droughts and substantially reduced soil moisture that will negatively impact crops. Higher latitude regions with short, cool growing seasons may see increased agricultural production as summer temperatures increase.

Changes in the Global Ice Cover Because warmer summers favor increased ablation, worldwide recession of low- and middle-latitude mountain glaciers is likely in a warmer world. On the other hand, warmer air in high latitudes can evaporate and transport more moisture from the oceans to adjacent ice sheets, which may cause them to grow larger.

The greatly enhanced heating projected for high northern latitudes (Fig. 18.14) favors the shrinkage of sea ice. A reduction in polar sea ice, which has a high albedo, should contribute to global warming by reducing the amount of short-wave solar radiation reflected back into space, thereby increasing the heat absorbed by the ocean. Models show much less heating in the high-latitude southern hemisphere, suggesting little change in sea-ice cover there.

Rising summer air temperatures will also begin to thaw vast regions of perennially frozen ground at high latitudes. The thawing will likely affect natural ecosystems as well as cities and engineering works built on frozen ground.

Worldwide Rise of Sea Level As the temperature of ocean water rises, its volume will expand, causing world sea level to rise. This rise in sea level, supplemented by meltwater from shrinking glaciers, is likely to increase calving along the margins of tidewater

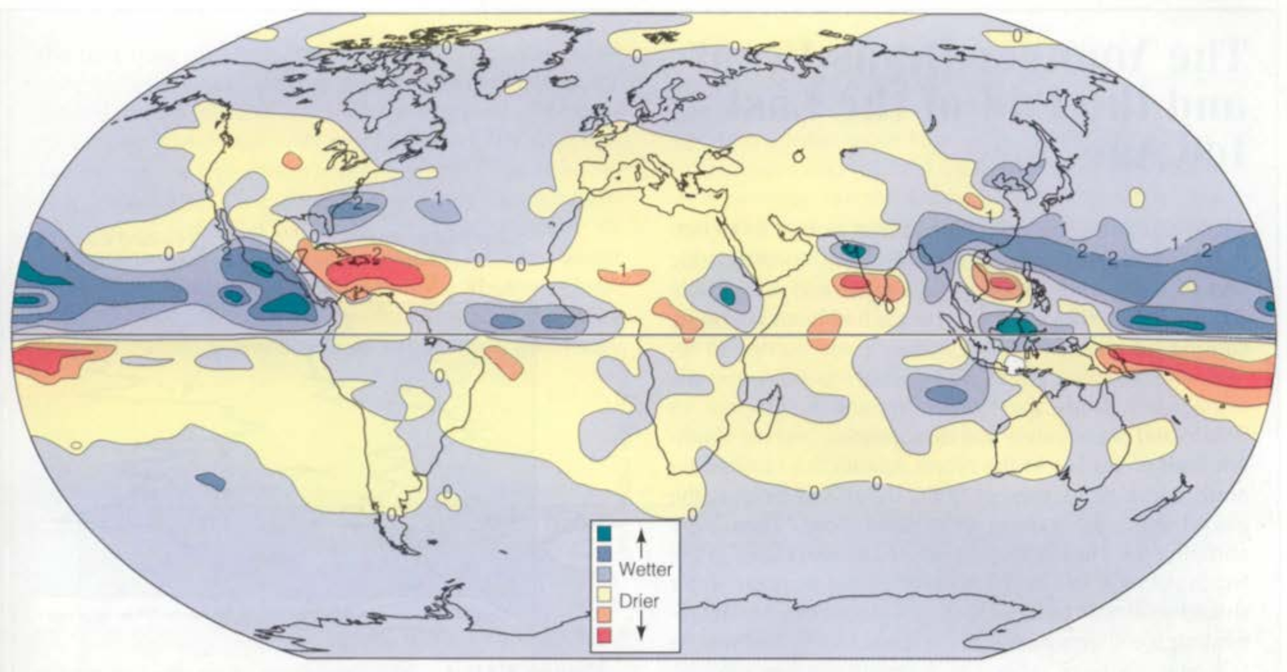


Figure 18.16 A computer simulation showing possible changes in summer (June, July, August) precipitation resulting from a doubling of atmospheric CO_2. Many of the areas of projected decreased precipitation (large parts of central North America, eastern and southern South America, Western Europe, Africa, the Middle East, and central Asia) are prime agricultural areas.

glaciers and ice sheets, thereby leading to additional sea-level rise. The rising sea will inundate coastal regions where millions of people live and will make the tropical regions even more vulnerable to larger and more frequent cyclonic storms.

Decomposition of Organic Matter in Soils As temperature rises, the rate of decomposition of organic matter in soil will increase. Soil decomposition releases CO_2 to the atmosphere, thereby further enhancing the greenhouse effect. If world temperature rises by 0.3°C (0.5°F) per decade, during the next 60 years soils will release an amount of CO_2 equal to nearly 20 percent of the projected CO_2 release due to combustion *of* fossil fuels, assuming the present rate of fuel consumption continues.

Breakdown of Gas Hydrates *Gas hydrates* are ice-like solids in which gas molecules, mainly methane, are locked up in water trapped in ocean sediments and beneath frozen ground. By one estimate, worldwide gas hydrates may hold 10,000 billion metric tons of carbon, twice the amount in all the known coal, gas, and oil reserves on land. They accumulate in ocean sediments beneath a water depth of 500 m (1640 ft), where the temperature is low enough and the pressure high enough to permit their formation. They also accumulate beneath permafrost, which acts as a seal to prevent upward migration and escape of the gas. When gas hydrates break down, they release methane. Global warming at high latitudes will thaw frozen ground, and this thawing may destabilize the hydrates there, releasing large volumes of methane and thus amplifying the greenhouse effect.

LESSONS FROM THE PAST

Examining changes to physical and biological systems that occurred when human influence was absent or minimal allows us to see how these systems responded to natural environmental change. Of particular interest are episodes of rapid or abrupt changes of climate—those occurring within a century or less—which may provide analogues of future climatic

A Closer Look

The Younger Dryas Event and the End of the Last Ice Age

At the end of the last glaciation, about 11,000 to 10,000 radiocarbon years ago, the climate in the North Atlantic and adjacent lands experienced a rapid and remarkable change. For 2000 years, the climate had been warming, causing ice sheets in North America and Europe to retreat and allowing plants and animals to reoccupy the deglaciated landscape. Many mountain glaciers in Britain and Scandinavia had disappeared, and the southern limit of sea ice in the North Atlantic had shifted far north, close to its present limit. By all indications, the glacial age was drawing to a rapid close. Then, very abruptly, the climate cooled. Water temperatures in the North Atlantic fell as the southern limit of polar water shifted southward, nearly to its full-glacial extent. The retreating ice sheets halted, then readvanced, and mountain glaciers were reborn in formerly ice-free cirques. Forests in northwestern Europe were rapidly replaced by low-growing herbaceous plants typical of full-glacial conditions. Among these plants was a distinctive flowering species, *Dryas octopetala,* now limited to polar latitudes and high altitudes. Pollen of this species is found abundantly in organic deposits dating to this interval and has provided the name used to identify this cold episode—the Younger Dryas. From oxygen-isotope data obtained from Swiss lake sediments and a Greenland ice core, we can see that the onset of the Younger Dryas episode was rapid (Fig. C18.1). These records also show that the event terminated equally rapidly. In fact, the ice core indicates that the climate over Greenland warmed about 7°C (12.6°F) in only 40 years, a rate that exceeds even the unusually rapid average global rate of temperature rise that climate models project for the coming century.

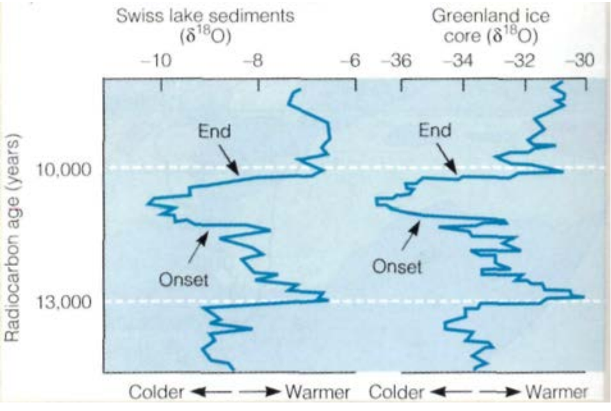


Figure C18.1 Measurements of oxygen isotopes in the sediments of a Swiss lake and an ice core from the Greenland Ice Sheet show an abrupt and rapid change of climate at the end of the last glaciation (arrows). The curves, which can be viewed as recording changes in temperature, show a sudden shift to colder climate followed by an abrupt return to warmer postglacial climate. Detailed studies of the ice core indicate that at the end of the Younger Dryas event average temperature in Greenland rose about 7 °C in only 40 years.

The effects of Younger Dryas cooling are most pronounced around the North Atlantic, and so the search for a cause has focused on this region. As pieces of the puzzle have been assembled, it has become clear that the solution likely lies in interactions of the Earth's natural systems: the cryosphere, the oceans, the atmosphere, and the biosphere.

Glacial-geologic studies have shown that as the ice sheet over eastern North America retreated, vast meltwater lakes were ponded beyond the glacier margin. When

change. We can examine rapid shifts in climate which occurred at the end of the last glacial age (see "A Closer Look: The Younger Dryas Event and the End of the Last Ice Age") and during the Little Ice Age. We can also search the geologic record for information about times when the climate was warmer than now. Because we may be on the brink of a warmer world, times of greatest interest are the early Holocene, from 10,000 to about 6000 years ago, when average temperatures were 0.5° to 1°C (0.9 to 1.8°F) warmer; the warmest part of the last interglaciation, about 120,000 years ago, when global temperatures were about 1° to 2°C (2 to 4°F) higher; and the Middle Pliocene, about 4.5 to 3 million years ago, when the Earth's climates may have been 3° to 4°C (5 to 7°F) warmer than present. These periods do not provide perfect analogs for present global warming because the distribution of solar radiation reaching the Earth's surface was different then (e.g., Figs. 14.24 and 14.25). Nevertheless, they enable us to see how plants and animals responded to climatic conditions that could have been broadly similar to those we may experience in the near future.

During the warmest parts of the Pliocene Epoch, the retreating ice uncovered a natural drainageway between these lakes and the North Atlantic, meltwater flowed rapidly into the ocean where it formed a freshwater lid over the denser salty marine water. The cold surface meltwater reduced evaporation from the ocean surface, thereby shutting down the ocean's thermohaline circulation system (Chapter 8). Eastward-flowing air masses traveling across the colder North Atlantic moved across Western Europe, bringing on the Younger Dryas cooling (Fig. CI 8.2). As the huge meltwater lakes drained and meltwater flow from the ice sheet eventually slowed, the ocean circulation system resumed, bringing warmer climate to the North Atlantic region and heralding the rapid termination of the ice age.

As we learn more about this remarkable natural climatic event and see how the Earth's physical and biological systems reacted to it, important insights will be gained that may help us anticipate future environmental changes resulting from a rapidly warming world.

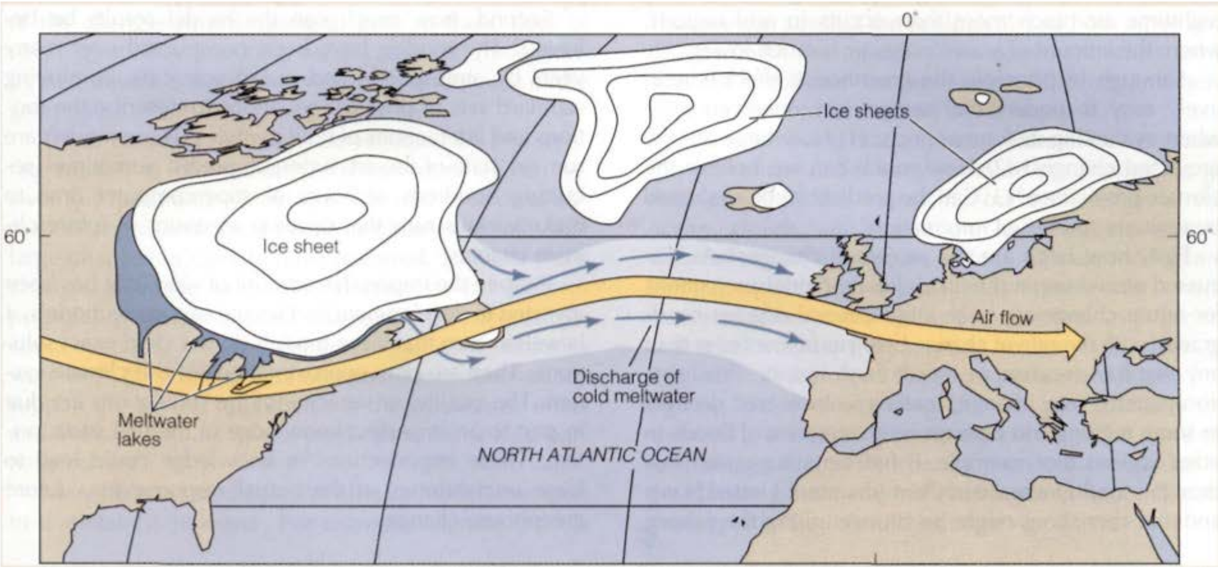


Figure C18.2 Distribution of ice sheets in the North Atlantic region and North American ice-margin lakes during the Younger Dryas event. Rapid drainage of large volumes of meltwater into the western North Atlantic cooled the ocean surface and reduced its salinity, shutting down the thermohaline conveyor system. Air passing over the cold North Atlantic brought colder conditions to northwestern Europe that led to the growth of glaciers and a major change in vegetation communities.

for example, world sea level was tens of meters higher than now, pointing to reduced global ice cover. Isotopic measurements of North Atlantic deep-sea cores show that winter sea-surface temperatures were at least 3°C (5°F) warmer than at present. In response to warmer ocean waters, temperature-sensitive marine organisms had different distributions. The boundaries between tropical, subtropical, and temperate assemblages of ostracodes, a type of crustacean, shifted northward along the eastern coast of the United States during the Middle Pliocene in response to warmer coastal waters, while in China a warm, moist climate produced deep weathering profiles and tropical to subtropical soils.

PERSPECTIVES OF GLOBAL CHANGE

Despite considerable research, we do not yet have a clear vision of our climatic future. At present, the best we can conclude is that the force of scientific evidence and theory makes it very probable that the cli-

Guest Essay

Assessing Projections for Future Climate Change

Thomas J. Crowley is a research scientist with the Applied Research Corporation in College Station, Texas. He specializes in the reconstruction of past climates and the use of climate and ocean models to interpret these changes.

Much has been written about the potential climate change that could result from the greenhouse effect. Continued emissions of carbon dioxide and other trace gases (such as methane and freons) should increase the trapping of radiation from within the earth, resulting in warming of the atmosphere. A simple examination of this effect occurs on warm humid summer evenings, when water vapor (another greenhouse gas) warms the nighttime air much more than occurs in arid regions, where the amount of water in the air is much lower.

Although in principle the greenhouse effect is relatively easy to understand, several key questions arise when evaluating its future effects: (1) How large are the projected changes? (2) How much can we believe the climate projections? (3) Can the predictions be evaluated by separate sources of information?

First, how large are the projected changes? As discussed elsewhere in the text, climate model projections for future change are quite large (several degrees centigrade), with the rate of change being as fast or faster than any that has occurred in recent earth history. Precipitation patterns may change, leading to increased drought in some regions and perhaps more storms and floods in other regions. For example, it has been suggested that droughts could increase in the midwestern United States and that hurricanes might be stronger along the eastern seaboard. But these regional predictions must be treated with greater caution (see below).

Second, how much can the model results be believed? The models have been constructed over many years by atmospheric and ocean scientists, employing standard sets of physical equations to describe the motions and interactions of fluid particles. These models are run on state-of-the-art supercomputers, sometimes requiring hundreds of hours of supercomputer time to make various runs that develop scenarios of future climate change.

Despite the impressive amount of work that has been invested in development and testing of climate models, it is well known that these models do not yield exact solutions. They are only approximations of the climate system. The quality, or veracity, of the simulations are due in part to an imperfect knowledge of the real earth system. These imperfections in knowledge could lead to large uncertainties in the actual response to a future greenhouse change.

mate is warming up and will continue to warm as we add greenhouse gases to the atmosphere. There also is a high probability that average global temperatures ultimately will increase by 2 to 4°C (4 to 7°F), leading to widespread environmental changes.

It is less probable that the temperature will increase steadily, for there are natural, and as yet largely unpredictable, modulations of the climate system on the time scale of years to decades. As an example, the huge Tambora eruption in the East Indies in 1815 was followed by the "year without a summer," during which midsummer snow and frost caused severe hardships in Europe and New England (Fig. 18.17). Other big eruptions, such as Krakatau (Indonesia), Katmai (Alaska), and Agung (Bali), also produced detectable atmospheric cooling. The eruption of the Philippine volcano Pinatubo in June 1991 (Chapter 5) introduced a vast quantity of fine ash and sulfurous gas into the stratosphere, where it quickly began to spread into the northern and southern hemispheres. As the veil of dust and gas spread throughout the atmosphere, it reflected incoming solar radiation, thereby reducing surface temperatures. Such a volcanic event may well reverse temporarily any upward trend in average global temperature attributable to continued emission of greenhouse gases.

While the short-term prospect (on the scale of human generations) is for a warmer world, if we stand back and look at our great geochemical "experiment" from a geological perspective, we will perceive that it is only a brief, very rapid, yet nonrepeatable perturbation in the Earth's climatic history. It is nonrepeatable because once the Earth's store of easily extractable fossil fuels is used up, most likely within the next several hundred years, the human impact on the atmosphere will inevitably decline and the climate system

Some of the most significant uncertainties in the design of climate models involve formulations of physical processes describing the formation of clouds, the ocean circulation, and the interaction between vegetation and the changing climate. Each of these processes can affect climate—clouds greatly affect the amount of heating and cooling, the ocean circulation affects the amount of heat that is transported into a region, and vegetation cover affects the amount of moisture released to the atmosphere (which in turn affects precipitation).

In addition to uncertainties in formulation of detailed physical processes, there are computational limitations for making model forecasts. Despite the many hours spent making supercomputer runs, many thousands more will be required to explore all the potential effects of the changes in environment. Even in these days of very high speed computing, we still have computing limits that prevent a more comprehensive assessment of the greenhouse effect.

Third, can the predictions be validated by any separate sources of information? To some extent we can answer this question by looking at the past records of climate change. Geological evidence clearly indicates that large changes in climate have occurred. Some of these changes coincide with natural (that is, not anthropogenic, or human-caused) changes in greenhouse gas concentrations. When greenhouse gas concentrations were high, the climate was warmer. These observations support the general importance of the greenhouse effect and indicate that the model predictions are at least partially correct.

More detailed predictions of climate models are open to uncertainty, however. For example, climate models predict that tropical oceans will get warmer and that, although winters will be warmer in arctic regions, subfreezing temperatures will still result. However, geological data do not provide much support for the scenario of warming tropical ocean temperatures. Conversely, geological data indicate that arctic regions were warmer in winter than predicated by the models. One explanation of this dilemma may be that ocean currents are transporting substantially more heat than predicted by the models, thereby cooling the tropics and warming the poles. Whatever the explanation, these results suggest that, although on the largest scale increasing greenhouse gas concentrations results in warming, the regional response to the greenhouse gases could differ from model predictions (for example, less drought than predicted). Thus we can say that the geological data both agree and disagree with the models. This agreement lends some level of credibility to the models; the disagreement means that room for caution is still required.

Efforts are being made in many research areas to narrow the uncertainties in greenhouse predictions. Large field programs are increasing our knowledge of the ocean, clouds, and vegetation, and their interactions with the rest of the climate system. Climate models are being constantly improved. Supercomputer time continues to increase in availability and new and faster generations of computers are coming on line every three to five years. Research on the geological record is also improving our understanding of how past climates responded to increased greenhouse forcing. All of these efforts suggest that in the coming years our knowledge of the climate system and our confidence in future climate model predictions should increase significantly.

should return to its natural state. The greenhouse perturbation may well last a thousand years, and perhaps more, but ultimately the changing geometry of the Earth's orbit will propel the climate system into the next glacial age (Fig. 18.15).

