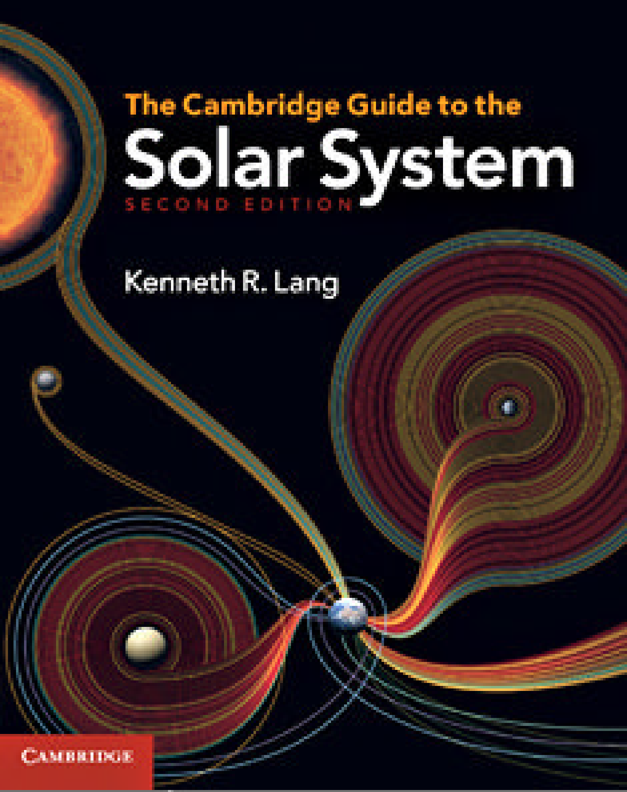


The Cambridge Guide to the Solar System

SECOND EDITION

Kenneth R. Lang

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521198578

This page intentionally left blank

The Cambridge Guide to the Solar System

Richly illustrated with full-color images, this book is a comprehensive, up-to-date description of the planets, their moons, and recent exoplanet discoveries.

The second edition of this classic reference is brought up-to-date with the fascinating new discoveries made during recent years from 12 new solar system missions. Representative examples include water on the Moon; widespread volcanism on Mercury's previously unseen half; vast buried glaciers on Mars; geysers on Saturn's active water moon Enceladus; lakes of methane and ethane on Saturn's moon Titan; the encounter with asteroid Itokawa; and an encounter and sample return from comet Wild 2. The book is further enhanced by hundreds of striking new images of the planets and moons.

Written at an introductory level appropriate for high-school and undergraduate students, it provides fresh insights that appeal to anyone with an interest in planetary science. A website hosted by the author contains all of the images in the book with an overview of their importance. A link to this can be found at www.cambridge.org/solarsystem/.

KENNETH R. LANG is a Professor of Astronomy at Tufts University. He is a well-known author and has published 25 books. *The Cambridge Encyclopedia of the Sun* (Cambridge University Press, 2001) was recommended by the *Library Journal* as one of the best reference books published that year. He has extensive teaching experience, and has served as a Visiting Senior Scientist at NASA Headquarters.

Several Circles. January–February 1926. The artist Vasily Kandinsky (1866–1944) seems to capture the essence of our space-age exploration of previously unseen worlds in this cosmic and harmonious painting. According to Kandinsky, “The circle is the synthesis of the greatest oppositions. It combines the concentric and the eccentric in a single form and in equilibrium. Of the three primary forms, it points most clearly to the fourth dimension.” (Courtesy of the Solomon R. Guggenheim Museum, New York City, New York.)

The Cambridge Guide to the Solar System

Second Edition

Kenneth R. Lang
Tufts University, Medford, Massachusetts, USA

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/solarsystem

First published 2003
Second edition 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-19857-8 Hardback

Contents

Part 3 The giant planets, their satellites and their rings: worlds of liquid, ice and gas

Part 4 Remnants of creation: small worlds in the solar system

Focus elements

Tables

Preface to the second edition

The second edition of *The Cambridge Guide to the Solar System* brings this comprehensive description of the planets and moons up to date, by extending it to include fascinating new discoveries made during the previous decade. As with the first edition, it is written at an introductory level appropriate for high-school and undergraduate students, while also providing fresh, current insights that will appeal to professionals as well as general readers with an interest in planetary science. This is accomplished in a light and uniform style, including everyday metaphors and many spacecraft images.

This second edition is filled with vital new facts and information, and lavishly illustrated in color throughout. Hundreds of new images have been provided. Most of these illustrations have never appeared together in print before, and many of them have a beauty comparable to works of art.

An Internet site for use by the instructor, students or casual reader also supports this book. It contains all of the images in this second edition, together with their legends and overview bullets of their seminal content. This site also includes similar material for the author's books about the Sun, including the second edition of *Sun, Earth and Sky* and the second edition of *The Sun from Space*. The website address is http://ase.tufts.edu/cosmos/.

Striking examples of new images from contemporary planetary spacecraft include the *Chandrayaan-1* and *LCROSS* missions to the Moon, the *MESSENGER* spacecraft that is viewing the unseen half of Mercury, the *2001 Mars Odyssey*, *Spirit* and *Opportunity Exploration Rovers* and *Phoenix* lander on Mars, the *Cassini–Huygens* mission to Saturn and its moons Enceladus and Titan, the *Deep Impact* and *Stardust* encounters with comets, with *Stardust*'s sample return to Earth, and the *Hayabusa* encounter with the asteroid Itokawa.

The more effective illustrations from previous spacecraft have been retained, without an excessive increase in the length of the book, including those from the *Apollo* missions to the Earth's Moon, the *Viking 1* and *2* missions to Mars, the *Mars Global Surveyor*, the *Voyager 1* and *2* missions to the four giant planets, the *Galileo* mission to Jupiter, and several spacecraft encounters with asteroids and comets.

We have not forgotten our home planet Earth, which continues to provide the reference background for discussions of volcanoes, water, geology, atmospheres and magnetospheres. The second edition also includes an updated appraisal of the effects of global warming, with attempts to combat it, and investigations of space weather that can threaten astronauts and influences the performance, reliability and lifetime of interplanetary spacecraft, Earth-orbiting satellites, and terrestrial communications and power systems.

A new chapter, entitled 'Beyond Neptune', takes us to the outer precincts of the planetary realm, with the discovery of several worlds that orbit the Sun beyond the orbit of Neptune. At least three of these dwarf planets are either larger than Pluto or comparable to it in size, and along with Pluto they have also been designated Plutoids. Three companion moons are also now known to orbit Pluto, and its tenuous atmosphere and surface markings have been scrutinized in anticipation of the encounter of the *New Horizons* spacecraft with Pluto in 2015. In the meantime, the *Voyager 1* and *2* spacecraft have traveled far beyond the planets and measured the termination shock of the Sun's winds at about 100 times the distance between the Earth and the Sun.

An entirely new end-chapter describes the origin of our solar system and the exciting new discoveries of hundreds of planets orbiting nearby stars other than the Sun. These extrasolar planets, or exoplanets for short, include multiplanet systems and are mainly hot Jupiters, which revolve unexpectedly close to their star and have masses comparable to that of Jupiter. But atmospheres have been found on some of these new exoplanets, and astronomers are also finding super-Earths of just a few times the mass of the Earth. Astronomers are now finding Earth-size planets that reside in the warm habitable zone where liquid water can exist on the planet's surface and living things might be present.

The main body of this second edition does not simply consist of a few updates or cosmetic patch ups of the material in the first edition. It instead contains all of the relevant new discoveries, ideas and information resulting from recent planetary spacecraft. They include: water on the Moon; evidence for widespread volcanic activity on young Mercury; ubiquitous evidence for ancient water flows and vast amounts of frozen water now on Mars; jets of ice particles, water vapor and organic compounds from Saturn's active water moon Enceladus; organic dunes and rain and lakes of liquid methane and ethane on Saturn's moon Titan; close-up details of Saturn's rings and its enigmatic moons Hyperion and Phoebe, immersed within a newly discovered ring; asteroid Itokawa's loose collection of rubble; the return of comet dust, containing organic molecules, from comet Wild 2 to Earth; and the *Deep Impact* collision with comet Tempel 1, with forced ejections of water vapor, water ice and other substances.

In addition to bringing images and discoveries up to date, the second edition also adds more scientific substance with set-aside *focus boxes* that emphasize basic planetary physics at the algebra level, such as the luminosity and temperature of the Sun; conservation of angular momentum in the Earth–Moon system; escape velocity, temperature and retention of planetary atmospheres; the Roche limit, and gravitational collapse. These interesting boxed set-asides can be used in introductory university courses that include fundamental scientific topics, but they are not a crucial aspect of the main text and can be bypassed by students with non-scientific interests or by the general reader.

In short, this is a fine, stimulating collection of exciting spacecraft images and marvelous discoveries about the planets and moons, as well as exoplanets, and I hope the reader derives as much pleasure as I have from finding out about them.

I am grateful to three bright, alert Tufts students, Laura Costello, Jeffrey Gottlieb and Nathaniel Eckman, for reading draft chapters of this book, offering insightful comments and spotting necessary corrections. This volume was also substantially improved by the careful editing of Sue Glover. Special thanks are extended to Joe Bredekamp at NASA Headquarters who has actively encouraged the writing of this book and helped fund it through NASA Grant NNX07AU93G with NASA's Applied Information Systems Research Program.

Kenneth R. Lang
Tufts University

Preface to the first edition

The planets have been the subject of careful observations and myth for millennia and the subject of telescopic studies for centuries. Our remote ancestors looked into the night sky, and wondered why the celestial wanderers or planets moved across the stellar background. They saw the planets as powerful gods, whose Greek and Roman names are still in use today. Then progressively larger telescopes enabled the detection of faint moons and remote planets that cannot be discerned with the unaided eye, and resolved fine details that otherwise remain blurred.

Only in the past half century have we been able to send spacecraft to the planets and their moons, changing many of them from moving points of light to fascinating real worlds that are stranger and more diverse than we could have imagined. Humans have visited the Moon, and robot spacecraft have landed on Venus and Mars. We have sent vehicles to the very edge of the planetary system, capturing previously unseen details of the remote giant planets, dropping a probe into Jupiter's stormy atmosphere, and perceiving the distant satellites as unique objects whose complex and richly disparate surfaces rival those of the planets. Probes have also been sent to peer into the icy heart of two comets, and robotic eyes have scrutinized the battered and broken asteroids.

The Cambridge Guide to the Solar System is a complete modern guide, updating and extending the prize-winning *Wanderers in Space* – Prix du livre de l'Astronomie in 1994. This book, written by the author and Charles A. Whitney, was completed before the *Clementine* and *Lunar Prospector* spacecraft were sent to the Moon, the *Magellan* orbiter penetrated the veil of clouds on Venus, the *Mars Pathfinder* landed on the red planet with its mobile roving *Sojourner*, the *Mars Global Surveyor* obtained high-resolution images of the surface of Mars, the *NEAR-Shoemaker* spacecraft orbited the asteroid 433 Eros, the *Galileo* orbiter and probe visited Jupiter and its four large moons, *Deep Space 1* peered into the nucleus of Comet Borrelly, and Comet Shoemaker–Levy 9 collided with Jupiter. *The Guide* updates *Wanderers* to include the captivating results of all these missions, presenting more than a half century of extraordinary accomplishment.

The Cambridge Guide to the Solar System provides comprehensive accounts of the most recent discoveries, from basic material to detailed concepts. It is written in a concise, light and uniform style, without being unnecessarily weighed down with incomprehensible specialized materials or the variable writing of multiple authors. Metaphors, similes and analogies will be of immense help to the lay person and they add to the enjoyment of the material. Vignettes containing historical, literary and even artistic material make this book unusual and interesting, but at a modest level that enhances the scientific content of the book and does not interfere with it.

The book is at once an introductory text of stature and a thorough, serious and readable report for general readers, with much compact reference data. The language, style, ideas and profuse illustrations will attract the general reader as well as students and professionals. In addition, it is filled with vital facts and information for astronomers of all types and for anyone with a scientific interest in the planets and their satellites.

The many full-color images, photographs, and line drawings help make this information highly accessible.

Each chapter begins with a set of pithy, one-sentence statements that describe the most important or interesting things that will be described in that chapter. A summary diagram, placed at the end of each chapter, captures the essence of our knowledge of the subject.

Set-aside *focus boxes* enhance and amplify the discussion with interesting details, fundamental physics and important related topics. They will be read by the especially curious person or serious student, but do not interfere with the general flow of the text and can be bypassed by the general educated reader who wants to follow the main ideas. Equations are kept to a minimum and, when employed, are almost always placed within the set-aside *focus elements*.

Numerous tables provide fundamental physical data for the planets and large moons. Many graphs and line drawings complement the text by summarizing what spacecraft have found. Guides to other resources are appended to the book as an annotated list of books for further reading, all published after 1990, and a list of relevant Internet addresses.

The Cambridge Guide to the Solar System has been organized into four main parts. The first introduces the planets and their moons, with a brief historical perspective followed by a discussion of their common properties. These unifying features include craters, volcanoes, water, atmospheres and magnetic fields. The second part discusses the rocky worlds found in the inner solar system – the Earth with its Moon, Mercury, Venus and Mars. The third part presents the giant planets, their satellites and their rings – worlds of liquid, ice and gas. The last part discusses the smaller worlds, the comets and asteroids, as well as collisions of these bodies with Jupiter, the Sun and Earth.

Chapter 1 traces our evolving understanding of the planets and their satellites made possible by the construction of ever-bigger telescopes. They resulted in the discovery of new planets and satellites, and resolved details on many of them. Here we include, in chronological order, the discoveries of Jupiter's moons, Saturn's rings, Uranus, Neptune, the asteroids, the icy satellites of the giant planets, tiny Pluto with its oversized moon, and the small icy objects in the Kuiper belt at the edge of the planetary realm. Other fundamental discoveries have been woven into the fabric of this chapter, including the realization that planets are whirling endlessly about the Sun, refinements of the scale and size of the planetary realm, and the spectroscopic discovery of the main ingredients of both the Sun and the atmospheres of the planets.

Chapter 2 begins with a description of how spacecraft have fundamentally altered our perception of the solar system, providing detailed close-up images of previously unseen landscapes and detecting incredible new worlds with sensors that see beyond the range of human vision. These new vistas have also resulted in a growing awareness of the similarities of the major planets and some moons. In the rest of Chapter 2 and in Chapter 3, they are therefore interpreted as a whole, rather than as isolated objects, by presenting comparative aspects of common properties and similar processes. This provides a foundation for subsequent examination of individual objects in greater detail.

Impact craters are found on just about every body in the solar system from the Moon and Mercury to the icy satellites of the distant planets, but in different amounts that depend on their surface ages and with varying properties. Ancient impacts on Venus have, for example, been erased by outpourings of lava, and the debris from subsequent impacts has been shaped by the planet's thick atmosphere. Numerous volcanoes have also been found throughout the solar system, including fiery outbursts on the Earth, towering volcanic mountains on Mars, numerous volcanoes that have resurfaced Venus, currently active volcanoes that have turned Jupiter's satellite Io inside out, and eruptions of ice on Neptune's largest moon, Triton.