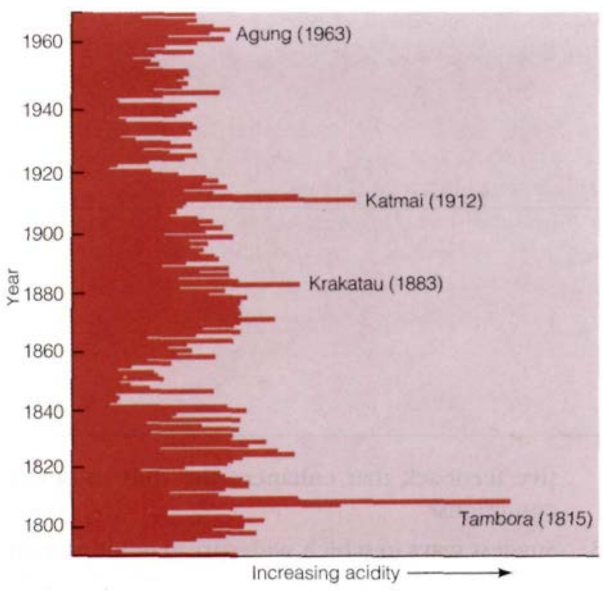


Figure 18.17 Acidity record from a Greenland ice core showing peaks in sulphuric-acid precipitation attributable to major volcanic eruptions. The largest acid peak dates to 1815-1816, the time of the huge Tambora eruption in the East Indies, which produced "the year without a summer" (1816), as the volcanic dust and gases in the stratosphere reduced northern hemisphere temperatures at least 0.7 °C. Subsequent eruptions of Krakatau, Katmai, and Agung also produced detectable climatic effects of smaller magnitude.

1. Synthetic chlorofluorocarbon (CFC) gases entering the upper atmosphere break down to chlorine, which destroys the protective ozone layer. Discovery of a vast ozone hole over Antarctica has led to international efforts to eliminate CFC production by the end of the century.
2. An expanding human population has led to global exploitation of natural resources and serious environmental deterioration. Human and natural geologic activities are intimately interlinked, and people have increasingly become a major agent of geologic change.
3. Changes affecting the Earth's climate system operate on time scales ranging from decades to millions of years.
4. Human activities have produced unanticipated poisoning of natural ecosystems, helped promote desertification in arid and semi-arid lands, and substantially reduced the extent of natural forests. In the process, complex positive and negative feedbacks have amplified the changes wrought by people.
5. The carbon cycle is among the most important of the Earth's biogeochemical cycles. Carbon resides in the atmosphere, the biosphere, the hydrosphere, and in the crust, and it cycles through these reservoirs at different rates.
6. The anthropogenic extraction and burning of fossil fuels perturbs the natural carbon cycle and has led to an increase in atmospheric CO_2 since the start of the Industrial Revolution about 1850.
7. The greenhouse effect, due to the trapping of long-wave infrared radiation by trace gases in the atmosphere, makes the Earth a habitable planet.
8. The increase in atmospheric trace gases (CO_2, CH_4, O_3, N_2O, and the CFCs) due to human activities is projected to warm the lower atmosphere between 2 and 4°C by the end of the next century. A probable 0.5°C increase in average global temperature since the mid-nineteenth century may reflect the initial part of this warming. The rate of warming may reach 0.3°C per decade and could lead to a "super interglaciation," making the Earth warmer than at any time in human history.
9. Potential physical and biological consequences of global warming include global changes in precipitation and vegetation patterns, melting of glaciers, sea ice, and frozen ground, a worldwide rise of sea level, increased rates of organic decomposition in soils, and breakdown of gas hydrates trapped beneath high-latitude permafrost.
10. Using information from the geologic record, geologists can determine the magnitude and geographic extent of past climatic changes, determine the range of climatic variability on different time scales, and test the accuracy of computer models that simulate past climatic conditions and provide insights into physical and biological responses to future global warming.
11. Although the Earth's surface environments may change substantially during the next several centuries in response to greenhouse warming, viewed from the geologic perspective, this interval will appear as only a brief perturbation in the Earth's climatic history.

Important Terms to Remember

deforestation (p. 476)
desertification (p. 474)
greenhouse effect (p. 479)

1. Why does chlorine have such an adverse affect on the ozone layer, despite the fact that it is released to the atmosphere in very small amounts?
2. How might desertification brought on by overgrazing in a semi-arid rangeland produce a positive feedback that enhances the shift to desert conditions?
3. Suggest ways in which widespread deforestation can affect (a) streams, (b) soils, and (c) local climate.

4. In what ways can carbon be trapped in the Earth and become part of the rock cycle? How can such stored carbon once again find its way into the atmosphere?
5. If atmospheric CO_2 can be dissolved in streams, lakes, groundwater, and the oceans, and also efficiently absorbed by vegetation, suggest why the burning of fossil fuels is causing the CO_2 content of the atmosphere to increase.
6. How is the Earth's atmosphere similar to a garden greenhouse, and why?
7. What are the anthropogenic sources of the principal greenhouse gases?
8. What geologic evidence indicates that the present concentrations of carbon dioxide and methane in the atmosphere are exceptional compared to those of the last several hundred thousand years?
9. Give an example of an environmental effect arising from global warming that could enhance the greenhouse effect and lead to additional warming.
10. What factors are likely to cause world sea level to rise in a warming climate? In what ways is rising sea level likely to impact the human population?
11. Why is the geologic record important in helping predict the environmental effects of greenhouse warming? Give two examples.

Questions for A Closer Look

1. What physical and biological evidence points to a sudden and rapid cooling in the North Atlantic region toward the end of the last glaciation?
2. How is ocean circulation involved in explaining the cause of the Younger Dryas cooling?

Questions for Discussion

1. Describe the carbon cycle. Why do we regard it as one of the most important biogeochemical cycles? In what ways are you, as an individual human, involved in the carbon cycle?
2. Why does uncertainty exist about the extent to which average global temperature will rise in the next century as a result of greenhouse warming?
3. Based on what you have learned about potential future greenhouse warming, what are some of the possible changes that could affect your community during the next 50 years if the average world climate warms by as much as 2°C?

APPENDIX A

Units and Their Conversions

ABOUT SI UNITS

Regardless of the field of specialization, all scientists use the same units and scales of measurement. They do so to avoid confusion and the possibility that mistakes can creep in when data are converted from one system of units, or one scale, to another. By international agreement the SI units are used by all, and they are the units used in this text. SI is the abbreviation of Systeme International d'Unites (in English, the International System of Units).

Some of the SI units are likely to be familiar, some unfamiliar. The SI unit of length is the meter (m), of area the square meter (m^2), and of volume the cubic meter (m^3). The SI unit of mass is the kilogram (kg), and of time the second (s). The other SI units used in this book can be defined in terms of these basic units. Three important ones are:

The newton (N), a unit of force defined as that force needed to accelerate a mass of 1 kg by 1 m/s^2; hence 1 N = 1 $kg{\cdot}m/s^2$. (The period between kg and m indicates multiplication.)

The joule (J), a unit of energy or work, defined as the work done when a force of 1 newton is displaced a distance of 1 meter; hence 1 J = 1 N·m. One important form of energy so far as the Earth is concerned is heat. The outward flow of the Earths internal heat is measured in terms of the number of joules flowing outward from each square centimeter each second; thus, the unit of heat flow is $J/cm^2/s$.

The pascal (Pa), a unit of pressure defined as a force of 1 newton applied across an area of 1 square meter; hence 1 Pa = 1 N/m^2. The pascal is a numerically small unit. Atmospheric pressure, for example (15 lb/in^2), is 101,300 Pa. Pressure within the Earth reaches millions or billions of pascals. For convenience, earth scientists sometimes use 1 million pascals (megapascal, or MPa) as a unit.

Temperature is a measure of the internal kinetic energy (expressed as movement) of the atoms and molecules in a body. In the SI system, temperature is measured on the Kelvin scale (K). The temperature intervals on the Kelvin scale are arbitrary, and they are the same as the intervals on the more familiar Celsius scale (°C). The difference between the two scales is that the Celsius scale selects 100°C as the temperature at which water boils at sea level, and 0°C as the freezing temperature of water at sea level. Zero degrees Kelvin, on the other hand, is absolute zero, the temperature at which all atomic and molecular motions cease. Thus, 0°C is equal to 273.15 K, and 100°C is 373.15 K. The temperatures of processes on and within the Earth tend to be at or above 273.15 K. Despite the inconsistency, earth scientists still use the Celsius scale when geological processes are discussed.

Appendix A provides a table of conversion from older units to Standard International (SI) units.

PREFIXES FOR MULTIPLES AND SUBMULTIPLES

When very large or very small numbers have to be expressed, a standard set of prefixes is used in conjunction with the SI units. Some prefixes are probably already familiar; an example is the centimeter (which is one hundredth of a meter, or 10^{2} m). The standard prefixes are

giga	1,000,000,000	=	10^9
mega	1,000,000	=	10^6
kilo	1,000	=	10^3
hecto	100	=	10^2
deka	10	=	10
deci	0.1	=	10^{-1}
centi	0.01	=	10^{-2}

milli	0.001	=	10^{-3}
micro	0.000001	=	10^{-6}
nano	0.000000001	=	10^{-9}
pico	0.000000000001	=	10^{-12}

One measure used commonly in geology is the nanometer (nm), a unit by which the sizes of atoms are measured; 1 nanometer is equal to 10^{-9} meter.

COMMONLY USED UNITS OF MEASURE

Length

Metric Measure

1 kilometer (km)	=	1000 meters (m)
1 meter (m)	=	100 centimeters (cm)
1 centimeter (cm)	=	10 millimeters (mm)
1 millimeter (mm)	=	1000 micrometers (urn) (formerly called microns)
1 micrometer (urn)	=	0.001 millimeter (mm)
1 angstrom (A)	=	10^{-8} centimeters (cm)

Nonmetric Measure

1 mile (mi)	=	5280 feet (ft) = 1760 yards (yd)
1 yard (yd)	=	3 feet (ft)
1 fathom (fath)	=	6 feet (ft)

Conversions

1 kilometer (km)	=	0.6214 mile (mi)
1 meter (m)	=	1.094 yards (yd) = 3.281 feet (ft)
1 centimeter (cm)	=	0.3937 inch (in)
1 millimeter (mm)	=	0.0394 inch (in)
1 mile (mi)	=	1.609 kilometers (km)
1 yard (yd)	=	0.9144 meter (m)
1 foot (ft)	=	0.3048 meter (m)
1 inch (in)	=	2.54 centimeters (cm)
1 inch (in)	=	25.4 millimeters (mm)
1 fathom (fath)	=	1.8288 meters (m)

Area

Metric Measure

1 square kilometers (km^2)	=	1,000,000 square meters (m^2)
	=	100 hectares (ha)
1 square meter (m^2)	=	10,000 square centimeters (cm^2)
1 hectare (ha)	=	10,000 square meters (m^2)

Nonmetric Measure

1 square mile (mi^2)	=	640 cares (ac)
1 acre (ac)	=	4840 square yards (yd^2)
1 square foot (ft^2)	=	144 square inches (in^2)

Conversions

1 square kilometer (km^2)	=	0.386 square mile (mi^2)
1 hectare (ha)	=	2.471 acres (ac)
1 square meter (m^2)	=	1.196 square yards (yd^2)
	=	10.764 square feet (ft^2)
1 square centimeter (cm^2)	=	0.155 square inch (in^2)
1 square mile (mi^2)	=	2.59 square kilometers (km^2)
1 acre (ac)	=	0.4047 hectare (ha)
1 square yard (yd^2)	=	0.836 square meter (m^2)
1 square foot (ft^2)	=	0.0929 square meter (m^2)
1 square inch (in^2)	=	6.4516 square centimeter (cm^2)

Volume

Metric Measure

1 cubic meter (m^3)	=	1,000,000 cubic centimeters (cm^3)
1 liter (1)	=	1000 milliliters (ml)
	=	0.001 cubic meter (m^5)
1 centiliter (cl)	=	10 milliliters (ml)
1 milliliter (ml)	=	1 cubic centimeter (cm^2)

Nonmetric Measure

1 cubic yard (yd^3)	=	27 cubic feet (ft^3)
1 cubic foot (ft^3)	=	1728 cubic inches (in^3)
1 barrel (oil) (bbl)	=	42 gallons (U.S.) (gal)

Conversions

1 cubic kilometer (km^3)	=	0.24 cubic miles (mi^3)
1 cubic meter (m^3)	=	264.2 gallons (U.S.) (gal)
	=	35.314 cubic feet (ft^3)
1 liter (1)	=	1.057 quarts (U.S.) (qt)
	=	33.815 ounces (U.S. fluid) (fl. oz.)

1 cubic centimeter (cm^3)	=	0.0610 cubic inch (in^3)
1 cubic mile (mi^3)	=	4.168 cubic kilometers (km^3)
1 acre-foot (ac-ft)	=	1233.46 cubic meters (m^3)
1 cubic yard (yd^3)	=	0.7646 cubic meter (m^3)
1 cubic foot (ft^3)	=	0.0283 cubic meter (m^3)
1 cubic inch (in^3)	=	16.39 cubic centimeters (cm^3)
1 gallon (gal)	=	3.784 liters (1)

Mass

Metric Measure

1000 kilograms (kg)	=	1 metric ton (also called a tonne) (m.t)
1 kilogram (kg)	=	1000 grams (g)

Nonmetric Measure

1 short ton (sh.t)	=	2000 pounds (lb)
1 long ton (l.t)	=	2240 pounds (lb)
1 pound (avoirdupois) (lb.)	=	16 ounces (avoirdupois) (oz) = 7000 grains (gr)
1 ounce (avoirdupois) (oz)	=	437.5 grains (gr)
1 pound (Troy) (Tr. lb)	=	12 ounces (Troy) (Tr. oz)
1 ounce (Troy) (Tr. oz)	=	20 pennyweight (dwt)

Conversions

1 metric ton (m.t)	=	2205 pounds (avoirdupois) (lb)
1 kilogram (kg)	-	2.205 pounds (avoirdupois) (lb)
1 gram (g)	=	0.03527 ounce (avoirdupois) (oz) 0.03215 ounce (Troy) (Tr. oz) = 15,432 grains (gr)
1 pound (lb)	=	0.4536 kilogram (kg)
1 ounce (avoirdupois) (oz)	=	28.35 grams (g)
1 ounce (avoirdupois) (oz)	=	1.097 ounces (Troy) (Tr. oz)

Pressure

1 pascal (Pa)	=	1 newton/square meter (N/m^2)
1 kilogram/square centimeter (kg/cm^2)	=	0.96784 atmosphere (atm) = 14.2233 pounds/square inch (lb/in^2) = 0.098067 bar
1 bar	-	0.98692 atmosphere (atm) = 10^5 pascals (Pa) = 1.02 kilograms/square centimeter (kg/cm^2)

Energy and Power

Energy

1 joule (J)	=	1 newton meter (N.m)
	=	2.390 X 10 ' calorie (cal)
	=	9-47 X 10^{-4} British thermal unit (Btu)
	=	2.78 X 10^{-7} kilowatt-hour (kWh)
1 calorie (cal)	=	4.184 joule (J)
	=	3-968 X 10^{-3} British thermal unit (Btu)
	=	1.16 X 10^{-6} kilowatt-hour (kWh)
1 British thermal unit (Btu)	=	1055.87 joules (J)
	=	252.19 calories (cal)
	=	2.928 X 10^{-4} kilowatt-hour (kWh)
1 kilowatt hour	=	3.6 X 10^6 joules (J)
	=	8.60 X 10^5 calories (cal)
	=	3.41 X 10^3 British thermal units (Btu)

Power (energy per unit time)

1 watt (W)	=	1 joule per second (J/s)
	=	3.4129 Btu/h
	=	1.341 X 10^{-3} horsepower (hp)
	=	14.34 calories per minute (cal/min)
1 horsepower (hp)	=	7.46 X 10^2 watts (W)

Temperature

To change from Fahrenheit (F) to Celsius (C)

$$°C = \frac{(°F-32°)}{1.8}$$

To change from Celsius (C) to Fahrenheit (F)

$$°F = (°C \times 1.8) + 32°$$

To change from Celsius (C) to Kelvin (K)

$K = °C + 273.15$

To change from Fahrenheit (F) to Kelvin (K)

$$K = \frac{(°F - 32)}{1.8} + 273.15$$

APPENDIX B

The Periodic Table of the Elements

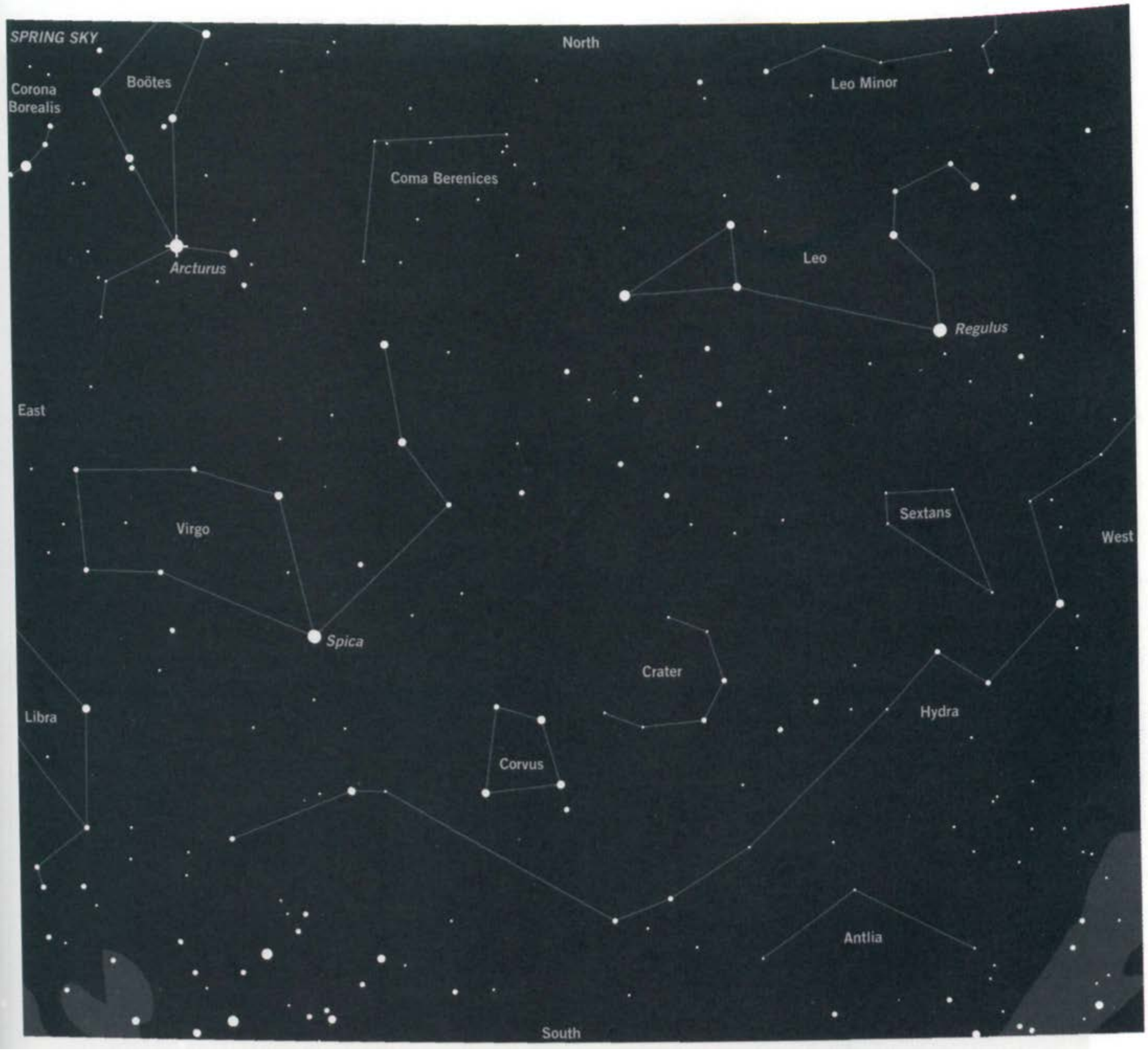
Atomic number: 11 — Name: Sodium — Symbol: **Na** — Approximate atomic weight (to nearest whole number): 23

Period	Group IA	IIA	IIIB	IVB	VB	VIB	VIIB	VIII	VIII	VIII	IB	IIB	IIIA	IVA	VA	VIA	VIIA	VIII
1	1 Hydrogen **H** 1																	2 Helium **He** 4
2	3 Lithium **Li** 7	4 Beryllium **Be** 9											5 Boron **B** 11	6 Carbon **C** 12	7 Nitrogen **N** 14	8 Oxygen **O** 16	9 Flourine **F** 19	10 Neon **Ne** 20
3	11 Sodium **Na** 23	12 Magnesium **Mg** 24											13 Aluminum **Al** 27	14 Silicon **Si** 28	15 Phosphorus **P** 31	16 Sulfur **S** 32	17 Chlorine **Cl** 35	18 Argon **Ar** 40
4	19 Potassium **K** 39	20 Calcium **Ca** 40	21 Scandium **Sc** 45	22 Titanium **Ti** 48	23 Vanadium **V** 51	24 Chromium **Cr** 52	25 Manganese **Mn** 55	26 Iron **Fe** 56	27 Cobalt **Co** 59	28 Nickel **Ni** 59	29 Copper **Cu** 64	30 Zinc **Zn** 65	31 Gallium **Ga** 70	32 Germanium **Ge** 73	33 Arsenic **As** 75	34 Selenium **Se** 79	35 Bromine **Br** 80	36 Krypton **Kr** 84
5	37 Rubidium **Rb** 85	38 Strontium **Sr** 88	39 Yttrium **Y** 89	40 Zirconium **Zr** 91	41 Niobium **Nb** 93	42 Molybdenum **Mo** 96	43 Technetium **Tc** 99	44 Ruthenium **Ru** 101	45 Rhodium **Rh** 103	46 Palladium **Pd** 106	47 Silver **Ag** 108	48 Cadmium **Cd** 112	49 Indium **In** 115	50 Tin **Sn** 119	51 Antimony **Sb** 122	52 Tellurium **Te** 128	53 Iodine **I** 127	54 Xenon **Xe** 131
6	55 Cesium **Cs** 133	56 Barium **Ba** 137	57 *Lanthanum **La** 139	72 Hafnium **Hf** 178	73 Tantalum **Ta** 139	74 Wolfran (tungsten) **W** 184	75 Rhenium **Re** 186	76 Osmium **Os** 190	77 Iridium **Ir** 192	78 Platinum **Pt** 195	79 Gold **Au** 197	80 Mercury **Hg** 201	81 Thallium **Tl** 204	82 Lead **Pb** 207	83 Bismuth **Bi** 209	84 Polonium **Po** 210	85 Astatine **At** 210	86 Radon **Rn** 222
7	87 Francium **Fr** 223	88 Radum **Ra** 226	89 **Actinium **Ac** 227	104 Unnil-quadium **Unq** 266	105 Unil-pentium **Unp** 262	106 Unil-hexium **Unh** 263	107 Unil-septium **Uns** 262	108 Uniloctium **Uno** 265	109 Unilennium **Une** 266									

Series	Period														
*Lanthanide series	6	58 Cerium **Ce** 140	59 Prase-odymium **Pr** 141	60 Neodymium **Nd** 144	61 Promethium **Pm** 147	62 Samarium **Sm** 150	63 Europium **Eu** 152	64 Gadolinium **Gd** 157	65 Terbium **Tb**	66 Dysprosium **Dy** 163	67 Holmium **Ho** 165	68 Erbium **Er** 167	69 Thulium **Tm** 169	70 Ytterbium **Yb** 173	71 Lutetium **Lu** 175
**Actinide series	7	90 Thorium **Th** 232	91 Protactinium **Pa** 231	92 Uranium **U** 238	93 Neptunium **Np** 237	94 Plutonium **Pu** 242	95 Americium **Am** 243	96 Curium **Cm** 247	97 Berkellium **Bk** 247	98 Californium **Cf** 251	99 Einsteinium **Es** 254	100 Fermium **Fm** 253	101 Mendelevium **Md** 256	102 Nobelium **No** 254	103 Lawrencium **Lr** 257

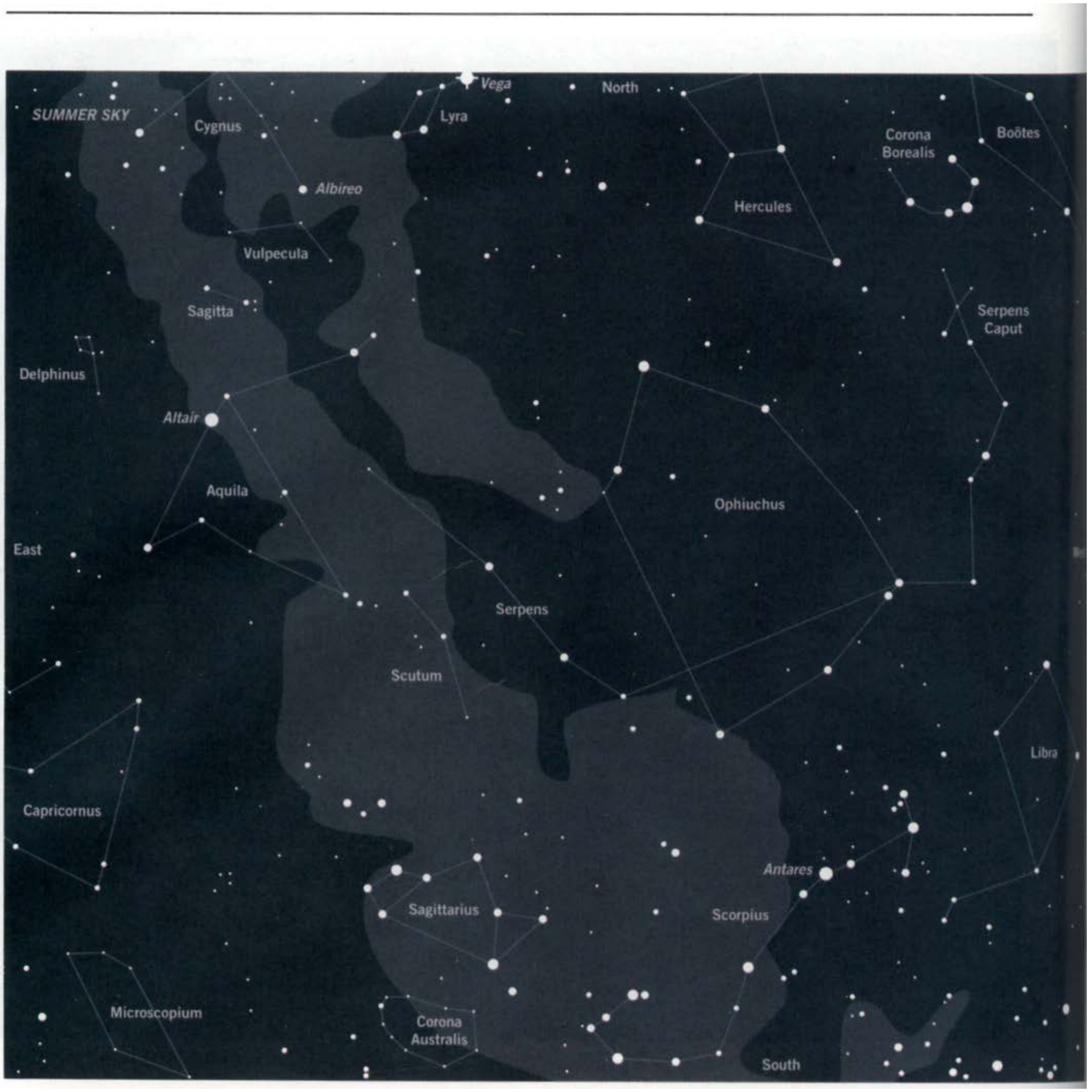
APPENDIX C

Seasonal Star Charts



SPRING SKY
North
Corona Borealis
Boötes
Leo Minor
Coma Berenices
Arcturus
Leo
Regulus
East
Sextans
Virgo
West
Spica
Crater
Libra
Hydra
Corvus
Antlia
South

Seasonal Star Charts *(continued)*




AUTUMN SKY
North
Triangulum
Aries
Pegasus
Pisces
Equuleus
East
Mira
Cetus
Aquarius
West
Eridanus
Capricornus
Fomalhaut
Piscis Austrinus
Fornax
Sculptor
Microscopium
Grus
South
Alnair

Seasonal Star Charts *(continued)*

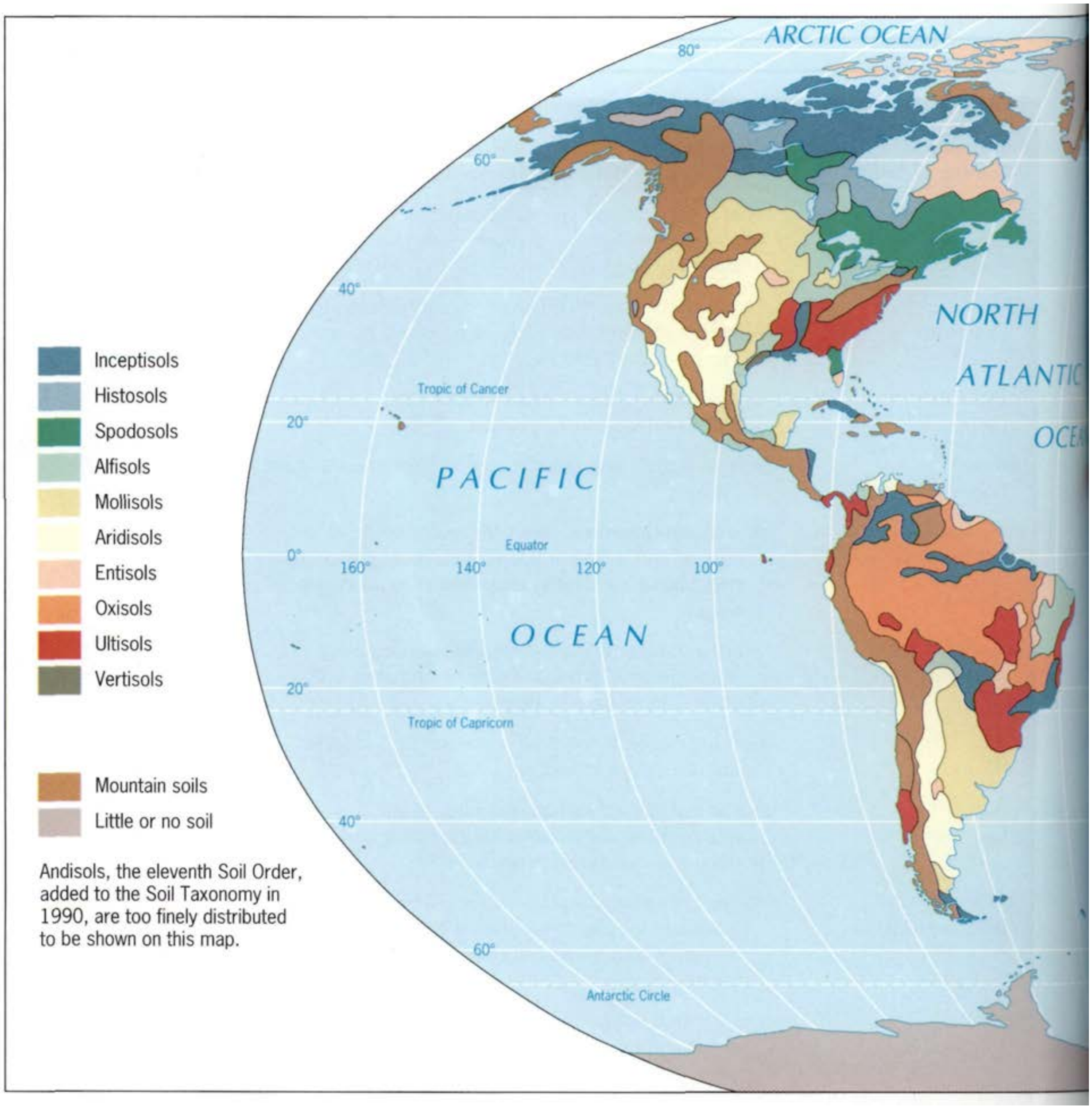
APPENDIX D

World Soils and Soil Classification System

TABLE D-l Orders of the Soil Classification System Used in the United States

Soil Order (Meaning of Name)	Main Characteristics
Alfisol (Pedalfer[a] soil)	Thin A horizon over clay-rich B horizon, in places separated by light-gray E horizon. Typical of humid middle latitudes.
Aridisol (Arid soil)	Thin A horizon above relatively thin B horizon and often with carbonate accumulation in K horizon. Typical of dry climates.
Entisol (Recent soil)	Soil lacking well-developed horizons. Only a thin incipient A horizon may be present.
Histosol (Organic soil)	Peaty soil, rich in organic matter. Typical of cool, moist climates
Inceptisol (Young soil)	Weakly developed soil, but with recognizable A horizon and incipient B horizon lacking clay or iron enrichment. Generally occurs under moist conditions.
Mollisol (Soft soil)	Grassland soil with thick dark A horizon, rich in organic matter. B horizon may be enriched in clay. E and K horizons may be present.
Oxisol (Oxide soil)	Relatively infertile soil with A horizon over oxidized and often thick B horizon.
Spodosol (Ashy soil)	Acidic soil marked by highly organic O and A horizons, an E horizon, and iron/aluminum-rich B horizon. Occurs in cool forest zones.
Ultisol (Ultimate soil)	Strongly weathered soil characterized by A and E horizons over highly weathered and clay-rich B horizon. Characteristic of tropical and subtropical climates.
Vertisol (Inverted soil)	Organic-rich soil having very high content of clays that shrink and expand as moisture varies seasonally.
Andisol (Dark soil)	Soil developed on pyroclastic deposits and characterized by low bulk density and high content of amorphous minerals.

[a] Soils rich in iron and aluminum.


ARCTIC OCEAN
NORTH
ATLANTIC
OCE
PACIFIC
OCEAN
Tropic of Cancer
Equator
Tropic of Capricorn
Antarctic Circle
80°
60°
40°
20°
0°
20°
40°
60°
160°
140°
120°
100°
Inceptisols
Histosols
Spodosols
Alfisols
Mollisols
Aridisols
Entisols
Oxisols
Ultisols
Vertisols
Mountain soils
Little or no soil
Andisols, the eleventh Soil Order, added to the Soil Taxonomy in 1990, are too finely distributed to be shown on this map.


ARCTIC OCEAN
PACIFIC OCEAN
INDIAN OCEAN
SOUTHERN OCEAN
UTH
LANTIC
OCEAN
80°
60°
40°
20°
0°
160°
20°
60°
0°
60°
80°

APPENDIX E

The Köppen Climatic Classification

There are five basic climatic categories in the Köppen system, symbolized as A, B, C, D, and E. These are further subdivided by adding lower-case letters to indicate lesser variations of temperature and moisture within the major groupings.

A. TROPICAL HUMID CLIMATES

Coolest month must be above 18° C (64.4° F).

Af—Tropical Rain Forest (f—feucht or moist) No dry season. Driest month must attain at least 6 cm (2.4 in.) of rainfall.
Aw—Tropical Savanna (w—winter) Winter dry season. At least one month must attain less than 6 cm (2.4 in.) of rainfall.

The folio-wing lower-case letters may be added for clarification in special situations:

m (monsoon)—despite a dry season, total rainfall is so heavy that rain forest vegetation is not impeded.
w'—autumn rainfall maximum.
w''—two dry seasons during a single year.
s (summer)—summer dry season.
i—annual temperature range must be less than 5°C (9° F).
g (Ganges)—hottest month occurs prior to summer solstice.

B. DRY CLIMATES

No specific amount of moisture makes a climate dry. Rather, the rate of evaporation (determined by temperature), relative to the amount of precipitation dictates how dry a climate is in terms of its ability to support plant growth. This is reckoned through the use of formulas that are not included here.

BW (W-wuste or wasteland)—desert.
BS (S—steppe)—semiarid.

The following lower-case letters may be added for clarification in special situations.

h (heiss or hot)—average annual temperature must be above 18° C (64.4°F).
k (kalt or cold)—average annual temperature must be under 18° C (*64.4°* F).
k'—temperature of warmest month must be under 18° C (64.4° F).
s—summer dry season. At last three times as much precipitation in the wettest month as in the driest.
w—winter dry season. At least ten times as much precipitation in the wettest month as in the driest.
n (nebel or fog)—frequent fog.

C. TEMPERATE HUMID CLIMATES

Coldest month average must be below 18° C (64.4° F), but above -3° C (26.6° F). Warmest month average must be above 10° C (50° F).

Cf—no dry season. Driest month must attain at least 3 cm (1.2 in.) of precipitation.
CW—winter dry season. At least ten times as much rain in the wettest month as in the driest.
Cs—summer dry season. At least three times as much rain in the wettest month as in the driest. Driest month must receive less than 3 cm (1.2 in.) of rainfall.

The following lower-case letters may be added for clarification in special situations:

a—hot summer. Warmest month must average above 22° C (71.6° F).
b—cool summer. Warmest month must average below 22° C (71.6° F). At least 4 months above 10° C (50° F).

c—short, cool summer. Less than four months over 10° C (50° F).
i—see A climate.
g—see A climate.
n—see B climate.
x—maximum precipitation in late spring or early summer.

D. COLD HUMID CLIMATES

Coldest month average must be below -3° C (26.6° F). Warmest month average must be above 10° C (50° F).

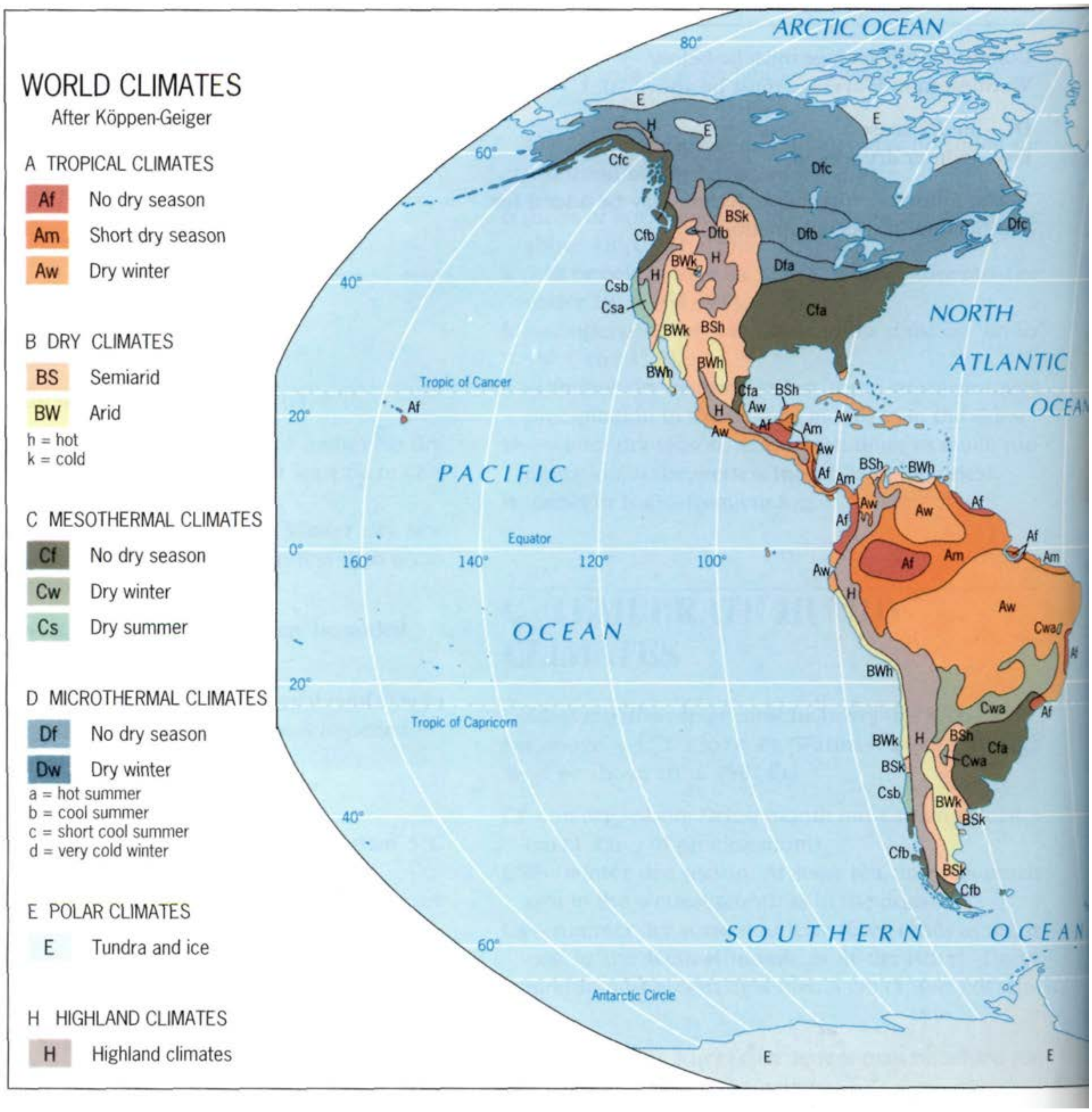
Df—no dry season.
Dw—winter dry season.