Liquid water, which is an essential ingredient of life, covers seventy-one percent of the Earth's surface. Catastrophic floods and deep rivers once carved deep channels on Mars, and spring-like flows have been detected in relatively recent times. Water ice is ubiquitous in the outer solar system, including the clouds of Jupiter, the rings of Saturn, and the surfaces of most satellites. There is even evidence for subsurface seas beneath the water-ice crusts of Jupiter's satellites Europa, Ganymede and Callisto, and liquid water might also reside beneath the frozen surface of Saturn's satellite Enceladus.

Chapter 3 describes the atmospheres and magnetic fields that form an invisible buffer zone between planetary surfaces and surrounding space. Venus has an atmosphere that has run out of control, smothering this nearly Earth-sized world under a thick blanket of carbon dioxide. Its greenhouse effect has turned Venus into a torrid world that is hot enough to melt lead and vaporize oceans. Mars now has an exceedingly thin, dry and cold atmosphere of carbon dioxide. The red planet breathes about one-third of its atmosphere in and out as the southern polar cap grows and shrinks with the seasons. Jupiter's powerful winds and violent storms have remained unchanged for centuries, and Neptune has an unexpectedly stormy atmosphere. Saturn's largest moon, Titan, has a substantial Earth-like atmosphere, which is mainly composed of nitrogen and has a surface pressure comparable to that of the Earth's atmosphere. Temporary, rarefied and misty atmospheres cloak the Moon, Mercury, Pluto, Triton, and Jupiter's four largest moons.

Magnetic fields protect most of the planets from energetic charged particles flowing in the Sun's ceaseless winds, but some electrons and protons manage to penetrate this barrier. Jupiter's magnetism is the strongest and largest of all the planets, as befits the giant, while the magnetic fields of Uranus and Neptune are tilted. Guided by magnetic fields, energetic electrons move down into the polar atmospheres of Earth, Jupiter and Saturn, producing colorful auroras there.

Our description of individual planets begins in Chapter 4, with our home planet Earth. Earthquakes have been used to look inside our world, determining its internal structure and locating a spinning, crystalline globe of solid iron at its center. At the surface, continents slide over the globe, colliding and coalescing with each other like floating islands, as ocean floors well up from inside the Earth.

A thin membrane of air protects life on this restless world, and that air is being dangerously modified by life itself. Synthetic chemicals have been destroying the thin layer of ozone that protects human beings from dangerous solar ultraviolet radiation, and wastes from industry and automobiles are warming the globe to dangerous levels. The world has become hotter in the last decade than it has been for a thousand years, and at least some of this recent rise in temperature is due to greater emissions of greenhouse gases by human activity. The politicized debate over global warming is also described in Chapter 4, as are the probable future consequences if we don't do something about it soon.

This fourth chapter also discusses how the Sun affects our planet, where solar light and heat permit life to flourish. The amount of the Sun's radiation that reaches the Earth varies over the 11-year solar cycle of magnetic activity, warming and cooling the planet. Further back in time, during the past one million years, our climate has been changed by the recurrent ice ages, which are caused by variations in the amount and distribution of sunlight reaching the Earth.

An eternal solar gale now buffets our magnetic domain and sometimes penetrates it. Forceful mass ejections can create powerful magnetic storms on Earth, and damage or destroy Earth-orbiting satellites. Energetic charged particles, hurled out during solar explosions, endanger astronauts and can also wipe out satellites that are so important to our technological society. Space-weather forecasters are now actively searching for methods to predict these threats from the Sun.

In Chapter 5 we continue on to the still, silent and lifeless Moon, a stepping stone to the planets. Most of the features that we now see on the Moon have been there for more than 3 billion years. Cosmic collisions have battered the lunar surface during the satellite's formative years, saturating much of its surface with impact craters, while lunar volcanism filled the largest basins to create the dark maria.

Twelve humans went to the Moon more than three decades ago, and brought back nearly half a ton of rocks. The rocks contain no water, have never been exposed to it, and show no signs of life. Yet, orbiting spacecraft have found evidence for water ice deposited by comets in permanently shaded regions at the lunar poles.

The fifth chapter also describes how the Moon generates tides in the Earth's oceans, and acts as a brake on the Earth's rotation, causing the length of day to steadily increase. The satellite also steadies our seasons by limiting the tilt of Earth's rotation axis. The story of the Moon's origin is given the latest and most plausible explanation: a glancing impact from a Mars-sized object knocked a ring of matter out of the young Earth; that ring soon condensed into our outsized, low-density Moon.

We discover in Chapter 6 that Mercury has an unchanging, cratered and cliff-torn surface like the Moon, but in a brighter glare from the nearby Sun. Although the planet looks like the Moon on the outside, it resembles the Earth on the inside. Relative to its size, Mercury has the biggest iron core of all terrestrial planets, and it also has a relatively strong magnetic field. Here we also mention tiny, unexplained motions of Mercury. As demonstrated by astronomers long ago, the planet does not appear precisely in its expected place. This discrepancy led Einstein to develop a new theory of gravity in which the Sun curves nearby space.

Chapter 7 discusses veiled Venus, the brightest planet in the sky. No human eye has ever gazed at its surface, which is forever hidden in a thick overcast of impenetrable clouds made of droplets of concentrated sulfuric acid. Radar beams from the orbiting *Magellan* spacecraft have penetrated the clouds and mapped out the surface of Venus in unprecedented detail, revealing rugged highlands, smoothed-out plains, volcanoes and sparse, pristine impact craters. Rivers of outpouring lava have resurfaced the entire surface of Venus, perhaps about 750 million years ago, and tens of thousands of volcanoes are now found on its surface. Venus exhibits every type of volcanic edifice known on Earth, and some that have never been seen before. Some of them could now be active. Unlike Earth, there is no evidence for colliding continents on Venus, its surface moves mostly up and down, rather than sideways. Vertical motions associated with upwelling hot spots have buckled, crumpled, deformed, fractured and stretched the surface of Venus.

Our voyage of discovery continues in Chapter 8 to the red planet Mars, long thought to be a possible haven for life. Catastrophic flash floods and deep ancient rivers once carved channels on its surface, and liquid water might have lapped the shores of long-vanished lakes and seas. But its water is now frozen into the ground and ice caps, and it cannot now rain on Mars. Its thin, cold atmosphere lacks an ozone layer that might have protected the surface from lethal ultraviolet rays from the Sun, and if any liquid water were now released on the red planet's surface it would soon evaporate or freeze. Yet underground liquid water may have been seeping out of the walls of canyons and craters on Mars in recent times.

Three spacecraft have landed on the surface of Mars, failing to detect any unambiguous evidence for life. Corrosive chemicals have destroyed all organic molecules in the Martian ground, which means that the surface now contains no cells, living, dormant or dead. A meteorite from Mars, named ALH 84001, exhibits signs that bacteria-like micro-organisms could have existed on the red planet billions of years ago, but most scientists now think that there is nothing in the meteorite that conclusively indicates whether life once existed on Mars or exists there now. The future search for life on Mars may include evidence of microbes that can survive in hostile environments, perhaps energized from the planet's hot interior.

Chapter 9 presents giant Jupiter, which is almost a star and radiates its own heat. Jupiter radiates nearly twice as much energy as it receives from the Sun, probably as heat left over from when the giant planet formed. Everything we see on Jupiter is a cloud, formed in the frigid outer layers of its atmosphere. The clouds are swept into parallel bands by the planet's rapid rotation and counter-flowing winds, with whirling storms that can exceed the Earth in size. The fierce winds run deep and are driven mainly from within by the planet's internal heat. The biggest storms and wind-blown bands have persisted for centuries, though the smaller eddies are engulfed by the bigger ones, deriving energy from them. The little storms pull their energy from hotter, lower depths. Jupiter has a non-spherical shape with a perceptible bulge around its equatorial middle, and this helps us determine what is inside the planet. It is almost entirely a vast global sea of liquid hydrogen, compressed into a fluid metal at great depths. And above it all, Jupiter has a faint, insubstantial ring system that is made of dust kicked off small nearby moons by interplanetary meteorites.

Chapter 9 additionally provides up-to-date accounts of the four large moons of Jupiter, known as the Galilean satellites. The incredible complexity and rich diversity of their surfaces, which rival those of the terrestrial planets, are only visible by close-up scrutiny from spacecraft. Although the *Voyager 1* and *2* spacecraft sped by with just a quick glimpse at them, it was time enough for their cameras to discover new worlds as fascinating as the planets themselves, including active volcanoes on Io, smooth ice plains on Europa, grooved terrain on Ganymede, and the crater-pocked surface of Callisto. Then the *Galileo* spacecraft returned for a longer look, gathering further data on the satellites' surfaces and using gravity and magnetic measurements to infer their internal constitution. Changing tidal forces from nearby massive Jupiter squeeze Io's rocky interior in and out, making it molten inside and producing the most volcanically active body in the solar system. Jupiter's magnetic field sweeps past the moon, picking up a ton of sulfur and oxygen ions every second and directing them into a doughnut-shaped torus around the planet. A vast current of 5 million amperes flows between the satellite Io and the poles of Jupiter and back again, producing auroral lights on both bodies. There are no mountains or valleys on the bright, smooth, ice-covered surface of Europa. The upwelling of dirty liquid water or soft ice has apparently filled long, deep fractures in the crust. Large blocks of ice float like rafts across Europa's surface, lubricated by warm, slushy material. A subsurface ocean of liquid water may therefore lie just beneath Europa's icy crust, perhaps even harboring alien life that thrives in the dark warmth. Ganymede has an intrinsic magnetic field. As far as we know, it is the only satellite known that now generates its own magnetism. Callisto is one of the oldest, most heavily cratered surfaces in the solar system. Both Callisto and Europa have a borrowed magnetic field, apparently generated by electrical currents in a subsurface ocean as Jupiter's powerful field sweeps by.

Our voyage of discovery continues in Chapter 10 with Saturn, second only to Jupiter in size. Like Jupiter, the ringed planet radiates almost twice as much energy as it receives from the Sun, but Saturn is not massive enough to have substantial heat left over from its formation. Its excess heat is generated by helium raining down inside the planet. It is Saturn's fabled rings that set the planet apart from the other wanderers. The astonishing rings consist of billions of small, frozen particles of water ice, each in its own orbit around Saturn like a tiny moon. They have been arranged into rings within rings by the gravitational influences of small nearby satellites that generate waves, sweep out gaps and confine the particles in the rings. Saturn's rings are thought to be relatively young, less than 100 million years old. They may have originated when a former moon strayed too close to the planet and was torn apart by its tidal forces.

Saturn's largest satellite, Titan, has a substantial atmosphere composed mainly of nitrogen molecules, also the principal ingredient of Earth's air. Clouds of methane, raining ethane, and flammable seas of ethane, methane and propane could exist beneath the impenetrable haze. We should find out what lies beneath the smog when the *Cassini* spacecraft arrives at Saturn, in July 2004, and parachutes the *Huygens* probe through Titan's atmosphere four months later. Six medium-sized moons revolve around Saturn, each covered with water ice. They are scarred with ancient impact craters, and some of them show signs of ice volcanoes and internal heat. A number of small irregularly shaped moons of Saturn have remarkable orbits. The co-orbital moons move in almost identical orbits, the Lagrangian moons share their orbit with a larger satellite, and the shepherd moons confine the edges of rings.

Uranus and Neptune are treated together in Chapter 11, because of their similar size, mass and composition. Unlike all the other planets, Uranus is tipped on its side and rotates with its spin axis in its orbital plane and in the opposite direction to that of most of the other planets. No detectable heat is emitted from deep inside Uranus, while Neptune emits almost three times the amount of energy it receives from the Sun. This internal heat drives Neptune's active atmosphere, which has fierce winds and short-lived storms as big as the Earth. Both planets are vast global oceans, consisting mainly of melted ice with no metallic hydrogen inside. The magnetic fields of both Uranus and Neptune are tilted from their rotation axes, and are probably generated by currents in their watery interiors. The ring systems of both planets are largely empty space, containing dark narrow rings with wide gaps. One of Neptune's thin rings is unexpectedly lumpy, with material concentrated in clumps by a nearby moon. The rings we now see around these planets will eventually be ground into dust and vanish from sight, but they can easily be replaced by debris blasted off small moons already embedded in them. The amazingly varied landscape on Miranda, the innermost mid-sized satellite of Uranus, indicates that the satellite may have been shattered by a catastrophic collision and reassembled, or else it was frozen into an embryonic stage of development. Neptune's satellite Triton revolves about the planet in the opposite direction to its spin. The glazed satellite has a very tenuous, nitrogen-rich atmosphere, bright polar caps of nitrogen and methane ice, frozen lakes flooded by past volcanoes of ice, and towering geysers that may now be erupting on its surface. Triton may have formed elsewhere in the solar system and was captured into obit around Neptune. Triton is headed for a future collision with Neptune as the result of tidal interaction with the planet.

Chapter 12 discusses the icy comets. They light up and become visible for just a few weeks or months when tossed near the Sun, whose heat vaporizes the comet's surface and it grows large enough to be seen. A million, million comets are hibernating in the deep freeze of outer space, and they have been out there ever since the formation of the solar system 4.6 billion years ago. We can detect some of them in the Kuiper belt reservoir at the edge of the planetary system, but billions of unseen comets reside in the remote Oort cloud nearly halfway to the nearest star. Two spacecraft have now passed close enough to image a comet nucleus, of Comet Halley and Comet Borrelly, showing that they are just gigantic, black chunks of water ice, other ices, dust and rock, about the size of New York City or Paris. When these comets come near the Sun, their icy nuclei release about a million tons of water and dust every day, from fissures in their dark crust. Some comets develop tails that flow away from the Sun, briefly attaining lengths as large as the distance between the Earth and the Sun, but other comets have no tail at all. Comets can have two kinds of tails: the long, straight, ion tails, that re-emit sunlight with a faint blue fluorescence, and a shorter, curved, dust tail that shines by reflecting yellow sunlight. They are blown away from the Sun by its winds and radiation, respectively. Meteor showers, commonly known as shooting stars, are produced when sand-sized or pebble-sized pieces of a comet burn up in the Earth's atmosphere, never reaching the ground. Any comet that has been seen will vanish from sight in less than a million years, either vaporizing into nothing or leaving a black, invisible rock behind. Some burned-out comets look like asteroids, and a few asteroids behave like comets, blurring the distinction between these two types of small solar-system bodies.

We continue in Chapter 13 with the rocky asteroids. There are billions of them in the main asteroid belt, located between the orbits of Mars and Jupiter, but they are so small and widely spaced that a spacecraft may safely travel though the belt. The combined mass of billions of asteroids is less than five percent of the Moon's mass. The Earth resides in a smaller swarm of asteroids, chaotically shuffled out of the main belt. Many of these near-Earth asteroids travel on orbits that intersect the Earth's orbit, with the possibility of an eventual devastating collision with our planet. The asteroids are the pulverized remnants of former, larger worlds that failed to coalesce into a single planet. The colors of sunlight reflected from asteroids indicate that they formed under different conditions prevailing at varying distance from the Sun. We could mine some of the nearby ones for minerals or water. An asteroid's gravity is too weak to hold on to an atmosphere or to pull most asteroids into a round shape. The close-up view obtained by passing spacecraft and radar images indicates that asteroids have been battered and broken apart during catastrophic collisions in years gone by. One spacecraft has circled the near-Earth asteroid 433 Eros for a year, examining its dusty, boulder-strewn landscape in great detail, obtaining an accurate mass for the asteroid, and showing that much of it is solid throughout. Other asteroids are rubble piles, the low-density, collected fragments of past collisions held together by gravity. Meteorites are rocks from space that survive their descent to the ground, and most of them are chips off asteroids. Organic matter found in meteorites predates the origin of life on Earth by a billion years; but the meteoritic hydrocarbons are not of biological origin.