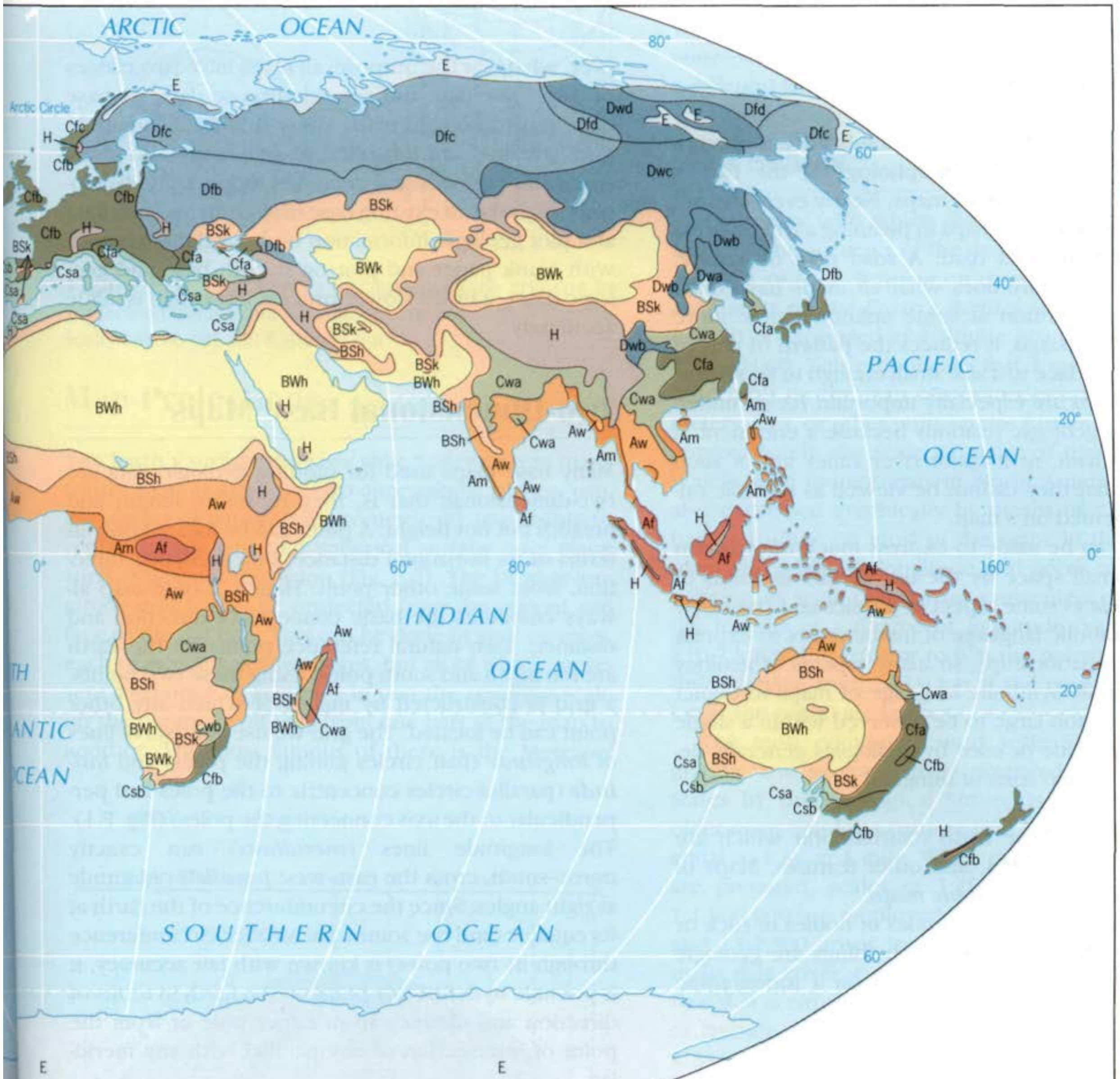
The following lower-case letters may be added for clarification in special situations:

a—see C climate.
b—see C climate.
c—see C climate.
d—coldest month average must be below —38° C (-36.4° F).
f—see A climate,
s—see A climate.
w—see A climate.

E. POLAR CLIMATES

Warmest month average must be below 10° C (50° F).


WORLD CLIMATES
After Köppen-Geiger
A TROPICAL CLIMATES
Af No dry season
Am Short dry season
Aw Dry winter
B DRY CLIMATES
BS Semiarid
BW Arid
h = hot
k = cold
C MESOTHERMAL CLIMATES
Cf No dry season
Cw Dry winter
Cs Dry summer
D MICROTHERMAL CLIMATES
Df No dry season
Dw Dry winter
a = hot summer
b = cool summer
c = short cool summer
d = very cold winter
E POLAR CLIMATES
E Tundra and ice
H HIGHLAND CLIMATES
H Highland climates
ARCTIC OCEAN
NORTH
ATLANTIC
PACIFIC
OCEAN
SOUTHERN
Tropic of Cancer
Equator
Tropic of Capricorn
Antarctic Circle


ARCTIC OCEAN
PACIFIC OCEAN
INDIAN OCEAN
SOUTHERN OCEAN
Arctic Circle
80°
60°
40°
20°
0°
160°
E
Dfc
Dfb
Dwd
Dfd
Dwc
Dwb
Dwa
Cfb
Cfa
Csa
Csb
Cwa
Cwb
BSk
BSh
BWk
BWh
H
Af
Am
Aw

APPENDIX F

Topographic Maps

USES OF MAPS

An important part of the accumulated information about the geology and morphology of the Earth's crust exists in the form of maps. Nearly everyone has used automobile road maps in planning a trip or in following an unmarked road. A road map of a state, province, or county does what all maps have done since their invention at some unknown time more than 5000 years ago; it reduces the pattern of part of the Earth's surface to a size small enough to be seen as a whole. Maps are especially important for an understanding of geologic relations because a continent, a mountain chain, or a major river valley are of such large size that they cannot be viewed as a whole unless represented on a map.

A map can be made to express much information within a small space by the use of various kinds of symbols. Just as some aspect of physics and chemistry use the symbolic language of mathematics to express significant relationships, so many aspects of geology use the simple symbolic language of maps to depict relationships too large to be observed within a single view. Maps made or used by geologists generally depict either of two sorts of things.

1. The shape of the Earth's surface, on which are shown hills, valleys, and other features. Maps of this kind are *topographic maps.*
2. The distribution and attitudes of bodies of rock or regolith. Maps showing such things are *geologic maps.* They are often plotted on a topographic base map.

BASE MAPS

Every map is made for some special purpose. Road maps, charts for sea or air navigation, and geologic maps are examples of three special purposes. However, whatever the purpose, all maps have two classes of data: base data and special-purpose data. As base data, most geologic maps show a latitude-longitude grid, streams, and inhabited places; many also show roads and railroads and details of topography. Geologists may take an existing base map contains such data and plot geologic information on it, or they may start with blank paper and plot on it both base and geologic data—a much slower process if the map is made accurately.

Two-Dimensional Base Maps

Many base maps used for plotting geologic data are two-dimensional; that is, they represent length and breadth but not height. A point can be located only in terms of its horizontal distance, in a particular direction, from some other point. Hence, a base map always embodies the basic concepts of direction and distance. Two natural reference points on the Earth are the north and south poles. Using these two points, a grid is constructed by means of which any other point can be located. The grid we use consists of lines of *longitude* (half circles joining the poles) and *latitude* (parallel circles concentric to the poles and perpendicular to the axis connecting the poles) (Fig. F. 1). The longitude lines (*meridians*) run exactly north-south, cross the east-west *parallels* of latitude at right angles. Since the circumference of the Earth at its equator (and the somewhat smaller circumference through its two poles) is known with fair accuracy, it is possible to define any point on the Earth in terms of direction and distance from either pole or from the point of intersection of any parallel with any meridian.

For convenience in reading, most maps are drawn so that the north direction is at the top or upper edge of the map. This is an arbitrary convention adopted mainly to save time. The north direction could just as well be placed elsewhere, provided its position is clearly indicated.

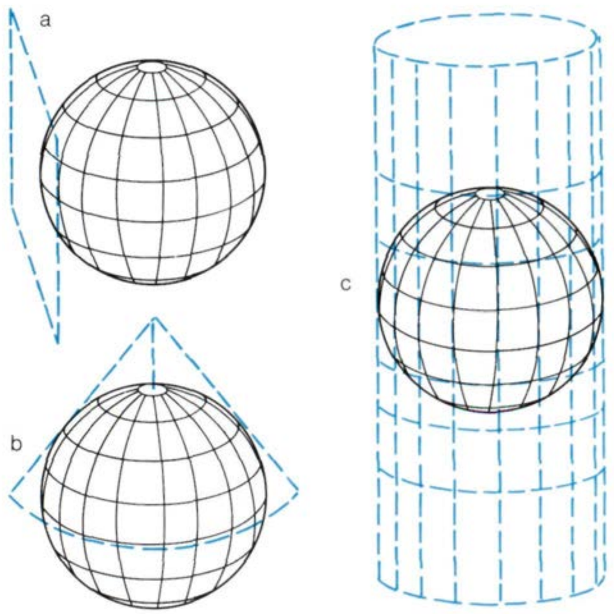


Figure F.1 The Earth's latitude-longitude grid can be projected onto a plane a, cylinder c, or core b that theoretically can be cut and flattened out.

Map Projections

The Earth's surface is nearly spherical, whereas maps (other than globes) are two-dimensional planes, usually sheets of paper. It is geometrically impossible to represent any part of a spherical surface on a plane surface without distortion (Fig. F.2). The latitude-longitude grid has to be projected from the curved surface to the flat one. This can be done in various ways, each of which has advantages, but all of which represent a sacrifice of accuracy in that the resulting scale on the flat map will vary from one part of the map to another. The most famous of these is the Mercator projection, prepared by projecting all points radially onto a cylinder (Fig. F.l*c*), then unfolding the cylinder; although it distorts the polar regions very greatly, compass directions drawn on a Mercator projection are straight lines. Because this is of enormous value in navigation, the Mercator projection is widely used in navigator's charts.

Other kinds of projections are shown in Figure F.l. A commonly used projection for small regions of the Earth's surface is the conic projection (Fig. F.1*b*). Some commonly used varieties are polyconic, in which not one cone, as in Figure F.l*b*, but several cones are employed, each one tangent to the globe at a different latitude. This device reduces distortion.

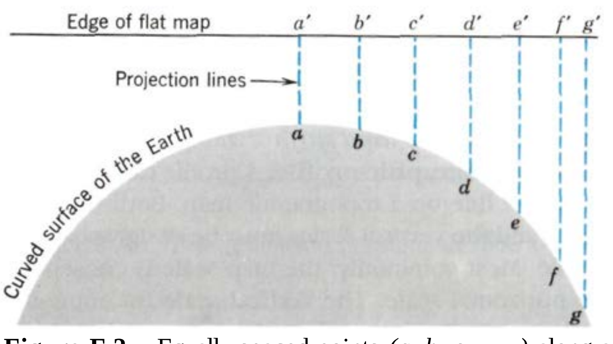


Figure F.2 Equally spaced points *(a, b, c, . . .)* along a line in any direction on the Earth's surface become unequally spaced when projected onto a plane. That is why all flat maps are distorted.

Map Scales

The accuracy with which distance is represented determines the accuracy of the map. *The proportion between a unit of distance on a map and the unit it represents on the Earth's surface* is the *scale* of the map. It is expressed as a simple proportion, such as 1:1,000,000. This ratio means that 1 meter, 1 foot, or other unit on the map represents exactly 1,000,000 meters, feet, or other units on the Earth's surface. It is approximately the scale of many of the road maps widely used by motorists in North America. Scale is also expressed graphically by means of a numbered bar, as is done on most of the maps in this book. A map with a latitude-longitude grid needs no other indication of scale (except for convenience), because the lengths of a degree of longitude (varying from 110.7 km at the equator to 0 at the poles) and of latitude (varying from 109.9 km at the equator to 110.9 km at the poles) are known.

The most commonly used scale for both topographic and geologic maps prepared in the United States by the Geological Survey is 1:24,000. Many older maps employ a scale of 1:62,500, approximately equal to 1 in. = 1 mile. When maps of larger regions are prepared, scales of 1:100,000, 1:250,000, and 1:1,000,000 are employed. Use of scales at 1:24,000 and 1:62,500 arises from the practice of preparing maps that cover a quadrangular segment of the surface that is either 15 minutes (15') of longitude by 15' of latitude (scale, 1:62,500), or 7 1/2' X 7 1/2' (scale, 1:24,000).

Contours and Topographic Maps

A more complete kind of base map is three-dimensional; it represents not only length and breadth but also height. Therefore, it shows **relief** *(the difference* *in altitude between the high and low parts of a land surface)* and also **topography**, defined as *the relief and form of the land. A map that shows topography* is a **topographic map.** Topographic maps can give the form of the land in various ways. The maps most commonly used by geologists show it by contour lines.

A **contour line** (often called simply a **contour**) is *a line passing through points having the same altitude above sea level.* If we start at a certain altitude on an irregular surface and walk in such a way as to go neither uphill nor downhill, we will trace out a path that corresponds to a contour line. Such a path will curve around hills, bend upstream in valleys, and swing outward around ridges. Viewed broadly, every contour must be a closed line, just as the shoreline of an island or of a continent returns upon itself, however long it may be. Even on maps of small areas, many contours are closed lines, such as those at or near the tops of hills. Many, however, do not close within a given map area; they extend to the edges of the map and join the contours on adjacent maps.

Imagine an island in the sea crowned by two prominent isolated hills, with much steeper slopes on one side than on the other and with an irregular shoreline. The shoreline is a contour line (the zero contour) because the surface of the water is horizontal. If the island is pictured as submerged until only the two isolated peaks project above the sea, and then raised above the sea 5 m at a time, the successive new shorelines will form a series of contour lines separated by 5 m contour intervals. (A **contour interval** is *the vertical distance between two successive contour lines* and is commonly the same throughout any one map.) At first, two small islands will appear, each with its own shoreline, and the contours marking their shorelines will have the form of two closed lines. When the main mass of the island rises above the water, the remaining shorelines or contours will pass completely around the landmass. The final shoreline is represented by the zero contour, which now forms the lowest of a series of contours separated by vertical distances of 5 m.

The following rules apply to contours:

1. All points on a contour have the same *elevation,* (also called *altitude),* which is *the vertical distance above mean sea level.*
2. A contour separates all points of higher elevation from all points of lower elevation.
3. In order to facilitate reading the contours on a map, certain contours (usually every fifth line) are drawn as a bolder line. Contours are numbered at convenient intervals for ready identification. The numbers are always multiples of the contour interval. For example, contour intervals of 5 m mean that successive contours are drawn at 10, 15, 20 m, etc.
4. Contours do not split or cross over, but at vertical cliffs they merge.
5. Because the contours that represent a depression without an outlet resemble those of an isolated hill, it is necessary to give them a distinctive appearance. Therefore, depression contours are *hatched;* that is, they are marked on the downslope side with short transverse lines called *hatchures.* An example is shown on one contour in Figure *¥3b.* The contour interval employed is the same as in other contours on the same map.
6. Closely spaced contours indicate steep slopes, and widely spaced contours indicate gentle slopes.
7. Contours crossing a valley form a V-shape pointing *up* the valley, while contours on a ridge form a V-shape pointing *down* the ridge.

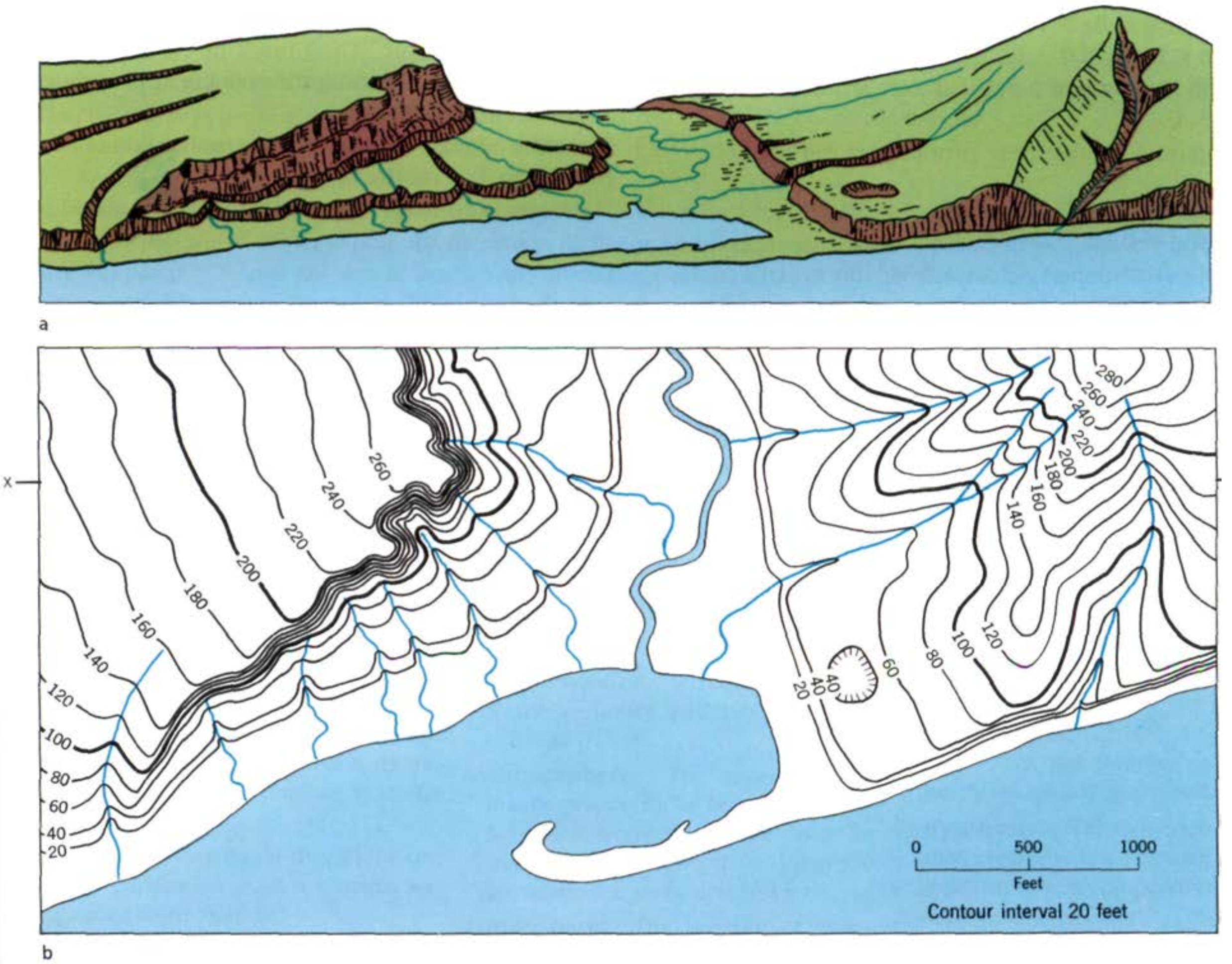
Idealized Example of a Topographic Map

Figures F.3*a* and *b* show the relation between the surface of the land and the contour map representing it. Figure F.3*a*, a perspective sketch, shows a stream valley between two hills, viewed from the south. In the foreground is the sea, with a bay sheltered by a curving spit. Terraces in which small streams have excavated gullies border the valley. The hill on the east has a rounded summit and sloping spurs. Some of the spurs are truncated at their lower ends by a wave-cut cliff, at the base of which is a beach. The hill on the west stands abruptly above the valley with a steep scarp and slopes gently westward, trenched by a few shallow gullies.

Each of the features on the map (Fig. F.3*b*) is represented by contours directly beneath its position in the sketch.

Topographic Profile

The *outline of the land surface along a given line* is called a **topographic profile.** A profile can be drawn along any line on a topographic map. Both the horizontal and the vertical scales must be designated for a profile. Most commonly, the map scale is chosen for the horizontal scale. The vertical scale is commonly made somewhat larger than the horizontal scale in order to exaggerate, or emphasize, the topography. The *ratio of the horizontal scale to the vertical scale in a topographic profile* is called the **vertical exaggeration.** If the horizontal scale is 1 cm = 1000 m,



a

b

Figure F.3 (a) Perspective sketch of a landscape. *(Source:* Modified from U.S. Geological Survey, *(b)* Topographic map of the area shown in Figure F.3*a*. Note that this map is scaled in feet and the contour interval is 20 ft. *(Source:* Modified from U.S. Geological Survey.)