The concluding Chapter 14 discusses colliding worlds, including pieces of a comet that hit Jupiter, comets that are on suicide missions to the Sun, and an asteroid that wiped out the dinosaurs when it hit the Earth 65 million years ago. The Earth is now immersed within a cosmic shooting gallery of potentially lethal, Earth-approaching asteroids and comets that could collide with our planet and end civilization as we know it. The lifetime risk that you will die as the result of an asteroid or comet striking the Earth is about the same as death from an airplane crash, but a lot more people will die with you during the cosmic impact. It could happen tomorrow or it might not occur for hundreds of thousands of years, but the risk is serious enough that astronomers are now taking a census of the threatening ones. With enough warning time, we could redirect its course.

The Cambridge Guide to the Solar System continues with an annotated list of books for further reading, all published after 1990, and a list of Internet addresses for the topics discussed.

The illustrator Sue Lee has combined artistic talent with a scientist's eye for detail in producing the fantastic line drawings and diagrams in this book. The text has been substantially improved by the careful attention of copy-editor Brian Watts.

This book was stimulated by the author's visit to the Jet Propulsion Laboratory, when the main results of the recent planetary missions were summarized by its director, Edward C. Stone, and the Project Scientists of many of them. Andrew P. Ingersoll, Torrence V. Johnson, Kenneth Nealson, R. Stephen Saunders, Donald K. Yeomans and Richard W. Zurek provided comprehensive scientific summaries matched only by the extraordinary accomplishments of the missions themselves. Planetary scientists with comprehensive knowledge have assured the accuracy, completeness and depth of individual chapters through critical review. I am grateful to my expert colleagues who have read portions of this book, and substantially improved it, either by thorough review or by expert commentary

on some isolated sections. They include Reta Beebe, Doug Biesecker, Mark A. Bullock, Owen K. Gingerich, Torrence V. Johnson, Brian G. Marsden, Steven J. Ostro, Carl B. Pilcher, Roger A. Phillips, David Senske, Paul D. Spudis, David J. Stevenson and Donald K. Yeomans.

Kenneth R. Lang
Tufts University

Principal units

This book uses the International System of Units (Système International, SI) for most quantities, but with two exceptions. As is the custom with planetary scientists, we often use the kilometer unit of length and the bar unit of pressure. The familiar kilometer appears on most automobile speedometers. There are one thousand meters in a kilometer, and a mile is equivalent to 1.6 kilometers. One bar corresponds to the surface pressure of the Earth's air at sea level. For conversion to the SI pressure unit of pascals, 1 bar $= 10^5$ pascals, or 1 pascal $= 10^{-5}$ bar.

Some other common units are the millibar, equivalent to 0.001 bar, the nanometer (nm) with 1 nm $= 10^{-9}$ meters, the micron or micrometer (μm) with 1 μm $= 10^{-6}$ m, the ångstrom unit of wavelength, where 1 ångstrom = 1 Å $= 10^{-10}$ meters, the nanotesla (nT) unit of magnetic flux density, where 1 nT $= 10^{-9}$ tesla $= 10^{-5}$ gauss, and the ton measurement of mass, where 1 ton $= 10^3$ kilograms $= 10^6$ grams.

The reader should also be warned that centimeter-gram-second (c.g.s.) units have been, and still are, widely employed in astronomy and astrophysics. The table provides unit abbreviations and conversions between units.

Quantity	SI units	Conversion to c.g.s. units
Length	meter (m)	100 centimeters (cm)
Mass	kilogram (kg)	1000 grams (g)
Time	second (s)	
Temperature	kelvin (K)	
Velocity	meter per second ($m\,s^{-1}$)	100 centimeters per second ($cm\,s^{-1}$)
Energy	joule (J)	10 000 000 erg
Power	watt (W) = joule per second ($J\,s^{-1}$)	10 000 000 erg s^{-1} (= 10^7 erg s^{-1})
Magnetic flux density	tesla (T)	10 000 gauss (G) (= 10^4 G)
Force	newton (N) (= $kg\,m\,s^{-2}$)	100 000 dyn (= 10^5 dyn)
Pressure	pascal (Pa) (= $N\,m^{-2}$) (= $kg\,m^{-1}\,s^{-2}$)	10 dyn cm^{-2} (= 10^{-5} bar)

Part 1 Changing views and fundamental concepts

1 Evolving perspectives: a historical prologue

- The wandering planets move in a narrow track against the unchanging background stars, and some of these vagabonds can suddenly turn around, apparently moving in the opposite direction before continuing on their usual course.
- The ancient Greeks noticed that the Earth always casts a curved shadow on the Moon during a lunar eclipse, demonstrating that our planet is a sphere.
- For centuries, astronomers tried to describe the observed planetary motions using uniform, circular motions with the stationary Earth at the center and with the distant celestial sphere revolving about the Earth once a day.
- Around 145 AD, Claudius Ptolemy devised an intricate system of uniform motion around small and large circles to model the motions of the Sun, Moon and planets around a stationary Earth; his model was used to predict their location in the sky for more than a thousand years.
- The stars seem to be revolving around the Earth each night, but the Earth is instead spinning beneath the stars. This rotation also causes the Sun to move across the sky each day.
- Mikolaj Kopernik, better known as Nicolaus Copernicus, argued in 1543 that the Earth is just one of several planets that are whirling endlessly about the Sun, all moving in the same direction but at different distances from the Sun and with speeds that decrease with increasing distance.
- Almost four centuries ago, Johannes Kepler used accurate observations, obtained by Tycho Brahe, to conclude that the planets move in ellipses, or ovals, with the Sun at one focus, and to infer a precise mathematical relation between the mean orbital distance and period of each planet.
- More distant planets take longer to move once around the Sun and they move with slower speeds; their orbital periods are in proportion to the cubes of their distances.
- Astronomy is an instrument-driven science in which novel telescopes and new technology enable us to discover cosmic objects that are otherwise invisible and hitherto unknown.

- Many major astronomical discoveries have been unanticipated and serendipitous, made while new telescopes were used to study other, known cosmic objects; the earliest of these accidental discoveries include the four large moons of Jupiter, the planet Uranus, and the first known asteroid, Ceres, discovered respectively by Galileo Galilei in 1610, William Herschel in 1781, and Giuseppe Piazzi in 1801.
- The asteroid belt between the orbits of Mars and Jupiter contains more than 500 000 asteroids, but it is largely empty space and has a total mass that is much less than that of the Earth's Moon.
- Two kinds of telescopes, the refractor and the reflector, enable astronomers to detect faint objects that cannot be seen with the unaided eye, and to resolve fine details on luminous planets that otherwise remain blurred.
- Jupiter, Saturn and Uranus have a retinue of large satellites, and Neptune has only one really large moon that moves in the opposite direction to all the other large satellites. Mercury and Venus have no moons, the Earth has one satellite, our Moon, and Mars has two very small ones.
- Christiaan Huygens discovered Saturn's rings in 1659; they are completely detached from the planet and consist of innumerable tiny satellites each with an independent orbit about Saturn.
- In his *Principia*, published in 1686, Isaac Newton showed how the laws of motion and universal gravitation describe the movements of the planets and everything else in the Universe.
- The solar system is held together by the Sun's gravitational attraction, which keeps the planets in their orbits; they move at precisely the right speed required to just overcome the pull of solar gravity.
- The gravitational attraction between two objects increases in proportion to the product of their masses and in inverse proportion to the square of the distance between them.
- The planet Neptune was discovered in 1846, near the location predicted by mathematical calculations under the assumption that the gravitational pull of a large, unknown world, located far beyond Uranus, was causing the observed positions of Uranus to deviate from its predicted ones.
- Estimates for the mean Earth–Sun distance, known as the astronomical unit or AU, were gradually refined over the centuries, eventually setting the scale of the solar system at $1\,\text{AU} = 149.6$ million kilometers. At this distance, it takes 499 seconds for light to travel from the Sun to the Earth.
- The nearest star other than the Sun is located at a distance of 4.24 light-years; it is about 270 000 times further away from the Earth than the Sun.
- The Sun is the most massive and largest object in our solar system. The Sun's mass, which is 333 000 times the Earth's mass, can be inferred from Kepler's third law using the Earth's orbital period of one year and the Earth's mean distance from the Sun, the AU.
- The Sun's size, at 109 times the diameter of the Earth, can be inferred from the Sun's distance and angular extent.
- The temperature of the Sun's visible disk is 5780 kelvin; it can be determined from the Sun's total irradiance of the Earth, the Earth–Sun distance or the AU, and the radius of the Sun.

- The temperature at the center of the Sun is 15.6 million kelvin, estimated from the speed a proton must be moving to counteract the gravitational compression of the massive Sun.
- The composition of the Sun is encoded in absorption lines that appear in the visible spectrum of sunlight.
- The lightest element, hydrogen, is the most abundant element in the Sun, and the next most abundant solar element, helium, was first discovered in the Sun.
- The regular spacing of hydrogen's spectral lines can be explained by quantum theory, in which the angular momentum and energy of an orbiting electron are quantized, depending on an integer quantum number.
- The eight major planets can be divided into two groups: the four rocky, dense, terrestrial planets, Mercury, Venus, Earth and Mars, located relatively near the Sun, and the four giant, low-density planets, Jupiter, Saturn, Uranus and Neptune, that are further from the Sun.
- The temperature and density increase systematically with depth in the giant planets, owing to the greater compression by overlying material.
- As the result of differentiation in their originally molten interiors, the rocky terrestrial planets contain dense iron cores surrounded by less-dense silicate mantles.
- The terrestrial planets contain partially molten, liquid cores, but their internal temperatures cool as time goes on due to the depletion of radioactive elements and the emission of internal heat.

1.1 Moving points of light

The ancient wanderers

Our remote ancestors spent their nights under dark skies, becoming intimately familiar with the stars. They looked up on any moonless night, and watched thousands of stars embedded in the black dome of the night, ceaselessly moving from one edge of the Earth to overhead and back down to another edge, night after night without end.

The brightest stars received names, and patterns, now called constellations, were noticed among groups of them. These permanent stellar beacons are always there, firmly rooted in the dark night sky, and the constellations remain unchanged over the eons.

As ancient astronomers watched the stars, they focused attention on seven objects that did not move with the stars. These celestial vagabonds changed position on the sphere of background stars from hour to hour or night to night, and unlike the stars, they would appear in the night sky at different times from year to year. Ranked in order of greatest apparent brightness, they are the Sun, Moon, Venus, Jupiter, Saturn, Mercury and Mars, each with the Latinized name of a Greek god or goddess. Our ancestors called them *planetes*, the ancient Greek word for "wanderers"; and the designation planet is still used for all but the Sun and Moon.

The Sun and Moon move with a rhythm, pattern and beat, marking out the time of our first clocks. The rising and setting Sun ticked off the days, the Moon's changing phase set the monthly cycle, and the seasons marked off the years.

The Sun does not rise at precisely the same point on the horizon each day. Instead, the location of sunrise drifts back and forth along the horizon in an annual cycle. Ancient astronomers used monuments to line up the limits of these excursions (Fig. 1.1). The length and height of the Sun's arc across the sky also change with a yearly rhythm. In the northern hemisphere, the Sun rises highest in the sky on the summer solstice, around June 21 each year, with its longest trajectory and the most daylight hours (Fig. 1.2).

Like the Sun, the Moon rises and sets at different points along the horizon, and reaches varying heights in the sky. Since the full Moon always lies nearly opposite to the Sun, the winter full Moon rises much higher in the sky than the summer full Moon.

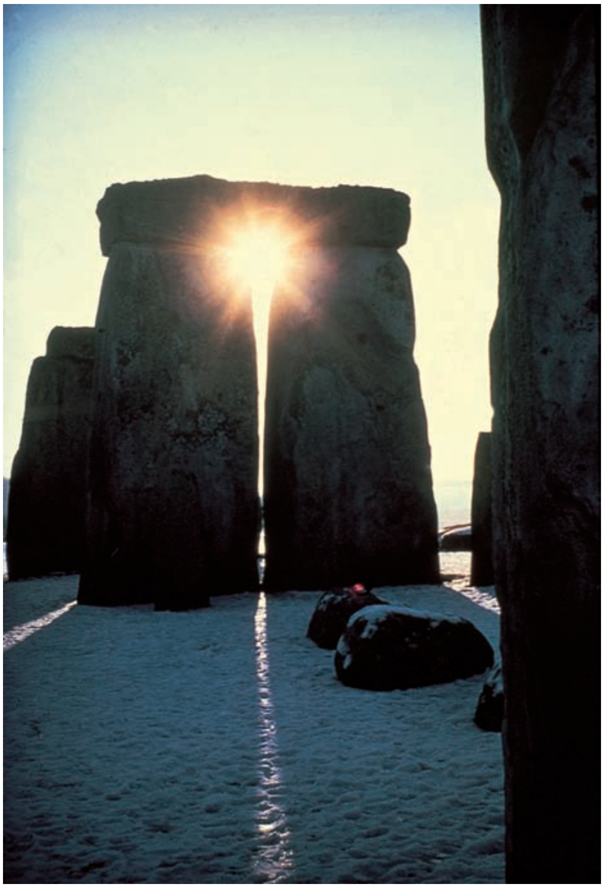


Fig. 1.1 Stonehenge The ancient stone pillars of Stonehenge in southern England, shown in this photograph, frame the rising Sun. This monument was used to find midsummer and midwinter 4000 years ago – before the invention of writing and the calendar. The Sun rises at different points on the horizon during the year, reaching its most northerly rising on Midsummer Day (summer solstice on 21 June). After this, the rising point of the Sun moves south along the horizon until it reaches its most southerly rising on Midwinter Day (winter solstice on 22 December). An observer located at the center of the main circle of stones at Stonehenge watched midsummer sunrise over a marker stone located outside the circle; other stones within the circle framed midwinter sunrise and sunset. (Courtesy of Owen Gingerich.)

The Moon repeats its motion around the Earth on a monthly cycle, periodically changing its appearance (Fig. 1.3). Once each month, the Moon comes nearly in line with the Sun, vanishing into the bright daylight. On the next night the Moon has moved away from this position, and a thin lunar crescent is seen. The crescent thickens on successive nights, reaching the rotund magnificence of full Moon in two weeks. Then, in another two weeks, the Moon disappears into the glaring Sun, completing the cycle of the month and providing another natural measure of time.

Even the earliest sky-watchers must have noticed that the wanderers are confined to a narrow track around the sky. Babylonian astronomers noticed it thousands of years ago, identifying constellations that lay along its path. Twelve of these constellations subsequently became known as the zodiac, from the Greek word for "animal". The Sun's annual path, called the ecliptic, runs along the middle of this celestial highway, and the paths of all the other wanderers lie within it. Its narrowness is a sign that the planets move almost like marbles on a table because the planes of their orbits are closely aligned with each other.

It was obvious to astronomers from the earliest times that the wanderers do not move at uniform speeds or follow simple paths across the sky. Mars apparently moved in a backwards loop for weeks at a time, seemingly disrupting its uniform progress across the night sky. It gradually came to a stop in its eastward motion, moved backward toward the west, and then turned around again and resumed moving toward the east (Fig. 1.4). Jupiter and Saturn also displayed such a temporary backwards motion in the westward retrograde direction before continuing on in the eastward prograde direction.