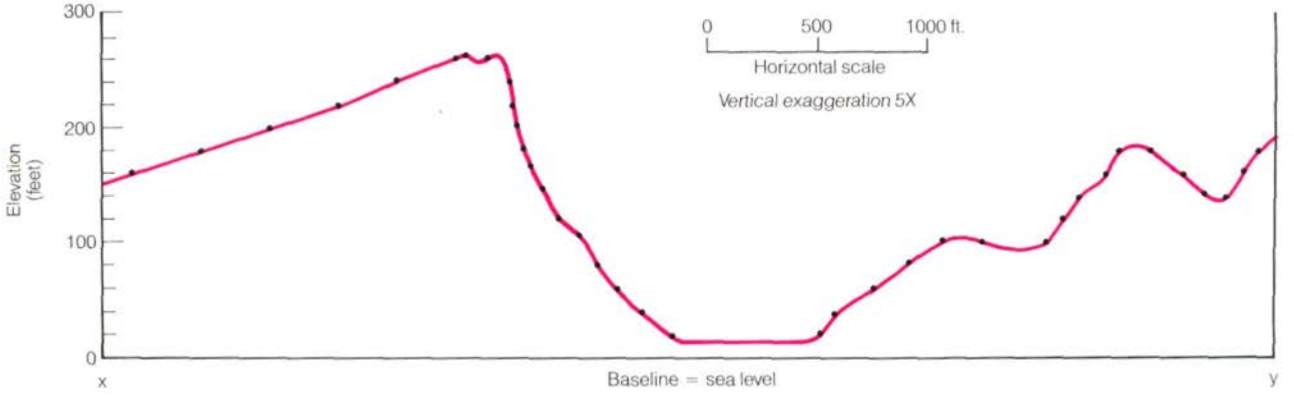


Figure F.4 Topographic profile along the line X-Y in Figure F.3*b*.

and the vertical scale is 1 cm = 200m, the vertical exaggeration is 1000/200 = *5X. A* topographic profile of Figure F.3*b* is shown at a vertical exaggeration of *5X* in Figure F.4.

To prepare a topographic profile perform the following steps:

1. Select the vertical and horizontal scales.
2. On a sheet of paper, select one of the horizontal lines as a baseline. Choice of a base varies from profile to profile; it can be sea level or any convenient height, such as the contour below the lowest point on the profile. Then mark in elevations on the graph paper, choosing the spacing appropriate to the vertical scale.
3. Place the edge of the graph paper on the line of the profile marked on the map.
4. Wherever a contour crosses the line of the profile, make a line on the graph paper at the appropriate elevation.
5. Join the points on the graph paper with a smooth line.

GLOSSARY

Chapter numbers in parentheses at the end of each definition indicate the chapter in which a term is first defined and used as a key term.

Ablation. The loss of mass from a glacier. (Ch. 10)

Abyssal plain. A large flat area of the deep seafloor having slopes less than about 1 m/km, and ranging in depth below sea level from 3 to 6 km. (Ch. 6)

Accreted terrane. Block of crust moved laterally by strike-slip faulting or by a combination of strike-slip faulting and subduction, then accreted to a larger mass of continental crust. Also called a *suspect terrane.* (Ch. 7)

Accumulation. The addition of mass to a glacier. (Ch. 10)

Adiabatic lapse rate. The way temperature changes with altitude in rising or falling air. (Ch. 12)

Adiabatic process. A process that happens without the addition or subtraction of heat from an external source. (Ch. 12)

Aerosol. A tiny liquid droplet or tiny solid particle so small it remains suspended in air. (Ch. 12)

Agglomerate. A pyroclastic rock consisting of bomb-sized tephra, i.e., tephra in which the average particle diameter is greater than *64* mm. (Ch. 5)

Air. The invisible, odorless mixture of gases and suspended particles that surrounds the Earth. (Ch. 12)

Air pressure gradient. The air pressure drop per unit distance. (Ch. 13)

Albedo. The reflectivity of the surface of a planet. (Ch. 10)

Alluvial fan. A fan-shaped body of alluvium typically built where a stream leaves a steep mountain valley. (Ch. 11)

Alluvium. Sediment deposited by streams in nonmarine environments. (Ch. 9)

Amphibolite. A metamorphic rock of intermediate grade, generally coarse-grained, containing abundant amphibole. (Ch. 7)

Amino acid. Organic molecule containing an amino (NH2) group; the building block of proteins. (Ch. 16)

Anaerobic. Without oxygen. (Ch. 15)

Andesite. A fine-grained igneous rock with the composition of a diorite. (Ch. 5)

Angiosperm. A plant whose seeds are surrounded by fruit. (Ch. 16)

Anion. An ion with a negative electrical charge. (Ch. 4)

Anticyclone. Air spiraling outward away from a high-pressure center. (Ch. 13)

Aquifer. A body of permeable rock or regolith saturated with water and through which groundwater moves. (Ch. 9)

Artesian aquifer. An aquifer in which water is under hydraulic pressure. (Ch. 9)

Asthenosphere. The region of the mantle where rocks become ductile, have little strength, and are easily deformed. It lies at a depth of 100 to 350 km below the surface. (Ch. 1)

Atmosphere. The mixture of gases, predominantly nitrogen, oxygen, carbon dioxide, and water vapor that surrounds the Earth. (Introduction)

Atom. The smallest individual particle that retains all the properties of a given chemical element. (Ch. 4)

Autotrophs. Organisms that can get energy directly from sunlight. (Ch. 15)

Barometer. A device that measures air pressure. (Ch. **12)**

Basalt. A fine-grained igneous rock with the composition of a gabbro. (Ch. 5)

Batholith. The largest kind of pluton. A very large, igneous body of irregular shape that cuts across the layering of the rock it intrudes. (Ch. 5)

Beach. Wave-washed sediment along a coast, extending throughout the surf zone. (Ch. 11)

Bed. The smallest formal unit of a body of sediment or sedimentary rock. (Ch. 7)

Bedding. The layered arrangement of strata in a body of sediment or sedimentary rock. (Ch. 7)

Bed load. Coarse particles that move along the bottom of a stream channel. (Ch. 9)

Bergeron process. The evaporation of supercooled water droplets in a cloud to release water vapor that is then deposited on ice crystals within the cloud, leading to precipitation. (Ch. 12)

Biogeochemical cycle. A natural cycle describing the movements and interactions through the Earth's spheres of the chemicals essential to life. (Ch. 16)

Biosphere. The totality of the Earth's organisms and, in addition, organic matter that has not yet been completely decomposed. (Introduction)

Biosynthesis. The polymerization of small organic molecules within a living organism to form biopolymers, particularly proteins. (Ch. 16)

Blackbody radiator. A (hypothetical) perfect radiator of light that absorbs all light that strikes it and reflects none; its light output depends only on its temperature. (Ch. 2)

Body waves. Seismic waves that travel outward from an earthquake focus and pass through the Earth. (Ch. 3)

Boundary current. A current that flows generally poleward, parallel to a continental coastline. (Ch. 8)

Bowen's reaction series. A schematic description of the order in which different minerals crystallize during the cooling and progressive crystallization of a magma. (Ch. 5)

Braided stream. A channel system consisting of a tangled network of two or more smaller branching and reuniting channels that are separated by islands or bars. (Ch. 9)

Breeder reactor. A nuclear reactor in which fission takes place, specifically, the pile in which the conversion of ^{238}U takes place. (Ch. 17)

Burial metamorphism. Metamorphism caused solely by the burial of sedimentary or pyroclastic rocks. (Ch. 7)

Caldera. A roughly circular, steep-walled volcanic basin several kilometers or more in diameter. (Ch. 5)

Calorie. The amount of heat energy needed to raise the temperature of 1 gram of water by 1 degree Celsius; 1000 calories equal 1 Calorie. (Introduction)

Calving. The progressive breaking off of icebergs from a glacier that terminates in deep water. (Ch. 10)

Carnivores. Meat-eating heterotrophs. (Ch. 15)

Carrying capacity. The limit on the population that an ecosystem can carry, imposed by the limited resources of that ecosystem. (Ch. 15)

Catastrophism. The concept that all of the Earth's major features, such as mountains, valleys, and oceans, have been produced by a few great catastrophic events. (Introduction)

Cation. A positive ion. (Ch. 4)

Cell. The basic structural unit of all living organisms. (Ch. 16)

Channel. The passageway in which a stream flows. (Ch. 9)

Chemical sediment. Sediment formed by precipitation of minerals from solutions in water. (Ch. 7)

Chemical weathering. The decomposition of rocks through chemical reactions such as hydration and oxidation. (Ch. 11)

Chemoautotrophs. Organisms that derive energy from the oxidation of hydrogen sulfide in the water discharged from black smokers. (Ch. 15)

Chemosynthesis. The synthesis of small organic molecules such as amino acids. (Ch. 16)

Cirque. A bowl-shaped hollow on a mountainside, open downstream, bounded upstream by a steep slope (headwall), and excavated mainly by frost wedging and by glacial abrasion and plucking. (Ch. 10)

Clastic sediment. The loose fragmented debris produced by the mechanical breakdown of older rocks. (Ch. 7)

Cleavage. The tendency of a mineral to break in preferred directions along bright, reflective plane surfaces. (Ch. 4)

Climate. The average weather conditions of a place or area over a period of years. (Ch. 12)

Cloud. Visible aggregations of minute water droplets, tiny ice crystals, or both. (Ch. 12)

Coal. A black, combustible, sedimentary or metamorphic rock consisting chiefly of decomposed plant matter and containing more than 50 percent organic matter. (Ch. 7, 17)

Cold front. A front in which dense, cold air flows in and displaces warmer air by pushing it upward, producing clouds and possibly rain. (Ch. 12)

Condensation. The formation of a more ordered liquid from a less ordered gas. (Ch. 12)

Conduction. The means by which heat is transmitted through solids without deforming the solid. (Introduction)

Cone of depression. A conical depression in the water table immediately surrounding a well. (Ch. 9)

Confined aquifer. An aquifer bounded by aquicludes. (Ch. 9)

Conglomerate. A sedimentary rock composed of clasts of rounded gravel set in a finer grained matrix. (Ch. 7)

Constellation. A distinctive star pattern in the sky, named mostly for animals and mythical characters. (Ch. 2)

Continental crust. The part of the Earth's crust that comprises the continents, which has an average thickness of 45 km. (Ch. 1)

Continental rise. A region of gently changing slope where the floor of the ocean basin meets the margin of a continent. (Ch. 6)

Continental shelf. A submerged platform of variable width that forms a fringe around a continent. (Ch. 6)

Continental shield. An assemblage of cratons and orogens that has reached isostatic equilibrium. (Ch. 7)

Continental slope. A pronounced slope beyond the seaward margin of the continental shelf. (Ch. 6)

Convection. The process by which hot, less dense materials rise upward, being replaced by cold, dense, downward flowing material to create a convection current. (Introduction)

Convection current. The flow of material as a result of convection. (Introduction)

Convergence. The coming together of air masses, caused by the inward spiral flow in a cyclone and leading to an upward flow of air at the center of the low-pressure center. (Ch. 13)

Convergent margin. The zone where plates meet as they move toward each other. See *subduction zone.* (Ch. 6)

Core. The spherical mass, largely metallic iron, at the center of the Earth. (Ch. 1)

Coriolis effect. An effect that causes any body that moves freely with respect to the rotating solid Earth to veer toward the right in the northern hemisphere and toward the left in the southern hemisphere, regardless of the initial direction of the moving body. (Ch. 8)

Craton. A core of ancient rock in the continental crust that has attained tectonic and isostatic stability. (Ch. 7)

Crevasse. A deep, gaping fissure in the upper surface of a glacier. (Ch. 10)

Cross bedding. Beds that are inclined with respect to a thicker stratum within which they occur. (Ch. 7)

Crust. The outermost and thinnest of the Earth's compositional layers, which consists of rocky matter that is less dense than the rocks of the mantle below. (Ch. 1)

Cryosphere. The part of the Earth's surface that remains perennially frozen. (Ch. 10)

Crystal. A solid compound composed of ordered, three-dimensional arrays of atoms or ions chemically bonded together and displaying crystal form. (Ch. 4)

Crystal faces. The planar surfaces that bound a crystal. (Ch. 4)

Crystal form. The geometric arrangement of crystal faces. (Ch. 4)

Crystal structure. The geometric pattern that atoms assume in a solid. Any solid that has a crystal structure is said to be *crystalline.* (Ch. 4)

Curie point. A temperature above

which permanent magnetism is not possible. (Ch. 6)

Cyclone. Air spiraling inward around a low-pressure center. (Ch. 13)

Cytoplasm. The main body of the cell, excluding the nucleus and the plasma membrane. (Ch. 16)

Deforestation. The process of forest clearing. (Ch. 18)

Delta. A body of sediment deposited by a streams where it flows into standing water. (Ch. 11)

Density. The average mass per unit volume. (Ch. 4)

Denudation. The sum of the weathering, mass-wasting, and erosional processes that result in the progressive lowering of the Earth's surface. (Ch. 11)

Deoxyribonucleic acid (DNA). A biopolymer consisting of two twisted, chainlike molecules held together by organic molecules called bases; the genetic material for all organisms except viruses, it stores the information on how to make proteins. (Ch. 16)

Desertification. The invasion of desert into nondesert areas. (Ch. 18)

Dew point. The temperature at which the relative humidity reaches 100 percent and condensation starts. (Ch. 12)

Differential stress. Stress in a solid that is not equal in all directions. (Ch. 7)

Dikes. Tabular, parallel-sided sheets of intrusive igneous rocks that cut across the layering of the intruded rock. (Ch. 5)

Diorite. A coarse-grained igneous rock consisting mainly of plagioclase and ferromagnesiam minerals. Quartz is sparse or absent. (Ch. 5)

Discharge. The quantity of water that passes a given point in a stream channel per unit time. (Ch. 9)

Discharge area. Area where subsurface water is discharged to streams or to bodies of surface water. (Ch. 9)

Dissolution. The chemical weathering process whereby minerals and rock material pass directly into solution. (Ch. 11)

Dissolved load. Matter dissolved in stream water. (Ch. 9)

Divergence. The separation of air masses in different directions, caused by the outward spiral flow in an anticyclone and leading to an outward flow of air from the center of a high-pressure center. (Ch. 13)

Divergent margin (of a plate). A fracture in the lithosphere where two plates move apart. Also called a *spreading center.* (Ch. 6)

Divide. The line that separates adjacent drainage basins. (Ch. 9)

Dolostone. A sedimentary rock composed chiefly of the mineral dolomite. (Ch. 7)

Downwelling. The process by which surface water thickens and sinks. (Ch. 8)

Drainage basin. The total area that contributes water to a stream. (Ch. 9)

Dune. A mound or ridge of sand deposited by wind. (Ch. 11)

Earthquake focus. The point of the first release of energy that causes an earthquake. (Ch. 3)

Earth system science. The science that studies the whole Earth as a system of many interacting parts and focuses on the changes within and between these parts. (Introduction)

Ecological niche. The sum of the conditions (including habitat and resources) that allows an organism and its offspring to sustain themselves and breed. (Ch. 15)

Ecosystem. A trophic pyramid and its habitat. (Ch. 15)

Ediacaran animals. The earliest fossils of multicellular organisms discovered in 600-million-year-old rocks in the Ediacara Hills of South Australia. (Ch. 16)

Ekman spiral. A spiraling current pattern from the water's surface to deeper layers, caused by the Coriolis effect, as each successive, slower moving layer of water is shifted to the right. (Ch. 8)

Ekman transport. The average flow of water in a current over the full depth of the Ekman spiral. (Ch. 8)

Elastic deformation. The reversible or nonpermanent deformation that occurs when an elastic solid is stretched and squeezed and the force is then removed. (Ch. 3)

Elastic rebound theory. The theory that earthquakes results from the release of stored elastic energy by slippage on faults. (Ch. 3)

El Niño/Southern oscillation (ENSO). A periodic climatic variation in which tradewinds slacken and surface waters of the central and eastern Pacific become anomalously warm. (Ch. 8)

Electromagnetic radiation. A self-propagating electric and magnetic wave, such as light, radio, ultraviolet, or infrared radiation; all types travel at the same speed and differ in wavelength or frequency, which relates to the energy. (Ch. 2)

Element (chemical). The most fundamental substance into which matter can be separated by chemical means. (Ch. 4)

Emergence. An increase in the area of land exposed above sea level resulting from uplift of the land and/or fall of sea level. (Chs. 8,11)

End moraine. A ridgelike accumulation of drift deposited along the margin of a glacier. (Ch. 11)

Energy-level shell. The specific energy level of electrons as they orbit the nucleus of an atom. (Ch. 4)

Enzyme. A protein that catalyzes a chemical reaction in an organism. (Ch. 16)

Epicenter. That point on the Earth's surface that lies vertically above the focus of an earthquake. (Ch. 3)

Equilibrium line. A line that marks the level on a glacier where net mass loss equals net gain. (Ch. 10)

Erosion. The complex group of related processes by which rock is broken down physically and chemically and the products are moved. (Introduction)

Eucaryotic cell (eucaryotes). A cell that includes a nucleus with a membrane, as well as other membrane-bound organelles. (Ch. 16)

Eutrophication. Bodies of water with a high level of plant nutrients and consequently high levels of algae growth. (Ch. 15)

Evaporite deposits. Layers of salts that precipitate as a consequence of evaporation. (Ch. 11)

Evolution. The changes that species undergo through time, eventually leading to the formation of new species. (Ch. 16)

Extrusive igneous rock. Rock formed by the solidification of magma poured out onto the Earth's surface. (Ch. 5)

Fault. A fracture in a rock along which movement occurs. (Ch. 3)

Fission. Controlled radioactive transformation. (Ch. 17)

Fjord. A deep, glacially carved valley submerged by the sea. Also spelled *fiord.* (Ch. 10)

Floodplain. The part of any stream valley that is inundated during floods. (Ch. 11)

Flux. The amount of energy flowing through a given area in a given time. (Ch. 2)

Foliation. The planar texture of mineral grains, principally micas, produced by metamorphism. (Ch. 7)

Food chain. The pathways by which energy (as food) is moved fron one trophic level to another. (Ch. 15)

Food web. The map of all interconnections among food chains for an ecosystem. (Ch. 15)

Fossil. The naturally preserved remains or traces of an animal or a plant. (Ch. 7)

Fossil fuel. Remains of plants and animals trapped in sediment that may be used for fuel. (Ch. 17)

Friction, The resistance to movement when two bodies are in contact. (Ch. 13)

Front. The boundary between air masses of different temperature and humidity, and therefore different density. (Ch. 12)

Gabbro. A coarse-grained igneous rock in which olivine and pyroxene are the predominant minerals and plagioclase is the feldspar present. Quartz is absent. (Ch. 5)

Galaxy. A cluster of a million or more stars, plus gas and dust, that is held together by gravity. (Ch. 2)

Gangue. The nonvaluable minerals of an ore. (Ch. 17)

Geocentric. A universe in which a stationary Earth is at the center and everything else revolves around it. (Ch. 1)

Geologic column. A composite diagram combining in chronological order the succession of known strata, fitted together on the basis of their fossils or other evidence of relative or actual age. (Ch. 7)

Geostrophic current. A flow of surface water around a gyre that is not deflected toward the center of the gyre. (Ch 8)

Geostrophic wind. A wind that results from a balance between pressure-gradient flow and the Coriolis deflection. (Ch. 13)

Geothermal gradient. The rate of increase of temperature downward in the Earth. (Introduction)

Glaciation. The modification of the land surface by the action of glacier ice. (Ch. 14)

Glacier. A permanent body of ice, consisting largely of recrystallized snow, that shows evidence of downslope or outward movement, due to the stress of its own weight. (Ch. 10)

Glacier ice. Snow that gradually becomes denser and denser until it is no longer permeable to air. (Ch. 10)

Global change. The changes produced in the Earth system as a result of human activities. (Introduction)

Gneiss. A high-grade metamorphic rock, always coarse-grained and foliated, with marked compositional layering but with imperfect cleavage. (Ch. 7)

Gradient. A measure of the vertical drop over a given horizontal distance. (Ch. 9)

Granite. A coarse-grained igneous rock containing quartz and feldspar, with potassium feldspar being more abundant than plagioclase. (Ch. 5)

Gravitation (law of). Every body in the universe attracts every other body. (Ch. 1)

Gravity anomaly. Variations in the pull of gravity after correction for latitude and altitude. (Ch. 3)

Greenhouse effect. The property of the Earth's atmosphere by which long wavelength heat rays from the Earth's surface are trapped or reflected back by the atmosphere. (Ch. 18)

Greenschist. A low-grade metamorphic rock rich in chlorite. (Ch. 7)

Groundwater. All the water contained in the spaces within bedrock and regolith. (Ch. 9)

Gymnosperms. Naked-seed plants. (Ch. 16)

Gyre. A large subcircular current system of which each major ocean current is a part. (Ch. 8)

Hadley cell. Convection cells on both sides of the equator that dominate the winds in tropical and equatorial regions. (Ch. 13)

Halocline. A zone of the ocean, below the surface zone, which is marked by a substantial increase of salinity with depth. (Ch. 8)

Heat capacity. The amount of heat required to raise or lower the temperature of a material. (Ch. 8)

Hardness. Relative resistance of a mineral to scratching. (Ch. 4)

Heat. The energy a body has due to the motions of its atoms. (Ch. 12)

Heat energy. The energy of a hot body. (Ch. 12)

Heliocentric. A universe in which a stationary Sun is at the center and everything else revolves around it. (Ch. 1)

Herbivores. Plant-eating heterotrophs. (Ch. 15)

Hertzsprung-Russell diagram (H-R diagram). A plot of a star's luminosity versus its temperature. (Ch. 2)

Heterotrophs. Organisms that are unable to use the energy from sunlight directly and so must get their energy by eating autotrophs or other heterotrophs. (Ch. 15)

High (H). An area of relatively high air pressure, characterized by diverging winds. (Ch. 13)

Homeostasis. The maintenance of fairly constant internal conditions; a balance within an ecosystem. (Ch. 15)

Humidity. The amount of water vapor in the air. (Ch. 12)

Hydrosphere. The totality of the Earth's water, including the oceans, lakes, streams, water underground, and all the snow and ice, including glaciers. (Introduction)

Hydrothermal mineral deposit. Any local concentration of minerals formed by deposition from a hydrothermal solution. (Ch. 17)

Hypothesis. An unproved explanation for the way things happen. (Introduction)

Igneous rock. Rock formed by the cooling and consolidation of magma. (Ch. 4)

Inner core. The central, solid portion of the Earth's core. (Ch. 1)

Insolation. The energy that reaches the surface of the Earth from the Sun. (Ch. 12)

Interglaciation. Period between glacial epochs. (Ch. 14)

Intertropical convergence zone. A low-pressure zone of convergent air masses caused by warm air rising in the tropics. (Ch. 13)

Intrusive igneous rock. Any igneous rock formed by solidification of magma below the Earth's surface. (Ch. 5)

Ion. An atom that has excess positive or negative charges caused by electron transfer. (Ch. 4)

Isobar. Lines on a map connecting places of equal air pressure. (Ch. 13)

Isostasy. The ideal property of flotational balance among segments of the lithosphere. (Ch. 3)

Isotopes. Atoms of an element having the same atomic number but differing mass numbers. (Ch. 4)

joule. The work done when a force of 1 Newton acts over a distance of 1 meter. (Introduction)

Jovian planets. Giant planets in the outer regions of the solar system that are characterized by great masses, low densities, and thick atmospheres consisting primarily of hydrogen and helium. (Ch. 1)

Karst topography. An assemblage of topographic forms resulting from dissolution of carbonate bedrock and consisting primarily of closely spaced sinks. (Ch. 11)

Laccolith. A lenticular pluton intruded parallel to the layering of the intruded rock, above which the layers of the invaded country rock have been bent upward to form a dome. (Ch. 5)

Landslide. Any perceptible downslope movement of a mass of bedrock or regolith, or a mixture of the two. (Ch. 11)

Latent heat. The amount of heat released or absorbed per gram during a change of state. (Ch. 12)

Lava. Magma that reaches the Earth's surface through a volcanic vent. (Ch. 5)

Law (scientific). A statement that some aspect of nature is always observed to happen in the same way and that no deviations have ever been seen. (Introduction)

Lead (in sea ice). A linear opening in thin ice cover caused by stresses resulting from the diverging movement of the ice cover. (Ch. 10)

Limestone. A sedimentary rock consisting chiefly of calcium carbonate, mainly in the form of the mineral calcite. (Ch. 7)

Lithosphere. The outer 100 km of the solid Earth, where rocks are harder and more rigid than those in the plastic asthenosphere. (Ch. 1)

Load. The material that is moved or carried by a natural transporting agent, such as a stream, the wind, a glacier, or waves, tides, and currents. (Ch. 9)

Loess. Wind-deposited silt, sometimes accompanied by some clay and fine sand. (Ch. 11)

Low (L). An area of relatively low air pressure, characterized by converging winds, ascending air, and precipitation. (Ch. 13)

Luminosity. The total amount of energy radiated outward each second by the Sun or any other star. (Ch. 2)

Luster. The quality and intensity of light reflected from a mineral. (Ch. 4)

Magma. Molten rock, together with any suspended mineral grains and dissolved gases, that forms when temperatures rise and melting occurs in the mantle or crust. (Ch. 5)

Magmatic mineral deposit. Any local concentration of minerals formed by magmatic process in an igneous rock. (Ch. 17)

Main sequence. The principal series of stars in the Hertzsprung-Russell diagram, which includes stars that are converting hydrogen to helium. (Ch. 2)

Mantle. The thick shell of dense, rocky matter that surrounds the core. (Ch. 1)

Marble. A metamorphic rock derived from limestone and consisting largely of calcite. (Ch. 7)

Mass balance (of a glacier). The sum of the accumulation and ablation on a glacier during a year. (Ch. 10)

Mass-wasting. The movement of regolith downslope by gravity without the aid of a transporting medium. (Ch. 11)

M-discontinuity. See *Mohorovicic discontinuity.* (Ch. 3)

Meander. A looplike bend of a stream channel. (Ch. 9)

Mesopause. The boundary between the mesosphere and the thermosphere. (Ch. 12)

Mesosphere (atmospheric science). One of the four thermal layers of the atmosphere, lying above the stratosphere. (Ch. 12)

Metabolism. The sum of all the chemical reactions in an organism, by which it grows and maintains itself. (Ch. 16)

Mesosphere (geology). The region between the base of the asthenosphere and the core/mantle boundary. (Ch. 1)

Metallogenic provinces. Limited regions of the crust within which mineral deposits occur in unusually large numbers. (Ch. 17)

Metamorphic facies. Contrasting assemblages of minerals that reach equilibrium during a metamorphism within a specific range of physical conditions belonging to the same metamorphic facies. (Ch. 7)

Metamorphic rock. Rock whose original compounds or textures, or both, have been transformed to new compounds and new textures by reactions in the solid state as a result of high temperature, high pressure, or both. (Ch. 4)

Metamorphism. All changes in mineral assemblage and rock texture, or both, that take place in sedimentary and igenous rocks in the solid state within the Earth's crust as a result of changes in temperature and pressure. (Ch. 7)

Midocean ridges. Continuous rocky ridges on the ocean floor, many hundreds to a few thousand kilometers wide with a relief of more than 0.6 km. Also called *oceanic ridges* and *oceanic rises.* (Ch. 4)

Mineral. Any naturally formed, crystalline solid with a definite chemical composition and a characteristic crystal structure. (Ch. 4)

Mineral assemblage. The variety and abundance of minerals present in a rock. (Ch. 4)

Mineral deposit. Any volume of rock containing an enrichment of one or more minerals. (Ch. 17)

Modified Mercalli Scale. A scale used to compare earthquakes based on the intensity of damage caused by the quake. (Ch. 3)

Mobo. See *Mohorovicic's discontinuity.* (Ch. 3)

Mohoroivicic discontinuity (also called *M-discontinuity* and *Mohd).* The seismic discontinuity that marks the base of the crust. (Ch. 3)

Moraine. An accumulation of drift deposited beneath or at the margin of a glacier and having a surface form that is unrelated to the underlying bedrock. (Ch. 10)

Negative feedback. The influence of a product on the process that produces it, such that production decreases with the growth of the product. (Ch. 15)

North Atlantic Deep Water (NADW). A deep-ocean mass in the North Atlantic that extends from the intermediate water to the ocean floor; dense and cold, it originates at several sites near the surface of the North Atlantic, flows downward, and spreads southward into the South Atlantic. (Ch. 8)

Oceanic crust. The crust beneath the oceans. (Ch. 1)

Oceanic ridges. See *midocean ridges.* (Ch. 4)

Omnivores. Heterotrophs that eat both meat and plants. (Ch. 15)

Organelles. A well-defined cell part that has a particular function in the operation of the cell. (Ch. 16)

Ore. An aggregate of minerals from which one or more minerals can be extracted profitably. (Ch. 17)

Original horizontality (law of). Waterlaid sediments are deposited in strata that are horizontal, or nearly horizontal, and parallel, or nearly parallel, to the Earth's surface. (Ch. 7)

Orogens. Elongate regions of the crust that have been intensively folded, faulted, and thickened as a result of continental collisions. (Ch. 7)

Outer core. The outer portion of the Earth's core, which is molten. (Ch. 1)

Paleomagnetism. Remanent magnetism in ancient rock recording the direction of the magnetic poles at some time in the past. (Ch. 6)

Paleontologist. A scientist who studies extinct organisms. (Ch. 16)

Panspermia. The hypothesis of the supposed origin of life in space, followed by a diaspora to various parts of the galaxy (including the Earth). (Ch. 16)

Peat. An unconsolidated deposit of plant remains that is the first stage in the conversion of plant matter to coal. (Ch. 7)

Periglacial. A land area beyond the limit of glaciers where low temperature and frost action are important factors in determining landscape characteristics. (Ch. 10)

Permafrost. Sediment, soil, or bedrock that remains continuously at a temperature below 0°C for an extended time. (Ch. 10)

Permeability. A measure of how easily a solid allows a fluid to pass through it. (Ch. 9)

Petroleum. Gaseous, liquid, and semisolid substances occurring naturally and consisting chiefly of chemical compounds of carbon and hydrogen. (Ch. 17)

Phyllite. A well-foliated metamorphic rock in which the component platy minerals are just visible. (Ch. 7)

Physical weathering. The disintegration (physical breakup) of rocks. (Ch. 11)

Pile. A device in which nuclear fission can be controlled. (Ch. 17)

Placer. A deposit of heavy minerals concentrated mechanically. (Ch 17)

Planetary accretion. The process by which bits of condensed solid matter were gathered to form the planets. (Ch. 1)

Plate tectonics. The special branch of tectonics that deals with the processes by which the lithosphere is moved laterally over the asthenosphere. (Introduction)

Pluton. Any body of intrusive igneous rock, regardless of shape or size. (Ch. 5)

Polar (cold) glacier. A glacier in which the ice is below the pressure melting point throughout, and the ice is frozen to its bed. (Ch. 10)

Porosity. The proportion (in percent) of the total volume of a given body of bedrock or regolith that consists of pore spaces. (Ch. 9)

Porphyry. Any igneous rock consisting of coarse mineral grains scattered through a mixture of fine mineral grains. (Ch. 5)

Positive feedback. The influence of a product on the process that produces it, such that production increases the growth of the product. (Ch. 15)

Pressure melting point. The temperature at which ice can melt at a given pressure. (Ch. 10)

Primary waves. See *P waves.* (Ch. 3)

Principle of stratigraphic superposition. See *stratigraphic superposition.*

Principle of Uniformitarianism. The same external and internal processes we recognize in action today have been operating unchanged, though at different rates, throughout most of the Earth's history. (Introduction)

Procaryotic cell (procaryotes). Cells without a nucleus; refers to single-celled organisms that have no membrane separating their DNA from the cytoplasm. (Ch. 16)

Protein. Molecule formed through the polymerization of an amino acid. (Ch. 16)

Provinciality. The extent to which the global ecosystem is divided into subsystems by barriers to the migration of organisms. (Ch. 15)

P waves. Seismic body waves transmitted by alternating pulses of compression and expansion. *P* waves pass through solids, liquids, and gases. (Ch. 3)

Pycnocline. An ocean zone beneath the surface zone in which water density increases rapidly, as a result of a decrease in temperature, an increase in salinity, or both. (Ch. 8)

Pyroclast. A fragment of rock ejected during a volcanic eruption. (Ch. 5)

Pyroclastic rocks. Rocks formed from pyroclasts. (Ch. 5)

Quartzite. A metamorphic rock consisting largely of quartz, and derived from a sandstone. (Ch. 7)

Radiation. Transmission of heat energy through the passage of electromagnetic waves. (Introduction)

Recharge. The addition of water to the saturated zone of a groundwater system. (Ch. 9)

Recharge area. Area where water is added to the saturated zone. (Ch. 9)

Red giant. A large, cool star with a high luminosity and a low surface temperature (about 2500 K), which is largely convective and has fusion reactions going on in shells. (Ch. 2)

Reflection. The bouncing of a wave off the surface between two media. (Ch. 2)

Refraction. The change in velocity when a wave passes from one medium to another; the process by which the path of a beam of light is bent when the beam crosses from one transparent material to another. (Ch. 2)

Regional metamorphism. Metamorphism affecting large volumes of crust and involving both mechanical and chemical changes. (Ch. 7)

Regolith. The irregular blanket of loose, noncemented rock particles that covers the Earth. (Introduction)

Relative humidity. The ratio of the vapor pressure in a sample of air to the saturation vapor pressure at the same temperature, expressed as a percentage. (Ch. 12)

Relative velocity (of a plate). The apparent velocity of one plane relative to another.

Relief (topographic). The range in altitude of a land surface. (Ch. 11)

Residual mineral deposit. Any local concentration of minerals formed as a result of weathering. (Ch. 17)

Rhyolite. A fine-grained igneous rock with the composition of a granite. (Ch. 5)

Ribonucleic acid (RNA). A single-stranded molecule similar to the DNA molecule, but with a slightly different chemical composition; it reads and executes the codes contained in the DNA. (Ch. 16)

Richter magnitude scale. A scale, based on the recorded amplitudes of seismic body waves, for comparing the amounts of energy released by earthquakes. (Ch. 3)

Rock. Any naturally formed, nonliving, firm, and coherent aggregate mass of mineral matter that constitutes part of a planet. (Introduction)

Runoff. The fraction of precipitation that flows over the land surface. (Ch. 9)

Salinity. The measure of the sea's saltiness; expressed in parts per thousand. (Ch. 8)

Saltation. The progressive forward movement of a sediment particle in a series of short intermittent jumps along arcing paths. (Chs. 9, 11)

Sand sea. Vast tract of shifting sand. (Ch. 11)

Sandstone. A medium-grained clastic sedimentary rock composed chiefly of sand-sized grains. (Ch. 7)

Saturated zone. The groundwater zone in which all openings are filled with water. (Ch. 9)

Schist. A well-foliated metamorphic rock in which the component platy minerals are clearly visible. (Ch. 7)

Scientific method. The use of evidence that can be seen and tested by anyone who has the means to do so, consisting often of observation, formation of a hypothesis, testing of that hypothesis and formation of a theory, formation of a law, and continual reexamination. (Introduction)

Seafloor spreading (theory of). A theory proposed during the early 1960s in which lateral movement of the oceanic crust away from midocean ridges was postulated. (Ch. 6)

Sea ice. A thin veneer of ice at the ocean surface in the polar latitudes; accounts for approximately two-thirds of the Earth's permanent ice cover. (Ch. 10)

Secondary waves. See *S waves.* (Ch. 3)

Sediment. Regolith that has been transported by any of the external processes. (Ch. 4)

Sedimentary mineral deposit. Any local concentration of minerals formed through processes of sedimentation. (Ch. 17)

Sedimentary rock. Any rock formed by chemical precipitation or by sedimentation and cementation of mineral grains transported to a site of deposition by water, wind, ice, or gravity. (Ch. 4)

Seismic sea waves (also called *tsunami).* Long wavelength ocean waves produced by sudden movement of the seafloor following an earthquake. Incorrectly called tidal waves. (Ch. 3)

Seismic waves. Elastic disturbances spreading outward from an earthquake focus. (Ch. 3)

Shale. A fine-grained, clastic sedimentary rock. (Ch. 7)

Shell fusion. The process of nuclear fusion in a star, in which the hydrogen in the shell around its core is converted into helium after the hydrogen in the core itself has already been depleted; such a star becomes a red giant. (Ch. 2)

Shield volcano. A volcano that emits fluid lava and builds up a broad dome-shaped edifice with a surface slope of only a few degrees. (Ch. 5)

Silicate (*-silicate mineral*). A mineral that contains the silicate anion. (Ch. 4)

Silicate anion. A complex ion $(SiO^4)^{-4}$, that is present in all silicate minerals. (Ch. 4)

Sills. Tabular, parallel-sided sheets of intrusive igneous rock that are parallel to the layering of the intruded rock. (Ch. 5)

Siltstone. A sedimentary rock composed mainly of silt-sized mineral fragments. (Ch. 7)

Sinkhole. A large solution cavity open to the sky. (Ch. 11)

Slate. A low-grade metamorphic rock with a pronounced slaty cleavage. (Ch. 7)

Snowline. The lower limit of perennial snow. (Ch. 10)

Soil. The part of the regolith that can support rooted plants. (Ch. 11)

Soil horizons. The subhorizontal weathered zones formed as a soil develops. (Ch. 11)

Soil profile. A vertical section through a soil that displays its component horizons. (Ch. 11)

Solar nebula. A flattened rotating disc of gas and dust surrounding the Sun. (Ch. 1)

Species. A population of individuals that can interbreed to produce offspring that are, in turn, interfertile with each other. (Ch. 15)

Specific gravity. A number stating the ratio of the weight of a substance to the weight of an equal volume of pure water. A dimensionless number numerically equal to the density. (Ch. 4)

Spectrum. A group of electromagnetic rays arranged in order of increasing or decreasing wavelength. (Ch. 2)

Spreading center (also called a *divergent margin).* The new, growing edge of a plate. Coincident with a mid-ocean ridge. (Chs. 4, 6)

Standard atmosphere. The model or average air pressure at sea level: 760 mm or 29.9 inches of mercury. (Ch. 12)

Stock. A small, irregular body of intrusive igneous rock, smaller than a batholith, that cuts across the layering of the intruded rock. (Ch. 5)

Strata. See *stratum.*

Stratification. The layered arrangement of sediments, sedimentary rocks, or extrusive igneous rocks. (Ch. 7)

Stratigraphic superposition (principle of). In a sequence of strata, not later overturned, the order in which they were deposited is from bottom to top. (Ch. 7)

Stratigraphy. The study of strata. (Ch. 7)

Stratopause. The boundary between the stratosphere and the mesosphere. (Ch. 12)

Stratosphere. One of the four thermal layers of the atmosphere, lying above the troposphere and reaching a maximum of about 50 km. (Ch. 12)

Stratovolcanoes. Volcanoes that emit both tephra and viscous lava, and that build up steep conical mounds. (Ch. 5)

Stratum (plural = *strata*). A distinct layer of sediment that accumulated at the Earth's surface. (Ch. 7)

Streak. A thin layer of powdered mineral made by rubbing a specimen on a nonglazed porcelain plate. (Ch. 4)

Stream. A body of water that carries detrital particles and dissolved substances and flows down a slope in a definite channel. (Ch. 9)

Striatums. Subparallel scratches inscribed on a clast or bedrock surface by rock debris embedded in the base of the glacier. (Ch. 11)

Subduction zone (also called a *convergent margin).* The linear zone along which a plate of lithosphere sinks down into the asthenosphere. (Chs. 4, 6)

Submergence. A rise of water level relative to the land so that areas formerly dry are inundated. (Chs. 8, 11)

Supernova. A stupendous explosion of a star, which increases its brightness hundreds of millions of times in a few days; a supernova releases heavy elements into space, and what remains of its core becomes a black hole. (Ch. 2)

Surf. Wave activity between the line of breakers and the shore. (Ch. 8)

Surface waves. Seismic waves that are guided by the Earth's surface and do not pass through the body of the Earth. (Ch. 3)

Surge. An unusually rapid movement of a glacier marked by dramatic changes in glacier flow and form. (Ch. 10)

Suspended load. Fine particles suspended in a stream. (Ch. 9)

S waves. Seismic body waves transmitted by an alternating series of sideways (shear) movements in a solid. 5 waves cause a change of shape and cannot be transmitted through liquids and gases. (Ch. 3)

Symbiotic. A close, long-term relationship between individuals of different species. (Ch. 15)

Tar (also called *asphalt).* An oil that is viscous and so thick it will not flow. (Ch. 17)

Temperate (warm) glacier. A glacier in which the ice is at the pressure-melting point and water and ice coexist in equilibrium. (Ch. 10)

Temperature. A measure of the average kinetic energy of all the atoms in a body. (Ch. 12)

Tephra. A loose assemblage of pyroclasts. (Ch. 5)