But why did these planets behave in such an unusual and singular manner? The ancient Greeks first proposed logical explanations, based on geometry and uniform motion, but modern explanations differ in both the locations and motions of the planets.

Circles and spheres

The ancient Greeks used geometrical models to visualize the cosmos, incorporating the symmetric forms of the circle and sphere. Their aim was to describe the regularities that underlay the planetary motions against the unchanging stellar background, and to thereby predict the locations of the planets at later times. They wanted to provide a reliable guide to the future, which is still one of the main points of science.

In arguments used by the Pythagoreans, and subsequently recorded by Aristotle (384–322 BC), it was shown that the Earth is a sphere. During a lunar eclipse, when the Moon's motion carries it through the Earth's shadow, observers at different locations invariably saw a curved shadow on the Moon (Fig. 1.5). Only a spherical body can cast a round shadow in all orientations. The curved surface of the ocean was also inferred by watching a ship disappear over the horizon; first the hull and then the mast disappear from view.

According to Plato, writing around 380 BC, the simplest and purest sort of motion was circular, so circles ought to describe the visible paths of the moving planets.

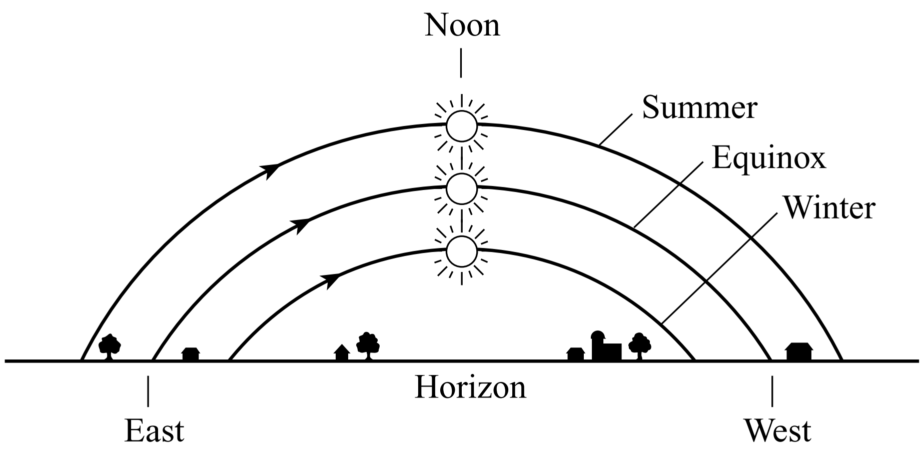


Fig. 1.2 The Sun's trajectory The Sun's motion across the sky as seen from the northern hemisphere. The maximum height of the Sun in the sky, and the Sun's rising and setting points on the horizon, change with the seasons. In the summer, the Sun rises in the northeast, reaches its highest maximum height, and stays up longest. The Sun rises southeast and remains low in the winter when the days are shortest. The length of day and night are equal on the Vernal, or Spring, Equinox (March 20) and on the Autumnal Equinox (September 23) when the Sun rises exactly east and sets exactly west.

After all, a wheel moves so well because it is round, without rough, sharp edges to get in the way. The circle also has no beginning or end, seemingly appropriate for describing the endless motion of the heavenly wanderers. And the central Earth would be separated from the heavens, like a magician who draws a boundary circle around him to seal off the region in which magical powers are brought into play.

Following Plato's suggestion, astronomers spent centuries trying to discover those uniform, perfectly regular, circular motions that would "save the appearances" presented by the planets. They supposed that the Earth stood still, an immobile globe at the center of it all. The imaginary celestial sphere of fixed stars wheeled around the central, stationary Earth once every day, with uniform circular motion and perfect regularity, night after night and year after year. Such a celestial sphere would also explain why people located at different places on Earth invariably saw just half of all the stellar heavens, and why travelers to new and distant lands would see new stars as well as new people.

The Sun, Moon and planets were once supposed to be carried on concentric, transparent crystalline spheres, which revolved around the stationary Earth, but their hypothetical uniform and circular motions contradicted observations. The Earth-centered model did not explain, for example, why each planet moved with changing speed across the sky, not at an unchanging, uniform rate.

So the Egyptian astronomer Claudius Ptolemy (fl. 150 AD) shifted the Earth from the exact center of the Universe by just a small amount, and described the planetary appearances with an intricate system of circles moving on other circles, like the gears of some fantastic cosmic machine. A planet in uniform circular motion about a center offset from the Earth would appear to a terrestrial observer to be moving with varying speed, faster when it is closest to Earth and slower when further away. Combinations of uniform circular motion were additionally required to account for the looping, or retrograde, paths of the planets (Fig. 1.6). Each planet was supposed to move with constant speed on a small circle, or epicycle, while the center of the epicycle rotated uniformly on a large circle, or deferent. In this way Ptolemy, in his *Mathematical Compilations*, or *Almagest*, written about 145 AD, was able to predict the motions of every one of the seven wanderers, compounding them from circles upon circles. By selecting suitable radii and speeds of motion, Ptolemy reproduced the apparent motions of the planets with remarkable accuracy. He succeeded so well that his model was still being used to predict the locations of the planets in the sky more than a thousand years after his death.

The Earth moves

The ancient Indians of Asia had a different point of view, supposing that the Earth moves around the Sun, as did the Greek mathematician and astronomer Aristarchos, born on the island of Samos in 310 BC. Aristarchos moved the center of the Universe from the Earth to the Sun, and set the Earth in motion, supposing that the Earth and other planets travel in circular orbits around the stationary Sun. He further stated that the fixed stars do not move, and that their apparent daily motion is due to the Earth's rotation on its axis.

Waxing
Crescent

First
Quarter

Waxing
Gibbous

Full
Moon

Waning
Gibbous

Third
Quarter

Waning
Crescent

Fig. 1.3 The Moon's varying appearance During the monthly cycle, the illuminated part of the Earth's Moon waxes (grows) from crescent to gibbous, and then after full Moon, it wanes (decreases) to a crescent again. The term crescent is applied to the Moon's shape when it appears less than half-lit; it is called gibbous when it is more than half-lit but not yet fully illuminated. The reason for the Moon's changing shape is described in Fig. 1.8. (Lick Observatory Photographs.)

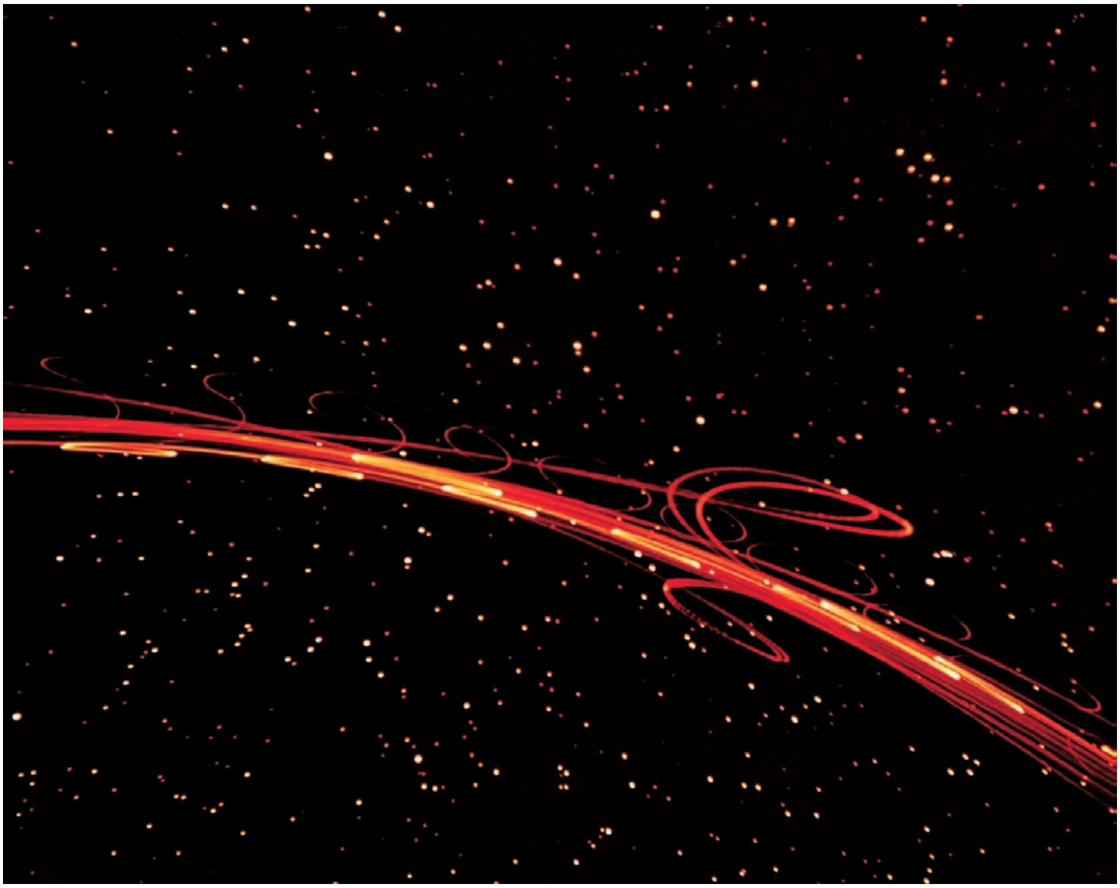


Fig. 1.4 Retrograde loops This photograph shows the apparent movements of the planets against the background stars. Mars, Jupiter and Saturn appear to stop in their orbits, then reverse direction before continuing on – a phenomenon called retrograde motion by modern astronomers. Ancient and modern explanations for this temporary backward motion are illustrated in Figs. 1.6 and 1.9, respectively. (Courtesy of Erich Lessing/Magnum.)

Fig. 1.5 Curved shadow of Earth This multiple-exposure photograph of a total lunar eclipse reveals the curved shape of the Earth's shadow, regarded by ancient Greek astronomers as evidence that the Earth is a sphere. Only a spherical body will cast the same circular shadow on the Moon when viewed from different locations on Earth or during different lunar eclipses. This photograph was taken by Akira Fujii during the lunar eclipse of 30 December 1982.

As we now know, Aristarchos was right. The stars seem to be revolving about the Earth each night, but appearances can be deceiving. The Earth could instead be spinning beneath the stars. As the Earth rotates, the stars slide by, accounting for the wheeling night sky, which just seems to be revolving.

And the Sun might not be moving across the bright blue sky each day, for the Earth's rotation could produce this motion. Every point on the surface of a spinning Earth can be carried across the line of sight to an unmoving Sun, from sunrise to sunset, producing night and day (Fig. 1.7). Since the Earth rotates from west to east, the Sun appears

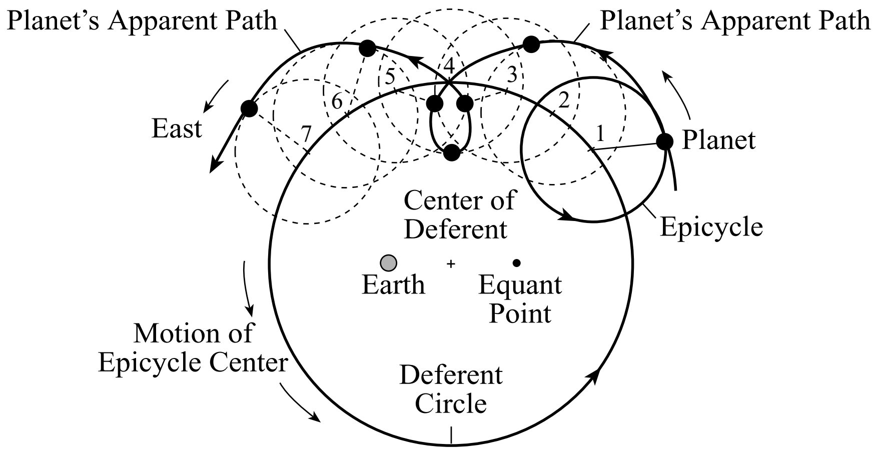


Fig. 1.6 Circles upon circles to explain retrograde loops To explain the occasional retrograde loops in the apparent motions of Mars, Jupiter and Saturn, astronomers in ancient times imagined that each planet travels with uniform speed around a small circle, known as the epicycle. The epicycle's center moves uniformly on a larger circle, the deferent. A similar scheme was used by Ptolemy (fl. 150 AD) to explain the wayward motions of the planets in his *Almagest*. In the Ptolemaic system, the Earth was displaced from the center of the large circle, and each planet traveled with uniform motion with respect to another imaginary point, the equant, appearing to move with variable speed when viewed from the Earth.

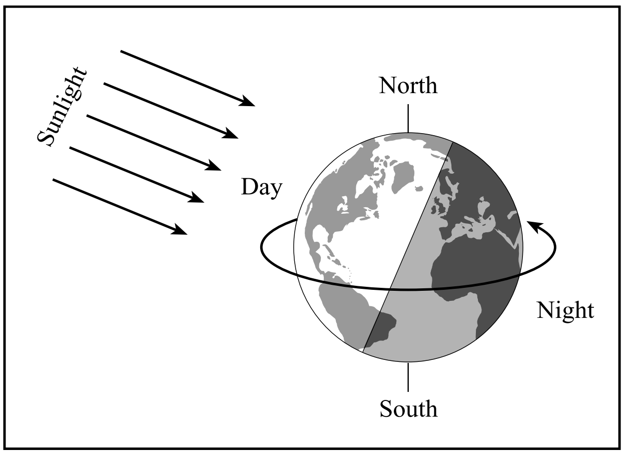


Fig. 1.7 Night and day The Earth rotates with respect to the Sun once every 24 hours, causing the sequence of night and day. Each point on the Earth's surface moves in a circular track parallel to the equator, and each track spends a different time in the Sun depending on the season. This drawing depicts summer in the northern hemisphere and winter in the southern hemisphere. Because the northern part of the Earth's rotational axis is tipped toward the Sun, circular tracks in the northern hemisphere spend a longer time in the Sun than southern ones.

to rise in the east and set in the west. Such a perspective involves a certain amount of detachment – the ability to separate yourself from the ground and use your mind's eye to look down on the spherical, rotating Earth, like a spinning ball suspended in space.

The Moon's motion from horizon to horizon each night could also be neatly explained by the rotation of the Earth, and the Moon's monthly circuit against the background stars could be ascribed to its slower orbital motion around the Earth. This would also account for the Moon's varying appearance (Fig. 1.8). The Moon borrows its light from the Sun, and the Sun illuminates first one part of the Moon's face and then another as the Moon orbits the Earth. On any given night, all observers on Earth will see the same phase of the Moon as our planet's rotation brings it into view.

The concept of a moving Earth nevertheless seems to violate common sense. The ground certainly seems to be at rest beneath our feet, providing the terra firma on which we carry out our daily lives. As Aristotle noticed, an arrow shot vertically upward falls to the ground where the archer stands, suggesting that the ground has not moved while the arrow was in flight. Moreover, if the Earth is rotating, then its surface regions have to be moving at high speeds (Focus 1.1).

Yet the globe on which we live might not only spin on its axis; it could also be whirling ceaselessly around the Sun, completing one circuit each year as Aristarchos had supposed. But his proposals had little impact on his contemporaries. It took another eighteen centuries before the Polish cleric and astronomer Mikolaj Kopernik (1473–1543), better known as Nicolaus Copernicus, revived this heliocentric, or Sun-centered, model. By 1514 Copernicus was privately circulating a manuscript, the *Commentariolus*, or *Little Commentary*, in which the planets were placed in uniform motion about a central Sun. His longer, more influential book, *De Revolutionibus Orbium Coelestium Libri VI*, or *Six Books Concerning the Revolutions of the Celestial*

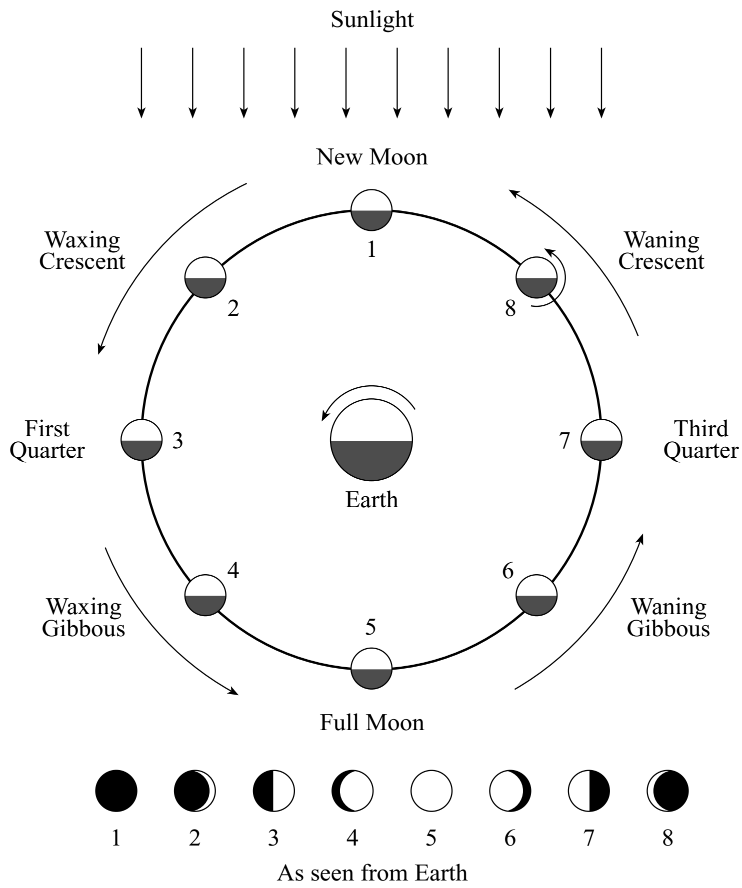


Fig. 1.8 Phases of the Moon Light from the Sun illuminates one half of the Moon, while the other half is dark. As the Moon orbits the Earth, we see varying amounts of its illuminated surface. The phases seen by an observer on Earth (*bottom*) correspond to the numbered points along the lunar orbit. The period from new Moon to new Moon is 29.53 days, the length of the month. As the Earth completes its daily rotation, all night-time observers see the same phase of the Moon.

Focus 1.1 Location and rotation speed on the Earth

The length of the day and the rotation period is the same for every place on Earth, but the speed of rotation around its axis depends on the surface location. A grid of great circles on the spherical Earth defines this location. A great circle divides the sphere in half, and the name comes from the fact that no greater circles can be drawn on a sphere. A great circle halfway between the North and South Poles is called the equator, because it is equally distant between both poles. Circles of longitude are great circles that pass around the Earth from pole to pole perpendicular to the equator, with 0 degrees at the Prime Meridian that passes through the Royal Observatory in Greenwich, England. The latitude is the angle measured northward (positive) or southward (negative) along a circle of longitude from the equator to the point.

The surface speed of rotation is greatest at the equator and reduces to almost nothing at the poles. Using an equatorial radius of about 6378 kilometers, which is close to the value inferred long ago (by Eratosthenes about 200 BC), the Earth would have to be rotating at a velocity of about 460 meters per second to spin about its equatorial circumference once every 24 hours. To calculate this speed, just multiply the equatorial radius by 2π to get the equatorial circumference, and divide by 24 hours and 3600 seconds per hour. At higher latitudes, closer to the poles, the circumferential distance around the Earth, and perpendicular to a great circle of longitude, is less, so the speed is less. The speed diminishes to almost nothing at the geographic poles, which are pierced by the rotation axis.

Bodies, was published almost thirty years later, in 1543, the year of its author's death.

For Copernicus, the Sun was located at the heart of the planetary system and the center of the Universe. The only thing to orbit the Earth was the Moon, and the Earth was supposed to rotate on its axis to make the stars swing by. In this model, the Earth was just one of several planets revolving around the Sun, in the same direction but at different distances and with various speeds, always passing each other without ever intersecting. In order of increasing

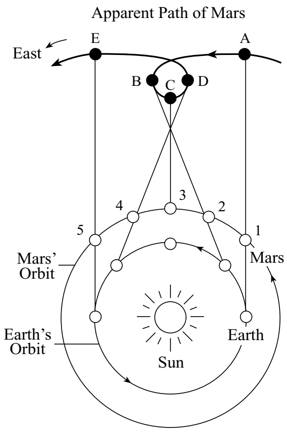


Fig. 1.9 Retrograde loops in a Copernican Universe A Sun-centered model of the solar system explains the looping path of Mars in terms of the relative speeds of the Earth and Mars. The Earth travels around the Sun more rapidly than Mars does. As Earth overtakes and passes the slower moving planet (*points 2 to 4*), Mars appears to move backward (*points B to D*) for a few months.

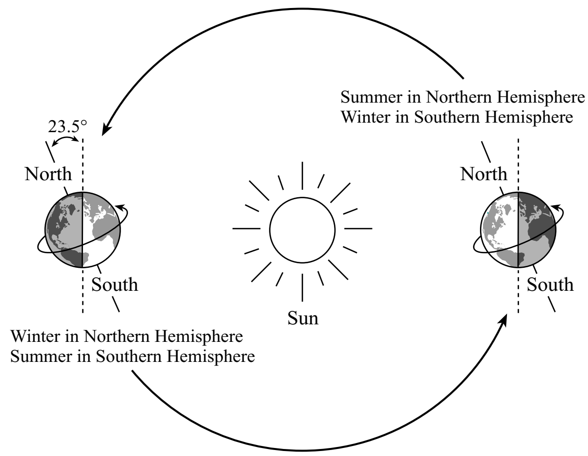


Fig. 1.10 The seasons As the Earth orbits the Sun, the Earth's rotational axis in a given hemisphere is tilted toward or away from the Sun. This variable tilt produces the seasons by changing the angle at which the Sun's rays strike different parts of the Earth's surface. The greatest sunward tilt occurs in the summer when the Sun's rays strike the surface most directly. In the winter, the relevant hemisphere is tilted away from the Sun and the Sun's rays obliquely strike the surface. When it is summer in the northern hemisphere, it is winter in the southern hemisphere and vice versa. (Notice that the radius of the Earth and Sun and the Earth's orbit are not drawn to scale.)

distance from the Sun, they are Mercury, Venus, Earth, Mars, Jupiter and Saturn. As Copernicus noticed, the further a planet is from the Sun, the longer it takes to complete a circuit.

There was no definite proof of this heliocentric hypothesis; but it did provide natural explanations for observed phenomena. Venus and Mercury, for example, are never to be seen far from the Sun. They rise and set with the Sun, unlike Mars, Jupiter and Saturn, which can be seen at any time of night. Since the orbits of Venus and Mercury lie inside that of Earth and closer to the Sun, these planets are only seen around dawn or dusk. In contrast, the orbits of Mars, Jupiter and Saturn lie outside that of the Earth, so they are visible throughout the night.

The Sun-centered view also provides a simple explanation of the retrograde motions that were so hard to reproduce using an Earth-centered, or geocentric, model. The jerky backwards motions were attributed to the uniform motion of the Earth and other planets at different speeds around the Sun. Planets moving at a slower speed than the Earth would sometimes appear to move ahead of Earth, and sometimes fall behind (Fig. 1.9).

Most of the time we see Mars, Jupiter and Saturn moving around the Sun in the same direction as the Earth, but during the relatively short time that the Earth overtakes one of these planets, that planet appears to be moving backward (Fig. 1.9). Moreover, one could confidently predict when a planet's apparent motion would come to a halt and turn around, and for how long it would seem to move backwards.

We now realize that the tilt of the Earth's rotational axis and the annual orbit of the Earth cause sunlight to fall differently on our planet at different times of year, explaining the seasons (Fig. 1.10). As the Earth orbits the Sun, its rotational axis remains pointed in the same direction in space, toward the North Star Polaris, so the orientation of the rotation axis in space remains unchanged throughout the year. But the orientation of the axis toward the Sun changes over the course of each orbit when the northern and southern hemispheres are tilted toward or away from the Sun by up to 23.5 degrees. The greatest sunward tilt in a given hemisphere occurs in summer when the Sun is more nearly overhead and its rays strike the surface more directly. Winter occurs half an orbit later, when that hemisphere is at its greatest tilt away from the Sun. Notice that the seasons are caused by the change in tilt, toward or away from the Sun, and not by any noticeable change in distance from the Sun.

The semi-annual alteration in incident sunlight is less pronounced in the equatorial regions, where the seasonal weather changes are not as great as they are at higher latitudes. People living near the equator have rainy and dry seasons, with the rains coming when the Sun is higher in the sky.

Copernicus' goal was to provide a geometric model that could replicate the planetary motions, but transforming their center to the Sun did not by itself improve the predictions. Proof of his Sun-centered model required improved

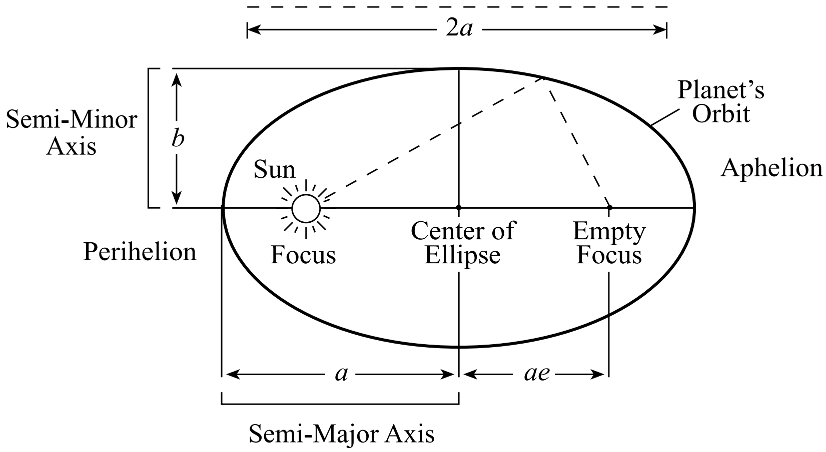


Fig. 1.11 Elliptical orbit Each planet moves in an ellipse with the Sun at one focus. The length of a line drawn from the Sun, to a planet and then to the empty focus, denoted by the dashed line, is always $2a$, or twice the semi-major axis, a. The eccentricity, or elongation, of the planetary ellipse has been greatly overdone in this figure; planetary orbits look much more like a circle.

observations and the introduction of non-circular motions. Yet Copernicus' book did become a symbol for a new perspective of the heavens, a view that was ultimately to unite the Earth and planets in the domain of terrestrial physics. He opened the way to the study of not only how the celestial bodies move, but to an investigation of the forces that propel them and the underlying laws that govern their motion.

The harmony of the world

In the hope of developing a more precise description of planetary motions, the Danish astronomer Tycho Brahe (1546–1601) built, with royal patronage, an observatory, Uraniborg, on the island of Hven, where he amassed a great number of observations that were more accurate and complete than any previous ones, including detailed records of the orbit of Mars. This was before the days of telescopes, and he used ingenious measuring instruments that resembled large gun sights with graduated circles. Johannes Kepler (1571–1630), Tycho's assistant and successor after his death, eventually interpreted these data and was able to determine precise mathematical laws from them.

Since circular motions could not describe Tycho's accurate observations, Kepler concluded that non-circular shapes were required. In 1605, after four years of computations, Kepler found that the observed planetary orbits could be described by ellipses, or ovals, with the Sun at one focus (Fig. 1.11; Focus 1.2). This ultimately became known as Kepler's first law of planetary motion. A planet also speeds up when it approaches the Sun, and slows down when it moves away from the Sun, and that accounts for a planet's varying speed when observed from Earth; this is described when the modern concept of conservation of angular momentum is applied to elliptical orbits (Focus 1.2).

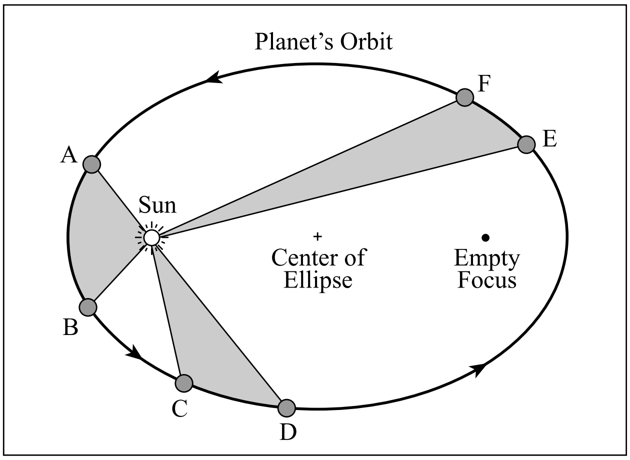


Fig. 1.12 Kepler's first and second laws Kepler's first law states that the orbit of a planet about the Sun is an ellipse with the Sun at one focus. The other focus of the ellipse is empty. According to Kepler's second law, the line joining a planet to the Sun sweeps out equal areas in equal times. This is also known as the law of equal areas. It is represented by the equality of the three shaded areas ABS, CDS and EFS. It takes as long to travel from A to B as from C to D and from E to F. A planet moves most rapidly when it is nearest the Sun (at perihelion); a planet's slowest motion occurs when it is farthest from the Sun (at aphelion).

Kepler was able to describe a planet's changing speed along its orbit in a precise mathematical form that can be explained with the help of Fig. 1.12. Imagine a line drawn from the Sun to a planet. As the planet swings about its elliptical path, the line (which will increase and decrease in length) sweeps out a surface at a constant rate. This is Kepler's second law of planetary motion, also known as the law of equal areas. During the three equal time intervals shown in Fig. 1.12, the planet moves through different arcs because its orbital speed changes, but the areas swept out are equal.

Focus 1.2 Elliptical planetary orbits

According to Kepler's first law, the planets move in elliptical orbits with the Sun at one focus. The planet's closest point to the Sun, when the planet moves most rapidly, is called the perihelion; and its most distant point is the aphelion, where the planet moves most slowly. The distance between the perihelion and aphelion is the major axis of the orbital ellipse. Half that distance is called the semi-major axis, designated by the symbol a.

The semi-major axis of the Earth's elliptical orbit about the Sun is called the astronomical unit, abbreviated AU. It sets the scale of the solar system, and when combined with the Earth's year-long orbital period permitted the determination of the Sun's mass and the Earth's orbital velocity, but only after astronomers had found out how large an AU is.