Tephra cone. A cone-shaped pile of tephra deposited around a volcanic vent. (Ch. 5)

Terrace. An abandoned floodplain formed when a stream flowed at a level above the level of its present channel and floodplain. (Ch. 11)

Terrestrial planets. The innermost planets of the solar system (Mercury, Venus, Earth, and Mars), which have high densities and rocky compositions. (Ch. 1)

Texture. The overall appearance that a rock has because of the size, shape, and arrangement of its constituent mineral grains. (Ch. 4)

Theory. A hypothesis that has been examined and found to withstand numerous tests. (Introduction)

Thermocline. A zone of ocean water lying beneath the surface zone, characterized by a marked decrease in temperature. (Ch. 8)

Thermohaline circulation. Global patterns of water circulation propelled by the sinking of dense cold and walty water. (Ch. 8)

Thermosphere. One of the four thermal layers of the atmosphere, reaching out to about 500 km. (Ch. 12)

Topographic relief. The difference in altitude between the highest and lowest points on a landscape. (Ch. 11)

Transform fault margin (of a plate). A fracture in the lithosphere along which two plates slide past each other. (Ch. 6)

Tributary. A stream that joins a larger stream. (Ch. 9)

Trophic pyramid. The hierarchy of organisms in which energy is moved from one level to the next. (Ch. 15)

Tropopause. The boundary between the troposphere and the stratosphere. (Ch. 12)

Troposphere. One of the four thermal layers of the atmosphere, which extends from the surface of the Earth to variable altitudes of 10 to 16 km. (Ch. 12)

Tsunami. See *seismic sea waves.* (Chs. 3, 8)

Tuff. A pyroclastic rock consisting of ash- or lapilli-sized tephra, hence *ash tuff* and *lapilli tuff.* (Ch. 5)

Unconfined aquifer. An aquifer with an upper surface that coincides with the water table. (Ch. 9)

Uncomformity. A substantial break or gap in a stratigraphic sequence that marks the absence of part of the rock record. (Ch. 7)

Uniform stress. Stress that is equal in all directions. Also called *confining stress* or *homogeneous stress.* (Ch. 7)

Uniformitarianism. See *Principle of Uniformitarianism.*

Upwelling. The process by which subsurface waters flow upward and replace the water moving away. (Ch. 8)

Varve. A pair of sedimentary layers deposited during the seasonal cycle of a single year. (Ch. 7)

Viscosity. The internal property of a substance that offers resistance to flow. (Ch. 5)

Volcanic neck. The approximately cylindrical conduit of igneous rock forming the feeder pipe of a volcanic vent that has been stripped of its surrounding rock by erosion. (Ch. 5)

Volcanic pipe. A cylindrical conduit of igneous rock below a volcanic vent. (Ch. 5)

Volcanic rock. Rock formed from the volcanic eruption of lava or tephra; often very fertile. (Ch. 5)

Volcano. The vent from which igneous matter, solid rock, debris, and gases are erupted. (Ch. 5)

Warm front. A front in which warm, humid air advances over colder air, producing clouds and possibly rain. (Ch. 12)

Water table. The upper surface of the saturated zone of groundwater. (Ch. 9)

watt. A unit of power at the rate of 1 joule per second. (Introduction)

Wave. An oscillatory movement of water characterized by an alternate rise and fall of the water surface.

Wave base. The effective lower limit of wave motion, which is half of the wavelength. (Ch. 8)

Wavelength. The distance between the crests or troughs of adjacent waves. (Ch. 8)

Wave refraction. The process by which the direction of a series of waves, moving into shallow water at an angle to the shoreline, is changed. (Ch. 8)

Weather. The state of the atmosphere at a given time and place. (Ch. 12)

Weathering. The chemical alteration and mechanical breakdown of rock materials during exposure to air, moisture, and organic matter. (Ch. 11)

Welded tuff (also called *ignimbrite).* Pyroclastic rocks, the glassy fragments of which were plastic and so hot when deposited that they fused to form a glassy rock. (Ch. 5)

Western boundary current. A current that flows generally poleward, parallel to a continental coastline; the poleward direction is caused by the deflection of westward-flowing equatorial currents as they encounter land. (Ch. 8)

White dwarf. A small, dense star that has exhausted its nuclear fuel and shines from residual heat; it has a high surface temperature but low luminosity. (Ch. 2)

Wind. Horizontal air movement arising from differences in air pressure. (Ch. 13)

Windchill factor. The heat loss from exposed skin as a result of the combined effects of low temperature and wind speed. (Ch. 13)

Zodiac. The 12 contellations through which the Sun passes. (Ch. 2)

Zone of aeration. The groundwater zone in which open spaces in regolith or bedrock are filled mainly with air. (Ch. 9)

CREDITS

PHOTOS

FRONTISPIECE: Science Source/ Photo Researchers.

FRONT MATTER (Top Left) Courtesy NASA. (Top Center) Frank Rossotto/Tom Stack & Associates. (Top Right) Hiroji Kubota/Magnum Photos, Inc. (Bottom Left) Impact Photos. (Bottom Center) Frans Lanting/Minden Pictures, Inc. (Bottom Right) Victor Englebert/Photo Researchers.

INTRODUCTION Opener: Andrew Holbrooke/Black Star. **Figure 1.3a:** John S. Shelton. **Figure I.3b:** William E. Ferguson. **Figure 1.4:** K.R. Gill. **Figure 1.5:** John S. Shelton. **Figure 1.6:** Professor W. Alvarez/Photo Researchers. **Figure 1.9:** Earth Satellite Corporation. **Figure 1.10:** John Bova/Photo Researchers. Essayist: Courtesy William Fyfe.

PART I OPENER: NASA JPL—Starlight.

CHAPTER 1 Opener: Courtesy NASA. **Figure 1.1:** Mary Evans Picture Library. **Figure 1.9:** Royal Observatory, Edinburgh, and Anglo-Australian Telescope Board. **Figure 1.15:** Courtesy Space Photography Laboratory, Arizona State University. **Jupiter Box Figure 1:** AP/Wide World Photos. **Jupiter Box Figure 2:** AP/Wide World Photos. **Essayist:** Courtesy Harrison Schmitt.

CHAPTER 2 Opener: Tony Hallas/Science Photo Library/ Photo Researchers. **Figure 2.7:** Courtesy NASA. **Figure 2.8:** Francois Gohier/Photo Researchers. **Figure 2.12A:** Courtesy NASA. **Figure 2.12B:** Science Photo Library/Photo Researchers. **Figure 2.12C:** Pekka Parviainen/Science Photo Library/ Photo Researchers. **Figure 2.14:** Jon Sanford/Science Photo Library/ Photo Researchers. **Figure C2.3A:** Richard Megna/Fundamental Photographs. **Figure C2.6:** Galen Rowell/Peter Arnold **Inc. Essayist:** Courtesy Barbara Welther.

PART II OPENER: Nicholas DeVore/Tony Stone Work Wide.

CHAPTER 3 Opener: J.C. Aunos/Gamma Liaison. **Figure 3.10:** Joe Caravetta/Gamma Liaison. **Figure 3.11:** Boris Yurchenko/AP/Wide World Photos. **Figure 3.12:** Stan Wayman/ Life Picture Service. **Essayist:** Courtesy Jeffrey Park.

CHAPTER 4 Opener: Jacques Jangoux/Peter Arnold, Inc. **Figure 4.7:** William Sacco. **Figures 4.9 A-D:** Brian J. Skinner. **Figure C4.1:** Wiilliam Sacco. **Figure C4.2:** Brian J. Skinner. **Figures C4.3, C4.4, and C4.5:** Williar Sacco. **Figure C4.6:** Breck Kent/Animals Animals. **Figur C4.7-.** William Sacco. **Essayist:** Courtesy Jill Schneiderman.

CHAPTER 5 Opener: Bettmann Archive. **Figure 5.2:** G. Brad Lewis/Gamma Liaison. **Figure 5.3:** J. D. Griggs/USGS. **Figure** 5.4: Krafft Explorer/Photo Researchers. **Figure 5.5:** J. D. Griggs/USGS. **Figure 5.6A:** William E. Ferguson. **Figure 5.6B:** J.P. Lockwood/USGS. **Figure 5.6C:** Steven L. Nelson/Alaska Stock Images. **Figure 5.8:** S. C. Porter. **Figure 5.9A:** Krafft Explorer/Photo Researchers. **Figure 5.9B:** Tom Bean/DRK Photo. **Figure 5.10:** Steve Vidler/Leo de Wys, Inc. **Figure 5.11:** Greg Vaughn/Tom Stack & Associates. **Figure 5.13:** David Hiser/Photographers Aspen. **Figure 5.14:** Ron Sanford/F-Stop Pictures. **Figure 5.16B:** Brian J. Skinner. **Figures C5.1A-D:** William Sacco. **Figure C5.2:** Brian J. Skinner. **Figures C5.3** A **and B:** William Sacco. **Essayist:** Courtesy Kathy Crane.

CHAPTER 6 Opener: Frank Rossotto/Tom Stack & Associates. **Figure 6.17:** Loren Mclntrye/Woodfin Camp & Associates. **Essayist:** Courtesy Charles Drake.

CHAPTER 7 Opener: Ric Ergenbright. **Figure 7.1:** Josef Muench. **Figures 7.6: and 7.7:** S.C. Porter. **Figure 7.8:** Jim Frazier. **Figures 7.11** A **and B:** William Sacco. **Figures 7.13** A **and B:** Schalk W. vand der Merwe. **Figures C7.3 A-D:** Brian J. Skinner. **Figures C7.5** A **and B:** Craig Johnson. Essayist: Courtesy Paul Lowman.

PART HI OPENER: C. Clapperton.

CHAPTER 8 Opener: Herb Kane. **Figure 8.15:** Arnulf Husmo/Tony Stone Images. **Figure 8.18:** Jeff Divine/FPG International. **Figure 8.19:** John S. Shelton. **Figures 8.25 A and B:** Greg Scott/Masterfile. **Figure 8.27A:** Arthur Bloom. **Figure C8.2A:** S.C. Porter. **Essayist:** Courtesy Victoria Kaharl.

CHAPTER 9 Opener: Lloyd Cluff. **Figure 9.1:** F. Stuart Westmorland/Photo Researchers. **Figure 9.4:** S.C. Porter. **Figure 9.7:** John Eastcott/YVA Momatiuk/DRK Photo. **Figure 9.8:** Courtesy NASA. **Figure 9.11:** Hiroji Kubota/Magnum Photos, Inc. **Figures 9.14** A **and B:** Earth Satellite Corporation. **Figures 9.15** A **and B:** S.C. Porter. **Figure 9.25:** The Bancroft Library, University of California. **Figure 9.26:** S.C. Porter. **Figure 9.29:** Michael Melford/The Image Bank. **Essayist:** Courtesy Waite Osterkamp.

CHAPTER 10 Opener: Jack Sayers Australian Antarctic Division. **Figure 10.1:** Courtesy NASA. **Figure 10.2:** S. C. Porter. **Figure 10.5A:** Breck P. Kent/Animals Animals. **Figure 10.5B:** Sam Abell/National Geographic Society. **Figure 10.5C:** John Price/Courtesy USDA. **Figure 10.6:** USGS. **Figure 10.7:** Courtesy NASA. **Figures 10.10** A **and B:** S.C. Porter. **Figures C10.1 and 10.17:** Austin Post. **Figure 10.18:** N. Untersteiner. **Figure 10.23:** Steve McCutcheon. **Figure 10.24:** Charlie Ott/Photo Researchers. **Essayist:** Courtesy Lonnie Thompson.

CHAPTER 11 Opener: Steve Raymer/National Geographic Society. **Figure 11.1:** William E. Ferguson. **Figures 11.2 and 11.3A:** S. C. Porter. **Figure 11.3B:** Fiorenzo Ugolini. **Figure 11.6:** Abril Imagens/ Nani Gois. **Figure 11.11:** Earth Satellite Corporation. **Figure 11.13:** S.C. Porter. **Figure 11.14:** Martin Miller. **Figure 11.15:** Earth Satellite Corporation. **Figure 11.16:** Michael Nichols/Magnum Photos, Inc. **Figure 11.17:** Jim Tuten/Black Star. **Figure 11.18:** Hiroji Kubota/Magnum Photos, Inc. **Figures 11.19, 11.20** A **and B:** S.C. Porter. **Figure 11.21:** Ole P. Rorvik/Aune Forlag. **Figure 11.22:** Tom Bean/AllStock, Inc. **Figures** 11.23 **and** 11.24: S.C. Porter. **Figure 11.25:** Joe Cornish/Tony Stone World Wide. **Figures 11.27 and 11.28:** S.C. Porter. **Figure 11.29A:** Spaceshots, Inc. **Figure 11.29B:** Courtesy NASA. Essayist: Courtesy Wes Ward.

PART IV OPENER: Science Source/Photo Researchers.

CHAPTER 12 Opener: Chris Bonington. **Figure 12.1:** Dr. Gerald L. Fisher/Science Photo Library/Photo Researchers. **Figure 12.8:** French Government Tourist Office. **Figure 12.9:** Courtesy NASA. **Figures 12.20** A **and B:** Tom Bean/DRK Photo. **Figure 12.20C:** S. Nielsen/ Bruce Coleman, Inc. **Figure 12.20D:** Joyce Photographers/ Photo Researchers. **Figure 12.20E:** Steve McCutcheon. **Figure 12.20F:** Darryl Torckler/Tony Stone World Wide. **Figure 12.20G:** Keith Gunnar/Bruce Coleman, Inc. **Figure 12.2011:** Barrie Rokeach. **Figure 12.201:** Stan Osolinski/Oxford Scientific Films Ltd. Essayist: Courtesy James C.G. Walker.

CHAPTER 13 Opener and Figure 13.1: Stephen Krasemann/DRK Photo. **Figure 13.8:** National Oceanic and Atmospheric Administration/DLR—FRG/Starlight, **Figure 13.20:** Warren Faidley/ Weatherstock. **Figure 13.21:** John H. Hoffman/Bruce Coleman, Inc. **Figure 13.22:** Impact Photos. **Figure 13.23:** Courtesy NASA Goddard Space Flight Center/A.F. Hasler, II. Pierce, K. Palaniappan, and M. Manyin. **Essayist:** Courtesy Howard Bluestein.

CHAPTER 14 Opener: Hanny Paul/Gamma Liaison. **Figures 14.4 and 14.8A:** S.C. Porter. **Figure 14.11:** Tom Bean. **Figure 14.12B:** Matsuo Tsukada. **Figure C14.1:** Deep Sea Drilling Project/Scripps Institute of Oceanography. **Figure 14.24:** Roger Werth/Woodfin Camp & Associates. **Essayist:** Courtesy L. Graumlich.

PART V OPENER: John Reader/Science Photo Library/ Photo Researchers.

CHAPTER 15 Opener: Frans Lanting/Minden Pictures, Inc. **Figure 15.1:** Wendy Stone/Gamma Liaison. **Figures 15.4** A **and** B: David Muench/AllStock, Inc. **Figure 15.5A:** Wm. Munoz/Photo Researchers. **Figure 15.5B:** Willard Clay. **Figure 15.9C:** M.I. Walker/Natural History Photographic Agency. **Figure 15.11:** Valerie Taylor/Ardea London. **Figure 15.12A:** Jany Sauvanet/Natural History Photographic Agency. **Figure I5.12B:** Jany Sauvanet/ Auscape International Pty. Ltd. **Figure 15.12C:** A. Cosmos Blank/Photo Researchers. **Figure 15.12D:** Karl Weidmann/Photo Researchers. **Figures 15.14 and 15.16:** The Natural History Museum, London. **Figure 15.18:** Robert R. Hessler/Scripps Institution of Oceanography. **Figure 15.19:** Mike Andrews/Earth Scenes. **Figure C15.1:** Fred Ward/Black Star. **Figure 15.20:** Michael P. Gadomski/Photo Researchers. **Essayist:** Courtesy Compton Tucker.

CHAPTER 16 Opener: Tom McHugh/Photo Researchers. **Figures 16.1** A **and** B: David M. Phillips/Visuals Unlimited. **Figure 16.1C:** Omikron/Science/Photo Researchers. **Figure 16.2:** Richard Feldman/Phototake. **Figure 16.3:** Manfred Kage/Peter Arnold, Inc. **Figure 16.5**: Reuters/ Bettmann Archive. **Figure 16.6A:** Dr. Jeremy Burgess/Photo Researchers. **Figure 16.6B:** CNRI/Photo Researchers. **Figure 16.7:** Prof. J. William Schopf, Dept. of Earth and Planetary Sciences, UCLA. **Figure 16.8A:** Ford Kristo/Planet Earth Pictures. **Figure 16.8B:** Francois Gohier/Ardea London. **Figure 16.10:** Brian J. Skinner. **Figure 16.12A:** D. Pauer and E. Paner-Cook/Auscape International Pty. Ltd. **Figure 16.12B:** Ken Lucas/Planet Earth Pictures. **Figure 16.13:** A. Kerstitch/Bruce Coleman, Inc. **Figure 16.16A:** Edward R. Degginger/Bruce Coleman, Inc. **Figure 16.16B:** John Shaw/Bruce Coleman, Inc. **Figure 16.17A:** Martin Land/Science Photo Library/Photo Researchers. **Figure 16.17B:** Breck P. Kent/Earth Scenes. **Figure 16.18:** Patti Murray/Animals Animals. **Figure 16.19:** S. Conway Morris. **Figure C16.2:** Phil Degginger/Bruce Coleman, Inc. Essayist: Courtesy Andrew Knoll.

PART VI OPENER: Courtesy NASA.

CHAPTER 17 Opener: Victor Englebert/Photo Researchers. **Figure 17.4B:** Dudley Foster/Woods Hole Oceanographic Institution. **Figures 17.5, 17.7** A **and** B: Brian J. Skinner. **Figure 17.8:** William Sacco. **Figures 17.10, 17.11** A **and B:** Brian J. Skinner. **Essayist:** Courtesy Pinar Yilmaz.

CHAPTER 18 Opener: Courtesy IBM. **Figure 18.3:** Courtesy Agency for International Development. **Figure 18.4:** George Gerster/COMSTOCK, Inc. **Figure 18.7:** Giorgio Gualco/Bruce Coleman, Inc. **Essayist:** Courtesy Thomas J. Crowley.

LINE ART AND TABLE CREDITS

CHAPTER 1 Figure 1.11, p. 33, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 12.23, p. 247.

CHAPTER 2 **Figure** 2.3, p. 48, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 1.11, p. 8. **Figure 2.4**, p. 48, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 13.4, p. 264. **Figure 2.6**, p. 52, based on John Gribbin (1993, March 13), Inside the sun, *New Scientist*, p. 1; Edward J. Tarbuck and Frederick J. Lutgens (1994), *Earth science*, 7th ed. (New York, Macmillan), Fig. 19.18, p. 656; Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 13.31, p. 284; Cesare Emiliani (1992), *Planet earth: Cosmology, geology, and the evolution of life and environment* (Cambridge, Cambridge University Press), Fig. 6.3. **Figure 2.9**, p. 54, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 13.6, p. 266. **Figure 2.10**, p. 55, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 13.5, p. 265. **Figure 2.11**, p. 55, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 13.6, p. 266. **Figure 2.13**, p. 57, adapted by permission from Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 13.20, p. 278. **Figure 2.15**, p. 58, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 14.8, p. 294. **Figure 2.16**, p. 59, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Figs. 14.6, 14.7, p. 291. **Figure C2.4**, p. 60, based on Michael Zeilik (1991), *Astronomy: The evolving universe*, 6th ed. (New York, John Wiley & Sons), Figs. 6.4, 6.7A, pp. 94, 95. **Figure C2.5**, p. 61, based on Michael Zeilik (1991), *Astronomy- The evolving universe*, 6th ed. (New York, John Wiley & Sons), Figs. 6.6, 6.8, pp. 95, 96. **Figure 2.17**, p. 62, adapted by permission from Michael Zeilik (1991), *Astronomy The evolving universe*, 6th ed. (New York, John Wiley & Sons), Fig. 14.15, p. 300.