The shape of an ellipse is determined by its eccentricity, e. If $e = 0$ its shape is a circle. The ellipse becomes more elongated and squashed as its eccentricity increases toward $e = 1.0$. The eccentricity of the planetary ellipse has been greatly exaggerated in Fig. 1.11, with an eccentricity of about $e = 0.5$.

At perihelion the distance between the planet and the Sun is $a(1 - e)$ and at aphelion that distance is $a(1 + e)$. With the exception of Mercury, all of the major planets have orbits that are nearly circular, with eccentricities of less than $e = 0.1$. This means that the Sun is very near the center of each orbital ellipse. For Mercury, $a = 0.387$ AU and $e = 0.206$, so its distance from the Sun is just 0.307 AU at perihelion and quite a lot greater at aphelion, located at 0.467 AU.

One of the great laws of physics, known as the conservation of angular momentum, explains why a planet keeps on whirling around the Sun, and why its speed is fastest at perihelion. For a planet of mass M, orbiting the Sun with speed or velocity V at a distance D:

$$\text{Angular momentum} = M \times V \times D$$

By the way, in physics velocity has both a magnitude and a direction, and speed is the magnitude of the velocity. In astronomy the velocity is often just given by its magnitude, the speed, so the orbital velocity is its speed along the orbit.

The conservation law says that as long as no outside force is acting on the planet, its angular momentum cannot change. So the planet just keeps on moving along without anything pushing or pulling it. The mass does not change, so when the distance from the Sun decreases, at perihelion, the velocity increases to compensate and keep the angular momentum unchanged; at aphelion the distance increases so the speed has to decrease.

Kepler labored another decade before publication of *Harmonice mundi*, or *Harmony of the World*, in 1619, where he claimed to have listened to, and described mathematically, the music of the heavenly spheres. Kepler investigated arithmetic patterns between the periods and sizes of the planetary orbits, discovering the harmonic relation that is now known as Kepler's third law. It states that the squares of the planetary periods are in proportion to the cubes of their average distances from the Sun.

These periods and distances are given in Table 1.1 with other mean orbital parameters of the major planets. Here the periods are given in units of the Earth's orbital period of one Earth year, and the distances from the Sun are specified in units of the Earth's mean distance from the Sun, the astronomical unit (AU). One Earth year is equivalent to 31.557 million (3.1557×10^7) seconds; a precise measurement for the length of 1 AU, which is equal to 149.598 million ($1.495\,98 \times 10^8$) kilometers, took centuries to determine. The orbital velocities are given in kilometers per second ($\mathrm{km\,s^{-1}}$).

If P_P denotes the orbital period of a planet measured in Earth years, and a_P describes its semi-major axis (Focus 1.2) measured in AU, then Kepler's third law states that $P_P{}^2 = a_P{}^3$, where the subscript "P" denotes the planet under consideration. This expression is illustrated in Fig. 1.13 for the major planets and for the brighter moons of Jupiter. The mean orbital velocity of each planet is proportional to the ratio a_P/P_P, so the velocity varies inversely with the square root of the distance, or as $(a_P)^{-1/2}$.

In other words, the more distant planets have longer orbital periods and they move around the Sun with a slower speed. For example, Jupiter is 5.2 times as far away from the Sun as the Earth is, and it takes Jupiter 11.86 Earth years to travel once around the Sun. The Earth's mean orbital velocity is nearly 30 kilometers per second, while Jupiter's orbital velocity is about half that amount. Both planets are whirling around the Sun with awesome speed.

1.2 Telescopes reveal the hitherto unseen

Astronomers use telescopes to extend their vision to otherwise invisible cosmic objects. Without a telescope, for example, you can't see Neptune, or any of the numerous asteroids and planetary satellites – other than the Earth's

Table 1.1 Mean orbital parameters of the major planets[a]

Planet	Semi-major axis, a_P (AU)	Orbital period, P_P (years)	Eccentricity, e	Inclination to the ecliptic, i (degrees)	Orbital velocity ($km\ s^{-1}$)
Mercury	0.387 099	0.2409	0.2056	7.00	47.890
Venus	0.723 332	0.6152	0.0068	3.39	35.030
Earth	1.000 000	1.0000	0.0167	0.01	29.790
Mars	1.523 688	1.8809	0.0933	1.85	24.130
Jupiter	5.202 834	11.8622	0.048	1.31	13.060
Saturn	9.538 762	29.4577	0.056	2.49	9.640
Uranus	19.191 391	84.0139	0.046	0.77	6.810
Neptune	30.061 069	164.793	0.010	1.77	5.430

[a] The dashed line divides the six planets known in Kepler's time from the two major outer planets discovered later.

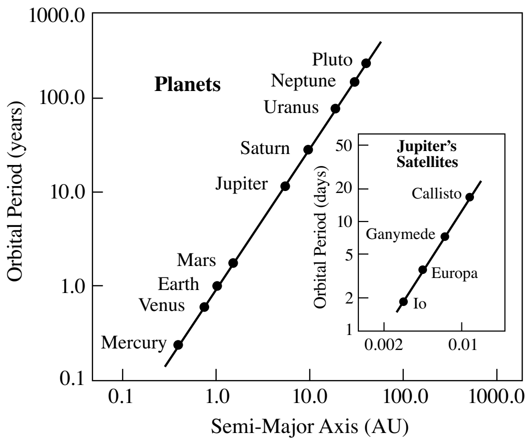


Fig. 1.13 Kepler's third law The orbital periods of the planets are plotted against their semi-major axes, using a logarithmic scale. The straight line that connects the points has a slope of 3/2, thereby verifying Kepler's third law that states that the squares of the orbital periods increase with the cubes of the planetary distances. This type of relation applies to any set of bodies in elliptical orbits about a much larger mass, including Jupiter's four largest satellites shown in the inset.

Moon. When we step outside to look at the dark night sky, our eyes can detect only about 3000 of the 100 billion stars in the Milky Way, and billions of galaxies similarly remain invisible.

The historical record indicates that the known cosmos which can be observed at any given time is just a modest part of a much vaster one that remains to be found. Moreover, several of the more significant astronomical discoveries have been unanticipated and serendipitous, made when new telescopes were used to study known cosmic objects (Focus 1.3).

Focus 1.3 Serendipitous astronomy

The unanticipated discovery of hitherto unknown cosmic objects began with the four large satellites of Jupiter, which were found when Galileo was observing the nearby full Moon. Serendipity continued with the discovery of Uranus, found while preparing to measure the distances of stars; the first asteroid Ceres, discovered when determining accurate positions of the stars; the high velocities of galaxies in the expanding Universe, discovered while measuring the rotations of spiral nebulae thought to be newborn stars with nascent planetary systems; and the energetic radio Universe, found while measuring interference with terrestrial radio communications.

More recently, the Nobel Prize in Physics has been awarded for four serendipitous astronomical discoveries, including cosmic X-ray sources, discovered while studying solar X-rays reflected from the Moon; the three-degree cosmic microwave background radiation, found when measuring noise sources in a microwave detection system; the discovery of pulsars while observing the twinkling of cosmic radio sources caused by the Sun's winds; and the indirect detection of gravitational radiation when carrying out a search for new pulsars. Although outside the general solar-system theme of this book, these discoveries demonstrate the instrument-driven, accidental nature of astronomy, in which significant discoveries result from new technology and novel telescopes, enabling us to "see" the invisible and permitting us to look at the Universe in new ways.

The unanticipated discovery of Jupiter's four large moons

One of the most fascinating and lively books in astronomy, *Sidereus Nuncius* or *Starry Messenger*, was published in 1610. In it, Galileo Galilei (1564–1642) described how he turned the newly devised spyglass, which shows faraway things as though nearby, toward the heavens, bringing the sky down to Earth and the Earth into the sky.

After presenting the Venetian Doge with a spyglass, or telescope, as a valuable tool of war, Galileo turned a telescope of his own making toward our Moon on 25 August 1609, discovering craters, rugged mountains and valleys that you need a telescope to see. At least one cosmic object, the Moon, was no longer the polished, smooth and perfectly spherical body imagined by the ancients. Even the Sun was found to be spotty and impure under telescopic scrutiny by Galileo, Thomas Harriot (1560–1621), and other pioneering observers.

Although few of Galileo's many telescopes survive, he probably used one with an objective lens of about 0.05 meters (2 inches) in diameter to make his startling discoveries. The collecting area of such a lens is roughly fifty times that of the pupil of the unaided eye, which is about 0.007 meters across. The increase in light-collecting power of even this small telescope enabled Galileo to resolve objects that remain blurred to the unaided eye, detecting previously unseen craters on the Moon, viewing thousands of stars in the Milky Way that had never been seen before, and detecting the four large moons of Jupiter.

When directing his rudimentary telescope at the nearly full Moon on 7 January 1610, Galileo must have naturally moved his spyglass just a little to look at Jupiter, which was then located just above the Moon and was also the next brightest object in the sky. As reported in the *Starry Messenger*, Galileo found four companions lined up on each side of Jupiter, accompanying the planet and orbiting it at different distances (Fig. 1.14). The smaller their orbit, the faster their orbital speed and the shorter their orbital period.

Any object that orbits a planet is now called a satellite, and a natural satellite is also now called a moon. We designate the Earth's Moon, or our Moon, with a capital M to distinguish it from all the other planetary moons. The four large moons that Galileo discovered are now often called the "Galilean satellites" in his honor. Galileo named them the "Medicean stars" after his patron Cosmo II de' Medici, but they now go by the names of four of Jupiter's lovers. Io is the innermost of the four Galilean satellites, succeeded by Europa, Ganymede, and Callisto.

The discovery of these satellites was most likely an unexpected result of Jupiter's proximity to the Moon when Galileo happened to be observing it with his spyglass. No one predicted the possible existence of moons orbiting any other object than the Earth, and Galileo's important discovery of more than one center of motion in the Universe contradicted the geocentric Ptolemaic system in which all astronomical objects move around the central Earth.

Galileo's *Dialogo Massimi Sistemi Del Mondo, Tolemaico e Copernicano*, or *Dialogue of the Two Great World Systems, Ptolemaic and Copernican*, published in 1632, demonstrated the advantages of the Sun-centered Copernican cosmology, and provided telescopic evidence in its favor, including his discovery that Venus exhibits phases like our Moon. If nearby Venus orbited the Earth inside the Sun's orbit, then it could never appear completely illuminated, but Venus could appear from Earth in all its phases if it orbited the Sun.

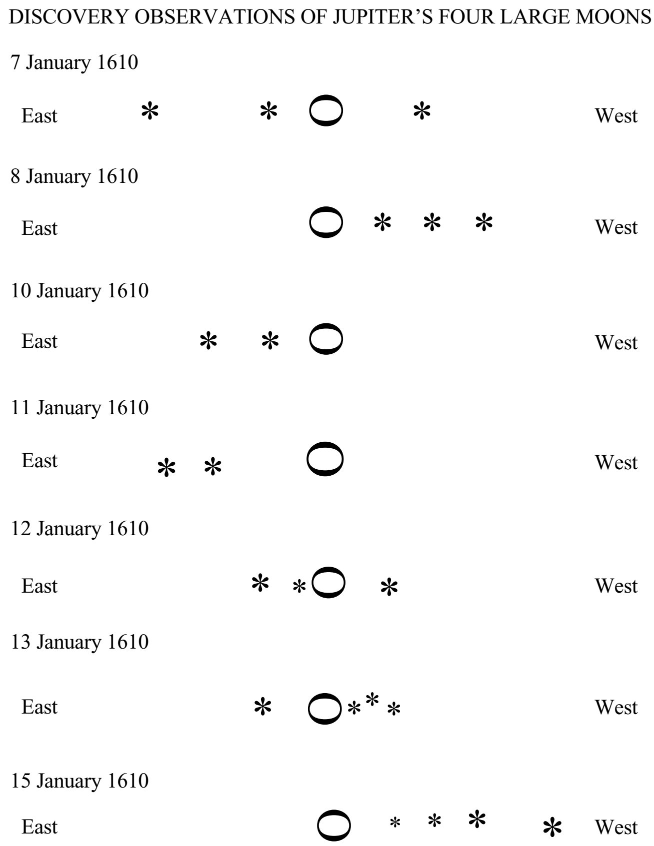


Fig. 1.14 Moons of Jupiter Some of Galileo Galilei's (1564–1642) observations of the "Medicean stars", lined up on each side of Jupiter's disk and changing position while accompanying Jupiter. (Adapted from Galilei, Galileo: *Sidereus Nuncius*, or *The Sidereal Messenger*, 1610. Translation by Albert Helden with Introduction, Conclusion and Notes. The University of Chicago Press, Chicago, 1989.)

Galileo's adoption of Copernicus' theory, in which the Earth moves around the Sun, was nevertheless opposed by theologians of the time, since a strict interpretation of the Bible (Psalm 104) indicated that "God fixed the Earth on its foundation, so it will never be moved". After trial by Inquisition in 1633, the Roman Catholic Church forced Galileo to recant his support of the Copernican system as "abjured, cursed and detested". He was banished to confinement at his house in Arcetri, in the hills surrounding Firenze, where he spent his last years. Legend has it that as Galileo rose from kneeling before his inquisitors, he murmured, "*e pu, si muove*" – "even so, it does move", but he would hardly have been foolish enough to risk even greater punishment. Not until 1992, more than 350 years after his trial, did Pope John Paul II in effect apologize for the harshness of Galileo's sentence.

The serendipitous discovery of Uranus

The first planet to be discovered since the dawn of history was found accidentally, by a professional musician and self-taught amateur astronomer, William Herschel (1738–1822), who moved from Germany to England in 1750 and became the organist in the town of Bath. He was both an excellent observer and a skilled mirror-maker, constructing telescopes with then unsurpassed light-collecting ability. Since they could not build such a fine instrument and obtain similar results, other astronomers were initially skeptical of Herschel's observations, but he simply replied that they would have to learn to see.

In 1781 Herschel was using one of his unique telescopes, with a metal mirror of 0.15-meters (6.2-inches) in diameter and 2.1-meters (7-feet) focal length, to examine all the brighter stars, of less than 8th magnitude, for faint companions. (In the peculiar magnitude system adopted by early astronomers, the brighter stars have smaller magnitudes.) He was hoping to determine the parallax, or distance, of the bright star from its changing position relative to the faint, adjacent one.

On 13 March 1781, Herschel unexpectedly found an uncommonly bright object, of 6th magnitude, that had a well-defined disk, unlike stars, and moved slowly from one night to another against the background stars. It was suspected by Herschel to be a comet without a tail. After several months of observations by Herschel and others, the moving object was recognized as a new planet, named Uranus, which orbits the Sun at about twice Saturn's orbital distance. Herschel became world-famous almost overnight. He was eventually appointed the King's Astronomer with a pension, which permitted him to give up music as a career and devote full time to astronomy.

Herschel proposed that the new planet be named the "Georgian Planet" in honor of King George III, England's reigning monarch and a patron of the sciences. After some controversy, the new planet was instead named Uranus, after the Greek personification of the sky. One consequence of this naming was that a newly discovered, heavy element was designated uranium in honor of the discovery of a new world.

When he found Uranus, Herschel was apparently unaware of a numerical sequence that predicted its relative distance from the Sun. Known as the Titius–Bode law, after the last names of the first persons to state it, the sequence describes the regular spacing of the planets, suggesting that the next planet beyond Saturn would be located at 19.6 AU, or at about twice Saturn's distance (Focus 1.4). The so-called "law" also indicated a missing planet at 2.8 AU, in the gap between Mars and Jupiter. Encouraged by the discovery of Uranus, astronomers began a search

Focus 1.4 The Titius–Bode law

In the inner solar system, each planet's orbit is about 1.5 times the distance of its inward neighbor, and this ratio increases to roughly a factor of 2.0 in the outer solar system. This relative spacing of the planets is described by the Titius–Bode law, first noted in 1766 by Johann Daniel Titius (1729–1796), and brought to prominence by Johann Elert Bode (1749–1826) in a 1772 edition of his popular book on astronomy.

The law states that the relative distances of the planets from the Sun can be approximated by taking the sequence 0, 3, 6, 12, 24, . . . , adding 4, and dividing by 10. Mathematically, the semi-major axis, a_n, of the nth planet, in order of increasing distance from the Sun, is given by:

$$a_1 = 0.4 \text{ for } n = 1$$
$$a_n = 0.1 \times [4 + 3 \times 2^{n-2}] \text{ for } n = 2, 3, \ldots, 9$$

where a_n is the relative distance compared with that of the Earth.

A comparison of the observed semi-major axes, a_P, of the planets with the distance predicted by this law is given in Table 1.2. The Titius–Bode law predates the discovery of Uranus at $n = 8$ by 15 years, the discovery of the first asteroid at $n = 5$ by 35 years, and the discovery of Neptune at $n = 9$ by 80 years. Although there is no well-accepted explanation for why this expression works so well, it probably has something to do with the dynamics, evolution or origin of the solar system.

Table 1.2 Comparison of measured planetary distances from the Sun with those predicted from the Titius–Bode law

Planet	n	Measured a_P (AU)	Predicted a_n (AU)
Mercury	1	0.387	$0.4 = (0 + 4)/10$
Venus	2	0.723	$0.7 = (3 + 4)/10$
Earth	3	1.000	$1.0 = (6 + 4)/10$
Mars	4	1.524	$1.6 = (12 + 4)/10$
Ceres (asteroid)	5	2.767	$2.8 = (24 + 4)/10$
Jupiter	6	5.203	$5.2 = (48 + 4)/10$
Saturn	7	9.537	$10.0 = (96 + 4)/10$
Uranus	8	19.19	$19.6 = (192 + 4)/10$
Neptune	9	30.07	$38.8 = (384 + 4)/10$

for the unknown planet that ought to be located at the predicted distance of 2.8 AU from the Sun. The first object to be found in this location was nevertheless discovered quite unexpectedly by the Sicilian monk and astronomer Giuseppe Piazzi (1746–1826) while he was preparing a catalog of accurate star positions.

The unexpected discovery of the first asteroid, Ceres

In the late 18th century, Giuseppe Piazzi (1746–1826) was compiling a catalog of the accurate positions of stars in the sky, using a finely calibrated telescope built by the celebrated instrument maker, Jesse Ramsden (1735–1800) of London, and installed in Piazzi's observatory at Palermo, Sicily (Fig. 1.15). On 1 January 1801, Piazzi unexpectedly discovered a new "star" of 8th magnitude that changed position from night to night. It was moving at a slow uniform rate against the background stars, and was thought to possibly be a comet without a nebulosity or tail.

Piazzi observed the new object for six weeks until it moved too close to the Sun to be observed, and when it returned to dark skies it could not be located. Hearing of the lost object, Carl Friedrich Gauss (1777–1855) developed a method of establishing an orbit from just a few observations, and used Piazzi's observations to predict a location. It was found a year after it had been first sighted, independently by Baron Franz Xaver von Zach (1754–1832) and Heinrich Wilhelm Olbers (1758–1840), close to Gauss's estimated position. A hitherto unknown, small solar-system body, named Ceres Ferdinandea by Piazzi, had been found, with an orbit between those of Mars and Jupiter. A few months later Olbers found another tiny object, named Pallas, orbiting the Sun at about the same distance as Ceres, and William Herschel named the two objects "asteroid" because they could not be resolved into disks and appeared to be "star-like" points of light.

The asteroids remained unresolved because they are very small and relatively nearby, rather than very large and distant like the stars. Even the largest asteroid, Ceres, has a radius of 476 kilometers, which is less than a third of the radius of the Moon, less than a tenth the radius of the Earth, and less than one hundredth the radius of Jupiter.

No other asteroids were identified for 38 years, but the hunt for new ones became something of an astronomical sport in the last half of the 19th century. More than 300 asteroids had been discovered by 1891, and the pace of discovery subsequently increased by using long-exposure photographs of several hours to detect the motions of asteroids against the stars.

Each asteroid is given a number corresponding to its chronological place in the discovery list, and a name that is usually provided by the discoverer, but they do not receive official numbers until their orbits are reliably known. For instance, 433 Eros was the 433rd asteroid to be discovered with a reliable orbit. The list of known asteroids reached the 2000 mark in 1977, and there were 50 000 known in the early 21st century. Astronomers estimate that there may be as many as half a million (500 000) faint asteroids smaller than one kilometer across, many with orbits that have now been determined.

Yet, despite their vast numbers, the combined mass of all the asteroids is estimated to be no more than 10 percent of that of the Earth's Moon, and nowhere near the mass of a single large planet.

Most of the asteroids with well-determined orbits lie in a great asteroid belt between the orbits of Mars and Jupiter (Fig. 1.16), at distances from the Sun of 2.2 to 3.3 AU and with orbital periods of 3 to 6 Earth years. The asteroids are so little, and distributed across such a large range of distances, that the asteroid belt is largely empty space. This leaves plenty of room for spacecraft to pass through to the giant planets, undamaged by collision with any asteroid.

Refractors and reflectors

Galileo's use of the telescope to extend the human senses marked the beginning of a new age in astronomy – an age in which telescopes are used to view objects hitherto unseen and unknown, and to scrutinize known ones in greater detail. Telescopes extend our vision by collecting enough light to detect intrinsically faint and otherwise invisible sources or to resolve bright sources whose individual features are too near to each other to separate with the unaided eye. This era continues today, as we build new

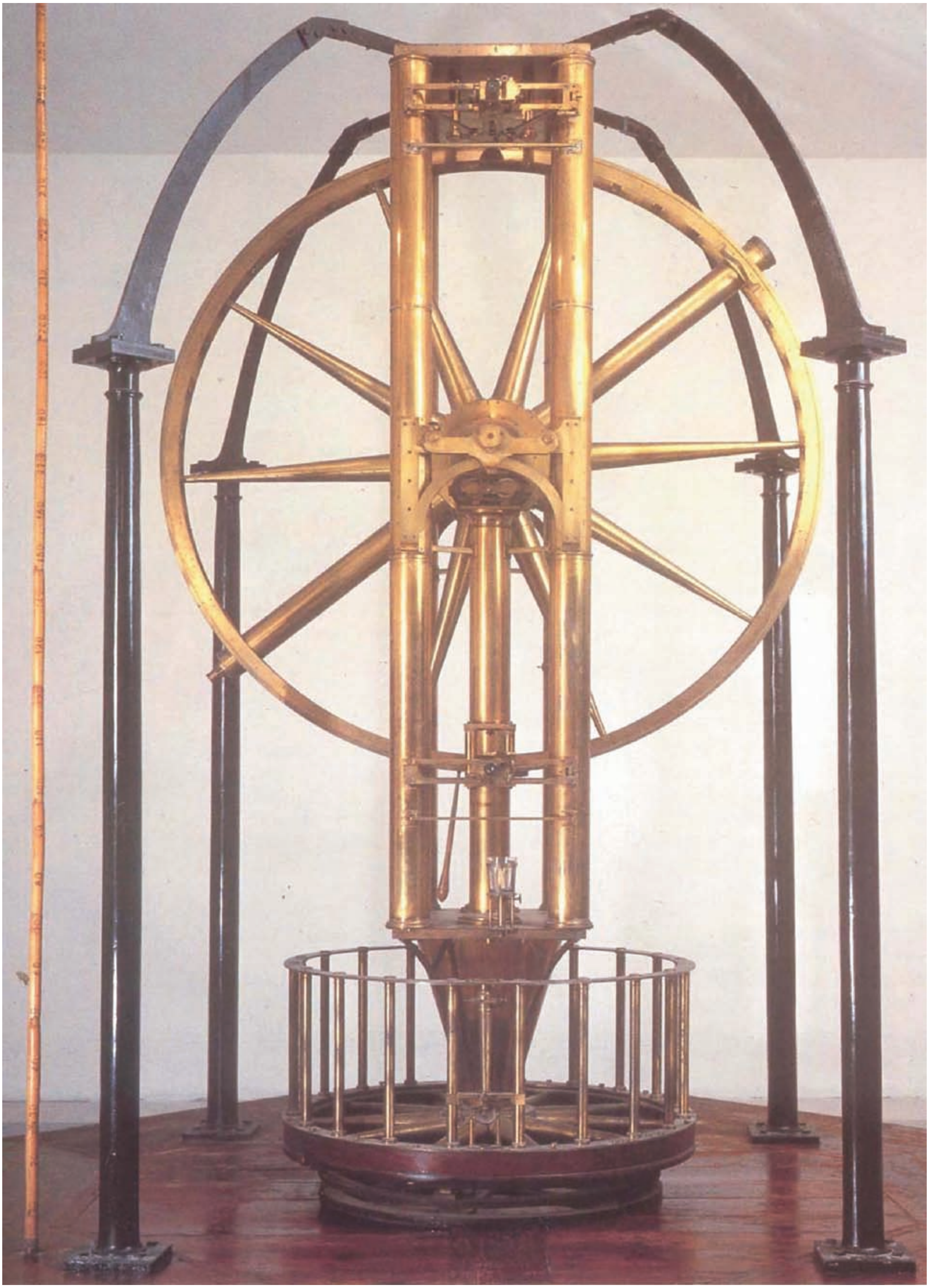


Fig. 1.15 The Palermo circle This instrument, built by Jesse Ramsden (1735–1800) of London, was used by Giuseppe Piazzi (1746–1826) to obtain precise measurements of stellar positions, with an accuracy of a few seconds of arc after observing each star for at least two nights. In the process, Piazzi unexpectedly discovered the first asteroid, 1 Ceres, in 1801. The telescope has a 7.5-cm objective lens. The altitude scale (5 feet in diameter) was read with the aid of two diametrically opposed micrometer microscopes; the azimuth scale (3 feet in diameter) was also read by means of a micrometer microscope.

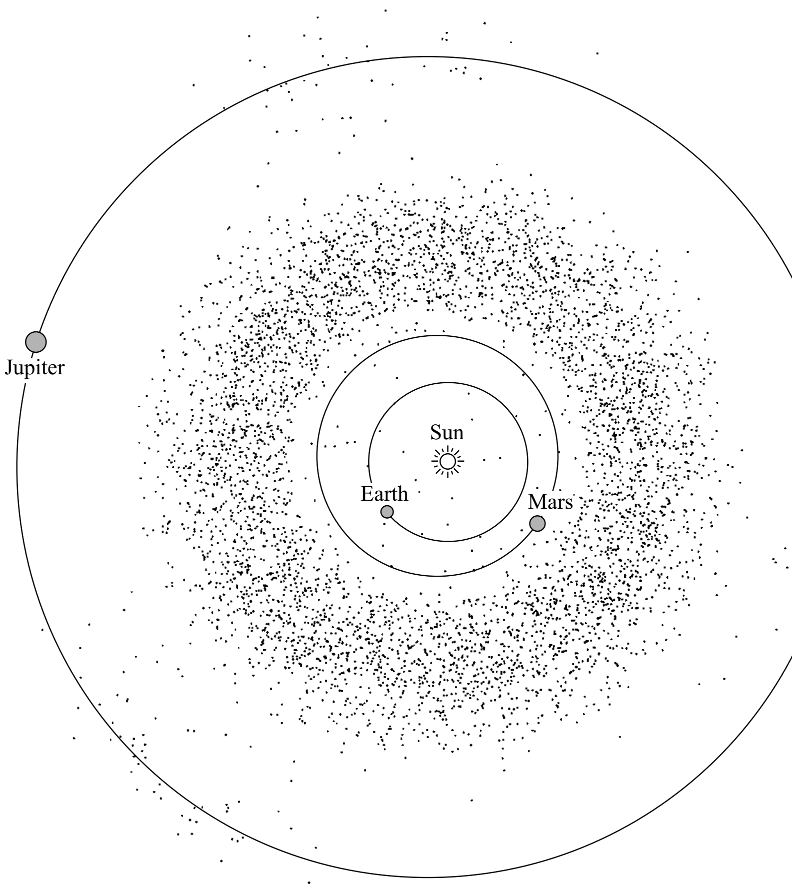


Fig. 1.16 Asteroid belt The exact locations of 5000 flying rocks, called asteroids or minor planets, whose orbits are accurately known. The vast majority of the asteroids orbit the Sun in the main belt located between the orbits of Mars and Jupiter. A few of them pass inside the orbit of Earth, while others move about 60 degrees ahead of and behind Jupiter in similar orbits. (Courtesy of Jeff Bytof, University of California at San Diego.)

telescopes on the ground and in space, to discover new worlds and to investigate familiar ones in different ways, and use spacecraft to carry instruments for close-up views of the planets and satellites, revealing unanticipated features that are far beyond the range of human vision with even the best telescope on Earth or in orbit around it.

There are two kinds of telescopes that are used to collect and focus radiation visible to the human eye. They are the refractor, used by Galileo, and the reflector, used by William Herschel. As the names suggest, the refractor uses a lens to gather, bend and focus light, employing the principle of refraction, while the reflector uses a curved mirror to collect, reflect and focus light (Fig. 1.17). Since the science of optics is used to describe the refraction or reflection of light rays, both kinds of telescopes are known as optical telescopes. They are used to carry out visible-light optical astronomy. Radio astronomers and X-ray astronomers use different kinds of telescopes that detect invisible radiation.

The first telescopes, such as the one Galileo used in 1610, were refractors using combinations of glass lenses to magnify and focus light. They were long, slender tubes with a convex objective lens at the far end that focused light on a second, smaller lens, termed the eyepiece, a few feet away. The eyepiece was used to magnify or enlarge the image before being observed, permitting more detail to be seen.

Galileo used an objective lens just one or two inches in diameter, and a smaller, concave eyepiece about 30 or 40 inches away, giving an erect image but a very narrow field of view. His telescopes had a magnification of about 20 and a field of view of about a quarter of a degree. Kepler introduced a convex eyepiece behind the focal plane, which widened the viewing angle and inverted the image. Many subsequent astronomers became accustomed to viewing this upside-down world.

The distance from the objective lens to the focal plane, called the focal length, determines the overall size of an image – the greater the focal length, the bigger the image. The diameter of the objective lens is called the aperture, and the focal ratio of the lens is the focal length divided by the aperture. The magnification, or power, of such a telescope is equal to the ratio of the focal lengths of the objective and the eyepiece.