CHAPTER 8 **Figure 8.1**, p. 196, map adapted by permission of Art Division, © National Geographic Society, from *National Geographic*, vol. 150 (October 1976), p. 517. **Figure 8.2**, p. 197, based on maps from *National Geographic Atlas of the World* (1963), p. 10. **Figure 8.6A**, p. 200, adapted by permission of Houghton Mifflin from *The Times Atlas of the World*, Comprehensive edition (1967), plate 3 (mean surface salinity). **Figure 8.6B**, p. 201, adapted by permission of Andrew Mclntyre from CLIMAP Project Members (1981), Map and Chart Series MC-36, Map 6A, published by the Geological Society of America. **Figure C8.1**, p. 207, adapted by permission from Colin S. Ramage (1986, June), El Niño, *Scientific American*, p. 76. Copyright © 1986 by Scientific American, Inc. All rights reserved. **Figure C8.2B**, p. 208, adapted by permission from J. Cole and R. Fairbanks (1990/91), The El Niño/Southern Oscillation system and the record of past climate change, *Lamont-Doherty Geological Observatory (Columbia University) Report 1990/91*, Fig. 3, p. 65. **Figure 8.12A**, p. 209, based on A. C. Duxbury and A. B. Duxbury (1989), *An introduction to the world's oceans* (Dubuque, Iowa; Wm. C. Brown), Fig. 7.IB. **Figure 8.12B**, p. 209, based in part on *National Geographic Atlas of the World* (1963), pp. 12-13. **Figure 8.13A**, p. 210, adapted by permission from A. L. Gordon (1990/91), The role of thermocline circulation in global climate change, *Lamont-Doherty Geological Observatory (Columbia University) Report 1990/91*, Fig. 3, p. 48. **Figure 8.13B**, p. 211, based on information from *Lamont-Doherty Geological Observatory (Columbia University) Report 1990/91*, Fig. 4, p. 50; J. Imbrie et al. (1992), On the structure and origin of major glaciation cycles, 1; Linear responses to Milankovitch forcing. *Paleoceanographyl*, Fig. lb, p. 704, published by the American Geophysical Union. **Figure 8.14**, p. 212, adapted by permission from S. J. Lehman and L. D. Keigwin (1992), Deep circulation revisited, *Nature*, vol. 358, pp. 197-198. Copyright © 1992 Macmillan Magazines Limited. **Figure 8.27B**, p. 220, adapted from N. J. Shackleton (1987), Oxygen isotopes, ice volume, and sea level, *Quaternary Science Reviews*, vol. 6, Fig. 5, p. 187, by permission of Elsevier Science Ltd., Pergamon Imprint, Oxford, England.

CHAPTER 9 **Figure 9.10**, p. 232, adapted from P. E. Potter et al. (1988), Teaching and field guide to alluvial processes and sedimentation of the Mississippi River, Fulton County, Kentucky, and Lake County Tennessee, *Kentucky Geological Survey*, Fig. 8. **Figure 9.12A**, p. 233, adapted by permission from C. F. Nordin, Jr., et al. (1980), Size distribution of Amazon River bed sediment, *Nature*, vol. 286, Fig. 2a. Copyright © 1980 Macmillan Magazines Limited. **Figure 9.12B**, p. 233, adapted from P. E. Potter et al. (1988), Teaching and field guide to alluvial processes and sedimentation of the Mississippi River, Fulton County, Kentucky, and Lake County Tennessee, *Kentucky Geological Survey*, Fig. 13. **Figure 9.13**, p. 233, adapted by permission from A. Foucault and D. J. Stanley (1989), Late quaternary palcoclimatic oscillations in East Africa recorded by heavy minerals in the Nile delta, *Nature*, vol. 339, Figs. 1, 2, p. 44. Copyright © 1989 Macmillan Magazines Limited. **Figure 9.16**, p. 236, reprinted by permission from Stanley A. Schumm, M. Paul Mosley, and William E. Weaver (1987), *Experimental fluvial geomorphology* (New York, John Wiley & Sons), Fig. 2.18.

CHAPTER 10 Figure 10.4, p. 254, adapted by permission from M. F. Meier and A. S. Post (1962), Recent variations in mass net budgets of glaciers in western North America, in *Symposium of Obergurgl: Variations of the Regime of Existing Glaciers*, IAHS Pub. 58, pp. 63-77. **Figure 10.19**, p. 267, adapted from Central Intelligence Agency (1978), *Polar regions atlas*, pp. 12, 38. **Figure 10.21**, p. 269, adapted by permission from T. L. Pewe (1983), The penglacial environment in North America during Wisconsin time, in S. C. Porter (ed.), *Late quaternary environments of the United States, vol. 1: The late Pleistocene* (Minneapolis, University of Minnesota Press), Fig. 9.1. **Figure 10.22**, p. 270, adapted by permission from S. W. Muller (1947), *Permafrost of permanently frozen ground and related engineering problems* (Ann Arbor, Mich., J. W. Edwards), Fig. 3.

CHAPTER 11 **Figure 11.26,** p. 299, reprinted by permission of HarperCollins Publishers Limited from K. Pye and H. Tsoar (1990), *Aeolian sand and sand dunes* (London, Unwin Hyman), Fig. 7.1. **Figure C11.1,** p. 303, based on P. F. Cerveny et al. (in press), History of uplift and relief of the Himalaya during the past 18 million years: Evidence from fission-track ages of detrital zircons from sandstones of the Siwalik Group, in Paola and Kleinspehn (eds.), *New perspectives in sedimentary Basin Analysis* (Springer Verlag), Fig. 3.4. **Figure C11.2,** p. 303, adapted by permission of the University of Chicago Press from J. D. Milliman and R. H. Meade (1983), World-wide delivery of river sediment to the oceans, *Journal of Geology,* vol. 91, pp. 1-21. Copyright© 1983 by the University of Chicago Press. **Figure 11.31,** p. 305, adapted by permission from W. F. Ruddiman and J. E. Kutzbach (1990), Late Cenozoic plateau uplift and climate change, *Transactions of the Royal Society of Edinburgh: Earth Sciences,* vol. 81, pt. 4, Fig. 1, p. 302.

CHAPTER 12 **Figure C12.1**, p. 319, based on J. M. Moran and M. D. Morgan (1991), *Meteorology,* 3d ed. (New York, Macmillan); R. G. Fleagle and J. Businger (1963), *An introduction to atmospheric physics* (New York, Academic Press), p. 153. **Figure 12.14**, p. 325, reprinted by permission from S. D. Gedzelman (1980), *The source and wonders of the atmosphere* (John Wiley & Sons, 1980), Fig. 10.6.

CHAPTER 13 **Figure 13.3**, p. 340, from National Weather Service. **Figure 13.6**, p. 341, from National Weather Service. **Figure 13.13**, p. 346, based on data from National Weather Service in Alan H. Strahler and Arthur N. Strahler (1992), *Modern physical geography,* 4th ed. (New York, John Wiley & Sons), p. 85; H. Riehl (1962), *Jet streams of the atmosphere* (Fort Collins, Colorado State University). **Figure C13.1**, p. 349, reprinted by permission from S. D. Gedzelman (1980), *The source and wonders of the atmosphere* (New York, John Wiley & Sons), Fig. 16.7. **Figure 13.16**, p. 350, reprinted by permission of HarperCollins Publishers Limited from K. Pye and H. Tsoar (1990), *Aeolian sand and sand dunes* (London, Unwin Hyman), Fig. 2.17. **Figure 13.17**, p. 350, reprinted by permission from L. K. Lustig (1974), Deserts, in *Encyclopedia Britannica* (15th ed.), Figs. 4, 5, 6, 7, 8. **Figure 13.19**, p. 353, reprinted with permission of Macmillan Publishing Company from J. M. Moran and M. D. Morgan (1991), *Meteorology*, 3d ed., Fig. 11.1. **Figure 13.24**, p. 356, reprinted with permission of MacMillan Publishing Company from J.M. Moran and M.D. Morgan (1991), *Meteorology,* 3d ed., p. 385.

CHAPTER 14 **Figure 14.2**, p. 363, data from Bernard Janin (1970), *he col du Grand-Saint-Bernard climat et variations climatiques* (Aoste, Italy, Marguerettaz-Musumeci), Tables 38, 40. **Figure 14.3**, p. 364, based on NCAR sea-level temperature records. **Figure 14.5A**, p. 365, adapted by permission from Zhang De-er (1982), Analysis of dust rain in the historic times of China. *Kexue Tongbao,* vol. 27, pp. 294-297. **Figures 14.5B and 14.5C**, p. 365, adapted by permission from H. H. Lamb (1977), *Climate—Present, past and future, vol. 2: Climatic history and the future* (London, Methuen). **Figure 14.5D**, p. 365, adapted by permission from H. H. Lamb (1966), *The changing climate: Selected papers* (London, Methuen), app. 3, p. 223. **Figure 14.6**, p. 366, adapted by permission from W. S. B. Paterson et al. (1977), An oxygen-isotope climatic record from the Devon Island ice cap, arctic Canada, *Nature,* vol. 266, Fig. 6. Copyright © 1977 Macmillan Magazines Limited. **Figure 14.8**, p. 367, reprinted by permission from H. H. Lamb (1977), *Climate—Present, past, and future, vol. 2: Climatic history and the future* (London, Methuen), Fig. 17.17, p. 462; derived from unpublished paper by L. M. Libby, R and D Associates, Santa Monica, Calif., 1974. **Figure 14.14**, p. 373, adapted by permission from Thompson Webb III et al. (1993), Vegetation, lake levels, and climate in eastern North America for the past 18,000 years, in H. E. Wright, Jr., et al. (eds.), *Global climates since the last glacial maximum* (Minneapolis, University of Minnesota Press), Fig. 17.10. **Figure C14.2**, p. 377, adapted by permission of Andrew Mclntyre from CLIMAP Project Members (1981), Seasonal reconstructions of the Earth's surface at the last glacial maximum, MC-36, maps 3A and 3B, published by the Geological Society of America. **Figure 14.17**, p. 378, based in part on N. J. Shackleton and J. P. Kennett (1975), Paleotemperature history of the Cenozoic and the initiation of Antarctic glaciation: Oxygen and carbon isotope analyses in Deep Sea Drilling Project sites, 277, 279, and 281, in J. P. Kennett et al. (1975), *Initial reports of the Deep Sea Drilling Project*, vol. 29, National Ocean Sediment Coring Program (Washington, D.C., National Science Foundation), Fig. 7. **Figure 14.18**, p. 379, adapted from Fig. 8.3 of *Paleoclimatology* by Thomas J. Crowley and Gerald R. North. Copyright © 1991 by Oxford University Press, Inc.—reprinted by permission; and adapted by permission of Elsevier Science Publishers and E. J. Barron from E. J. Barron and W. M. Washington (1982), Cretaceous climate: A comparison of atmospheric simulations with the geologic record. *Paleogeography Paleoclimatology, Paleoecology,* vol. 40, Fig. 9. **Figure 14.19**, the American Geophysical Union. **Figure 14.23**, p. 383, based in part on C. Lorius et al. (1990), The ice-core record: Climate sensitivity and future greenhouse warming, *Nature,* vol. 347, pp. 139-145; J. T. Houghton, G. J. Jenkins, and J. J. Ephraums (1990), *Climate change: The 1PCC scientific assessment* (Cambridge, Cambridge University Press), Fig. 2.

CHAPTER 15 **Figure 15.2**, p. 393, reprinted with permission of Macmillan Publishing Company from Fig. 4.10, p. 167, of *Communities and ecosystems,* by R. H. Whittaker. Copyright © 1970 by Robert H. Whittaker; as it appeared in W. B. Clapham (1973), *Natural ecosystems* (New York, Macmillan), p. 29. **Figure 15.3**, p. 393, adapted by permission from Ecological relations of jaegers and owls as lemming predators near Barrow, Alaska, by F. A. Pitelka, P. 0- Tomich, and G. W. Treichel, *Ecological Monographs,* 1955, vol. 25, Fig. 2, p. 89. Copyright © 1966 by the Ecological Society of America. As it appeared in C. J. Kriss (1972), *Ecology* (New York, Harper & Row), p. 441. **Figure 15.6,** p. 396, based on U.S. Department of State data in E. A. Keller (1992), *Environmental geology*, 6th ed. (New York, Macmillan), p. 10. **Figure 15.7**, p. 397, adapted from S. C. Kendeigh (1961), *Animal ecology,* Fig. 17.15, p. 239. Reprinted by permission of Prentice Hall, Inc., Englewood Cliffs, NJ. Data from D. A. MacLulich (1937), Fluctuations in the numbers of the varying hare (Lepus americanus), *University of Toronto Biological Series,* vol. 43, pp. 1-136. Figure appeared in Wang Jen-Yu, William B. Hemmers, and Frederick B. Puttie (1973), *Exploring man's environment* (Palo Alto, Calif., Field Educational Publications), p. 110. **Figures 15.9A and 15.9B,** p. 398, adapted by permission from G. F. Cause (1934), *The struggle for existence* (Baltimore, Williams & Wilkins); as they appeared in W. B. Clapham (1973), *Natural ecosystems* (New York, Macmillan), p. 66. **Figure 15.10**, p. 399, reprinted by permission of Yale University Press from F. G. Stehli (1968), Taxonomic diversity gradients in pole location: The recent model, in E. T. Drake (ed.), *Evolution and environment,* Fig. 6, p. 170. Copyright © 1968 by Yale University Press. **Figure 15.13**, p. 402, adapted by permission from E. Barghoorn (1971, May), The oldest fossils, *Scientific American,* p. 41. Copyright © 1971 by Scientific American, Inc. All rights reserved. **Figure 15.15**, p. 403, adapted by permission from Wann Langston, Jr. (1981, February), Pterosaurs, *Scientific American,* p. 130 (illustration by Tom Prentiss). Copyright © 1981 by Scientific American, Inc. All rights reserved.

CHAPTER 16 **Figure 16.4**, p. 418, adapted by permission from G. Brum, L. McKane, and G. Karp (1994), *Biology: Explaining life,* 2d ed. (New York, John Wiley & Sons), Fig. 14.4. **Figure 16.9**, p. 424, adapted by permission from G. Brum, L. McKane, and G. Karp (1994), *Biology: Explaining life,* 2d ed. (New York, John Wiley & Sons), Fig. 5.24 (a and b). **Figure 16.11**, p. 427, based on G. Brum, L. McKane, and G. Karp (1994), *Biology: Explaining life,* 2d ed. (New York, John Wiley & Sons), Fig. 35.3. **Figure 16.14**, p. 429, based on a drawing by Virge Kask in Derek E. G. Briggs (1991), Extraordinary fossils, *American Scientist,* vol. 79, no. 2, p. 133. **Figure 16.15**, p. 430, composite drawing adapted by permission from G. Brum, L. McKane, and G. Karp (1994), *Biology: Explaining life*, 2d ed. (New York, John Wiley & Sons), Figs. 35.9, 35.10, 35.12. **Figure 16.20**, p. 431, adapted by permission of the American Geological Institute from J. John Sepkoski, Jr. (1994, March), Extinction and the fossil record, *Geotimes*, p. 17.

CHAPTER 18 **Figure 18.1**, p. 473, adapted by permission of Cambridge University Press from B. L. Turner et al. (eds.) (1990), *The Earth as transformed by human action*, Fig. 3.1, p. 43. **Table 18.1,** p. 476, reprinted by permission of Routledge, ITPS Ltd., from M. K. Tolba and D. A. El-Kholy (eds.) (1992), *The world environment 1972-1992* (London, Chapman and Hall), p. 169. **Figure 18.5**, p. 477, adapted by permission of Routledge, ITPS Ltd. from M. K. Tolba and D. A. El-Kholy (eds.) (1992), *The world environment 1972-1992* (London, Chapman and Hall), Fig. 3, p. 164; by permission of Oxford University Press from T. C. Whitmore (1990), *An introduction to tropical forests* (Oxford, Clarendon Press); and by permission from Sader and Joyce (1988), Deforestation rates and trends in Costa Rica, 1940-1983, *Biotropica*, 20, 11-19. **Figure 18.8**, p. 480, adapted with permission from U.S. Committee for the Global Atmospheric Research Program, *Understanding climatic change: A program for action.* Fig. 3.2. Copyright 1975 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, D.C. **Table 18.2**, p. 480, reprinted by permission from Atmospheric trace gases that are relatively active and of significance to global change (1991), *EarthQuest* (UCAR/Office of Interdisciplinary Earth Studies), vol. 5, no. 1. **Figure 18.9**, p. 481, adapted by permission from C. D. Keeling et al. (1989), A three-dimensional model of atmospheric CO2 transport based on observed winds; 1. Analysis of observational data. *Aspects of climate variability in the Pacific and the Western Americas* (AGU Monograph 55), Fig. 16, p. 179. Copyright 1989 by the American Geophysical Union. **Figure 18.10A**, p. 481, adapted from A. M. Solomon et al. (1985), The global carbon cycle, in J. R. Trabalka (ed.), *Atmospheric carbon dioxide and the global carbon cycle,* U.S. Department of Energy Report DOE/ER-0239, Fig. 1.2, p. 7. **Figure I8.10B**. p. 481, adapted by permission of the Royal Swedish Academy of Sciences from International Geosphere Biosphere Programme (1990), Past global changes (PAGES), in *The International Geosphere-Biosphere Programme: A Study of Global Change,* IGBP Report 12, Fig. 1, pp. 7.1-4 (including data from Stauffer, 1988, andBarnett, 1988). **Figure 18.11**, Data provided by P.D. Jones, Climatic Research Unit (personal communication), updated from: Jones, P.D., Wigley, T.M.L. and Farmer, G. 1991: Marine and land temperature data sets: a comparison and a look at recent trends. (In) *Greenhouse-Gas-Induced Climatic Change: A Critical Apopraisal od Simulations and Observations* (ed. M.E. Schlesinger), Elsevier Science Publishers, Amsterdam, Nethrelands, 153-172 and Parker, D.E. and Folland, C.K., 1991: (In) *Greenhouse-Gas-Induced Climatic Change: A Critical Apopraisal od Simulations and Observations* (ed. M.E. Schlesinger), Elsevier Science Publishers, Amsterdam, Nethrelands, 173-194. **Figure 18.12**, p. 483, reprinted by permission from M. E. Schlesinger (1983), A review of climate model simulations of CO_2-induced climatic change, *Report No. 41,* Oregon State University, Climatic Research Institute, Fig. 3; as adapted by permission from M. E. Schlesinger and W. L. Gates (1981a), Preliminary analysis of four general circulation model experiments on the role of the ocean in climate, *Report No. 25,* Oregon State University, Climatic Research Institute. **Figure 18.13**, p. 484, reprinted by permission of IPCC Scientific Assessment from J. T. Houghton, G. J. Jenkins, and J. J. Ephraums (eds.) (1990), *Climate change: The IPCC scientific as-*

sessment (Cambridge, Cambridge University Press), Fig. 5. **Figure 18.14,** p. 485, adapted from M. Schlesinger and J. Mitchell (1987), Climate model simulations of the equilibrium climate response to increased carbon dioxide, *Reviews of Geophysics,* vol. 25, Figs. 14, 15, 16. Copyright 1987 by the American Geophysical Union. **Table 18.3,** p. 485, reprinted by permission of IPCC Scientific Assessment from J. T. Houghton, G. J. Jenkins, and J. J. Ephraums (eds.) (1990), *Climate change: The IPCC scientific assessment* (Cambridge, Cambridge University Press), Table 5.1. **Figure 18.15**, p. 486, from *Ice ages: Solving the mystery,* by John Imbrie and Katherine Palmer Imbrie, © 1979, Fig. 48, p. 186. Used by permission of Enslow Publishers, Inc. **Figure 18.16**, p. 487, reprinted by permission of IPCC Scientific Assessment from GCM output in J. T. Houghton, G. J. Jenkins, and J. J. Ephraums (eds.) (1990), *Climate change: The IPCC scientific assessment* (Cambridge, Cambridge University Press). **Figure C18.1**, p. 488, adapted by permission from W. Dansgaard, J. W. C. White, and S. J. Johnsen (1989), The abrupt termination of the Younger Dryas climate event, *Nature,* vol. 339, Fig. 1. Copyright © 1989 Macmillan Magazines Limited. **Figure 18.17**, p. 491, based on data of K. Hammer, personal communication.

INDEX

Page numbers in italics indicate illustrations.

INTERACTIONS AMONG THE FOUR MAJOR PARTS OF THE EARTH SYSTEM

Moorea, one of a chain of volcanoes that form the Society Islands in the South Pacific, is part of the *lithosphere*. In the background rises Tahiti, another member of the island chain. A line of white breakers marks the outer edge of a barrier reef that surrounds a shallow, light-blue lagoon, and beyond lies the dark blue of the deep ocean, which constitutes the largest part of the *hydrosphere*. The marine organisms that live on and build the coral reef, as well as the verdant blanket of tropical plants covering the steep slopes of the island, are parts of the *biosphere*. Above Tahiti, in the *atmosphere*, lies a blanket of altocumulus clouds.

ISBN 0-471-54021-8

90000

9 780471 540212

JOHN WILEY & SONS, INC.
New York Chichester Brisbane Toronto Singapore