Because the glass objective lens in a refractor could not be shaped into an accurate curve, these early telescopes yielded blurry images, distorted by spherical aberration. Moreover, because light is composed of different colors or wavelengths, the imperfect glass lens bent or refracted

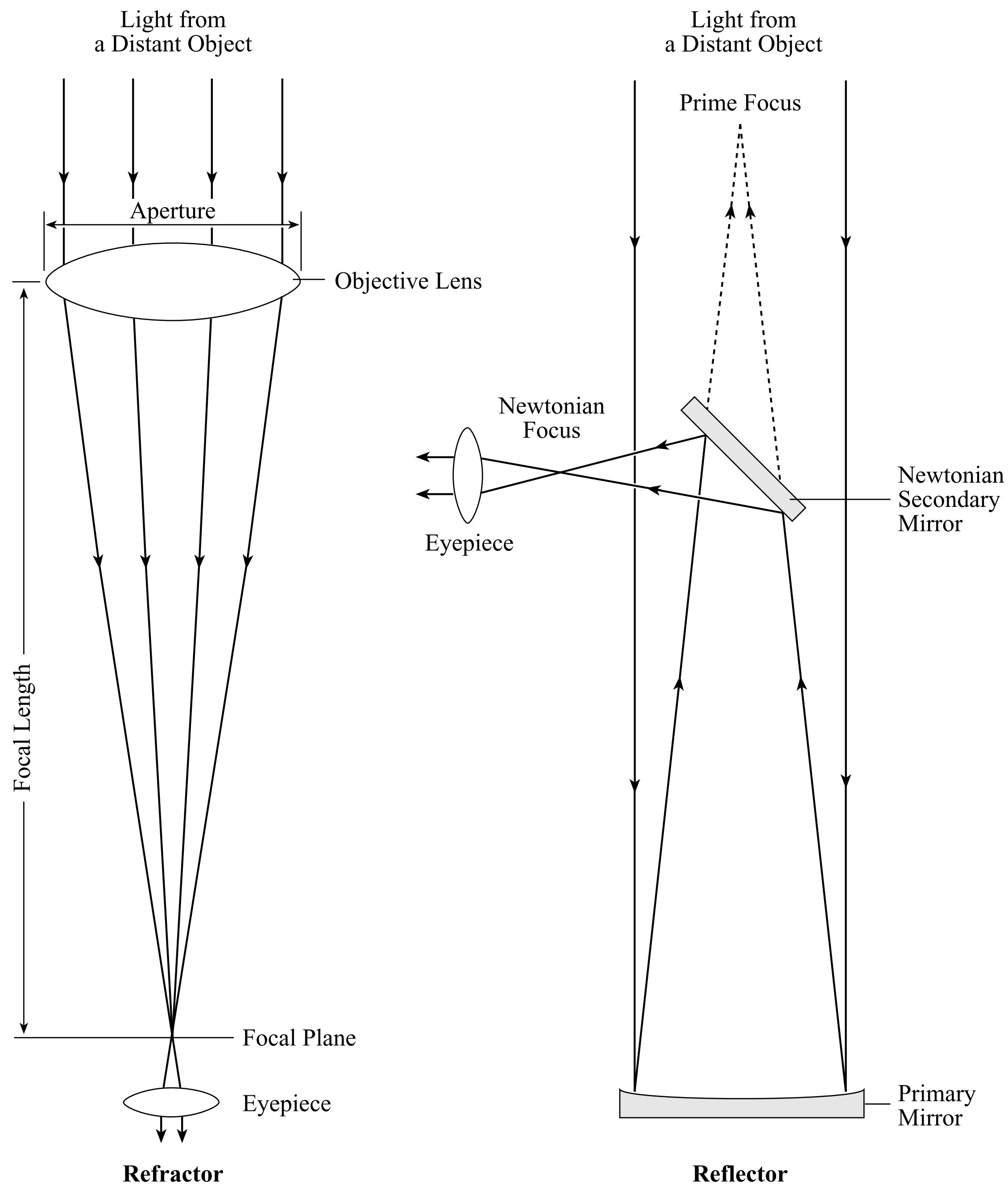


Fig. 1.17 Refractor and reflector Two kinds of optical telescopes, the refractor and the reflector, are used to gather and concentrate light at visible wavelengths. Light waves that fall on the Earth from a distant object are parallel to one another, and are focused to a point by the lens or mirror of a telescope. The earliest telescopes were refractors (*left*). The curved surfaces of the convex objective lens bend the incoming parallel light rays by refraction, and bring them to a focus at the center of the focal plane, where the light rays meet and an image is created. A second, smaller lens, called the eyepiece, was used to magnify the image in the early refractors; later versions placed photographic or electronic detectors at the focal plane. The reflecting telescope (*right*) uses a large, concave, or parabolic, primary mirror to collect and focus light. A small, flat secondary mirror, inclined at an angle of 45 degrees to the telescope axis, reflects the light sideways, at a place now known as the Newtonian focus. Other light-deflecting mirror arrangements can be used to obtain any desired focal length, which varies with the curvature and position of small convex mirrors.

each wavelength through a slightly different angle, focusing light of different colors differently and resulting in chromatic aberration.

Such blurring was avoided in a reflecting telescope that uses mirrors instead of lenses. A curved primary mirror gathers the parallel rays of light entering the open end of a telescope and focuses them to a point (Fig. 1.17). Light does not pass though a mirror, as it does through a lens, and the mirror concentrates light of all colors to the same focus, producing a sharp image.

In 1668 Isaac Newton (1642–1727), born on Christmas day the year of Galileo's death, placed a second, flat mirror, angled at 45 degrees, just before the focal point of the primary mirror, to direct the light to the side where an eyepiece was located. This Newtonian focus remains popular for many amateur astronomers because of its elegant simplicity, but other light-deflecting mirror arrangements are used with the biggest optical telescopes to obtain any desired focal length.

The critical parameter of a telescope is the diameter of the light-gathering lens or mirror. The larger the diameter, the more light is gathered, the brighter will be the image, and the fainter the objects that can be seen or recorded. The amount of light that can be gathered is proportional to the area of the lens or mirror, and consequently to the square of its diameter, so a small change in diameter can make a large change in light-collecting area.

Bigger telescopes also provide better angular resolution, which is the ability to detect the separation between things that are close together. This ability to discriminate fine details is called the resolving power of a telescope, and it depends on the diameter of the light-gathering lens or mirror and the wavelength of observation (Focus 1.5). At a given wavelength, a bigger objective lens or primary mirror provides better angular resolution.

The angular resolution of ground-based optical telescopes operating at visible wavelengths is limited by turbulence in the Earth's atmosphere. Similar variations cause the stars to twinkle at night. This atmospheric limitation to angular resolution is called "seeing". The best seeing, of about 0.2 seconds of arc in unusual conditions, is found only at a few high-altitude sites in the world, and observatories are located at most of them. Better visible images with even finer detail can be obtained from the unique vantage point of outer space, using satellite-borne telescopes unencumbered by our atmosphere.

Focus 1.5 Angular resolution

The angular resolution, θ, of a telescope is determined by the diameter, D, of the objective lens or primary mirror, as well as the wavelength, λ, of observation. The mathematical expression is:

$$\text{Angular resolution } \theta = \frac{\lambda}{D} \text{ radians}$$
$$= 2.063 \times 10^5 \frac{\lambda}{D} \text{ seconds of arc}$$

where one radian is equivalent to 206 265 ($2.062\,65 \times 10^5$) seconds of arc. This equation tells us that a bigger lens or mirror provides finer angular resolution at a given wavelength. The resolving power of a telescope operating at the wavelengths that we detect with our eye is about $0.13/D$ seconds of arc if D is in meters.

Atmospheric effects limit the resolution of any telescope operating at visible wavelengths to about one second of arc, so you cannot improve the angular resolution by building an optical telescope bigger than about 0.13 meters in diameter. Nevertheless, a bigger telescope still gathers more light than a smaller one, permitting the detection of fainter sources. If a large telescope is placed in space, above our distorting atmosphere, greater angular resolution can also be achieved.

The equation applies at radio wavelengths where very big telescopes are required to achieve significant angular resolution. At a radio wavelength of 0.1 meters, an angular resolution of 1 second of arc requires a telescope with a diameter of 20 kilometers. The advantage of radio signals is that the atmosphere does not distort them, or limit the angular resolution. We can observe cosmic radio sources on a cloudy day, just as your home radio or cell phone work even when it rains or snows outside.

Satellites and rings

In order to get greater magnification, and at the same time to reduce distortion of images and the colored halo around them, or to reduce spherical and chromatic aberrations, objective lenses of great focal length were necessary. As a result, early refractor telescopes became longer and longer, with an objective lens placed on a tower or tall pole and a separate eyepiece near the ground. This novel arrangement permitted the use of an objective lens with slight curvature and long focal length to help correct for aberration and bring the image into sharp focus.

Such "aerial" refractors enabled the discovery of new planetary satellites and the rings of Saturn. In 1655 the Dutch astronomer Christiaan Huygens (1629–1695), for example, discovered Titan, the first known and largest satellite of Saturn, named after Saturn's older brother. Huygen's telescope had a 0.05-meter (2-inch) objective lens with a focal length of 7 meters (23 feet), connected to an eyepiece by just a string for alignment.

Within a few decades the Italian astronomer Giovanni Domenico Cassini (1625–1712), working at the Paris Observatory, had used an aerial refractor to discover four more moons circling Saturn; they are named Iapetus, Rhea, Tethys and Dione. Like the Earth's Moon, Saturn's second-largest moon, Iapetus, always presents the same face to its planet. According to Greek mythology, Gaia (Earth) gave birth to Uranus (Heaven) without the aid of any male, and coupled with her son to conceive six male Titans, including Iapetus, and six female Titanesses, including Rhea and Tethys.

During the next three centuries, the discovery of planetary satellites progressed more or less in tandem with the development of increasingly large and more powerful reflecting telescopes, whose greater light-gathering capability permitted the detection of smaller satellites that reflected less sunlight. William Herschel, the discoverer of Uranus, used his 0.15-meter (6.2-inch) reflector to identify four new moons, two each of Uranus (Oberon and Titania in 1787) and Saturn (Mimas and Enceladus in 1789). It wasn't until 1851 that William Lassell (1799–1880) found two more Uranian satellites, Ariel and Umbriel, using a 0.61-meter (24-inch) reflector, and a fifth moon, Miranda, wasn't found until nearly a century later — in 1948 by Gerard P. Kuiper (1905–1973) at the McDonald Observatory in Texas using a 2.1-meter (82-inch) reflector.

Saturn's satellite Mimas has the name of one of the giants who fought against the gods in Greek mythology. The ringed planet's moon Enceladus is named for the giant who was crushed in a battle between the Olympian gods and the Titans. Earth that was piled on top of him became the island of Sicily. Two other large satellites of Saturn, discovered in the 19th century, are named after Hyperion, a Titan, and Phoebe, a Titaness.

The five large satellites of Uranus are named for characters in literature. Oberon and Titania are the king and queen of the fairies in Shakespeare's *A Midsummer Night's Dream*. Inside their orbits is Umbriel, a "dusky, melancholy sprite" in Alexander Pope's *Rape of a Lock*. Close to the planet is Ariel, described by Shakespeare as "an airy spirit" in *The Tempest*. Closer yet is Miranda, named for Prospero's daughter in *The Tempest*.

Lassell found Neptune's largest satellite in 1846, just a few weeks after the discovery of the planet. The satellite was named Triton – a sea god in Greek mythology, the son of Poseidon, the Greek equivalent of Neptune, the Roman god of the sea. A second, much smaller Neptunian satellite was not definitely known for more than a century; Kuiper located it on photographic plates in 1949. It was named Nereid for a sea nymph lured by Triton's conch-shell music in mythology.

Neptune's satellites differ from those of Jupiter, Saturn and Uranus. Each of these three giant planets has a group of large satellites that revolve in regularly spaced, circular orbits in the same direction as the rotation of the planet and close to the planet's equatorial plane, presumably because they share the rotation of the material from which the planet and its satellites formed. In contrast, Neptune's largest satellite, Triton, moves in the backward retrograde direction, opposite to the direction of the planet's rotation and the orbital direction of all planets and most satellites.

The large planetary satellites, with radii larger than 100 kilometers, or 10^5 meters, were all discovered by the mid 20th century, and most of them were known by the end of the 19th century (Table 1.3). Altogether 21 of them are known – one for Earth, four for Jupiter, nine for Saturn, five for Uranus, and two for Neptune.

In the meantime, back in the 17th century, Huygens turned his telescope toward Saturn itself, and explained its mysterious handle-like appendages. Galileo had noticed that the planet was not round, but had blurry objects on each side. When these objects disappeared two years later, Galileo wondered if Saturn "had devoured his own children". In 1656 Huygens, then only 27 years old, realized that the planet is surrounded by "a thin flat ring, nowhere touching it, and inclined to the ecliptic" (Fig. 1.18). Because the ring is tipped with respect to the plane of the Earth's orbit around the Sun, it changes its shape when viewed from Earth, slowly opening up and then turning edge-on as Saturn makes its 29.5-year orbit around the Sun. When the ring is opened up, it resembles handle-like appendages, but when it is viewed edge-on the ring virtually disappears.

Cassini also observed Saturn's ring, suggesting that it was composed of swarms of satellites too small to be resolved individually, circling the planet with different velocities. In 1675 he discovered a dark separation in the ring that is now known as the "Cassini Division".

In fact, there are three main rings of Saturn visible from the Earth, the outer A, central B and inner C rings (Table 1.4). The Cassini Division separates the A and B rings. The C ring is also known as the "crepe ring" since it is the most transparent of the three main rings.

But why don't Saturn's rings fall down onto the planet? The rings stay up because they are moving, supported by their motion against the downward pull of Saturn's gravity. That is the same reason our Moon stays apart from the Earth, yet always accompanying it. Except, as James Clerk Maxwell (1831–1879) showed in 1859, the wide, thin rings of Saturn are composed of a vast number of small particles, each pursuing its individual orbit in the

Table 1.3 Large planetary satellites

Name	Mean radius (km)[a]	Mass (10^{20} kg)[a]	Mean mass density (kg m^{-3})[a]	Distance from planet (10^3 km)	Period of revolution[b] (days)	Year of discovery
EARTH						
Moon	1738	734.8	3344	384.4	27.3217	
JUPITER						
Io	1822	843.2	3528	422	1.77	1610
Europa	1561	480.0	3013	671	3.55	1610
Ganymede	2631	1481.9	1942	1070	7.16	1610
Callisto	2410	1075.9	1834	1883	16.7	1610
SATURN						
Mimas	198	0.38	1150	186	0.94	1789
Enceladus	252	1.08	1608	238	1.37	1789
Tethys	533	6.18	973	295	1.89	1684
Dione	562	10.96	1476	377	2.74	1684
Rhea	764	23.07	1233	527	4.52	1672
Titan	2576	1345.5	1880	1222	15.9	1655
Hyperion	135	0.06	542	1481	21.3	1848
Iapetus	736	18.06	1083	3561	79.3	1671
Phoebe	107	0.08	1634	12 952	550 R	1898
URANUS						
Miranda	236	0.66	1214	130	1.41	1948
Ariel	579	12.6	1592	191	2.52	1851
Umbriel	585	12.2	1479	266	4.14	1851
Titania	789	34.2	1662	436	8.71	1787
Oberon	761	28.8	1559	583	13.5	1787
NEPTUNE						
Triton	1353	214.2	2064	354	5.88 R	1846
Nereid	170	0.3	1500	5510	360	1949

[a] The radii are given in units of kilometers (km), the mass is in kilograms (kg), and the mean mass density in kilograms per cubic meter (kg m^{-3}). By way of comparison, the equatorial radius of the planet Mercury is 2440 kilometers, so Ganymede and Titan are both bigger than Mercury.

[b] The letter R following the period denotes a satellite revolving about its planet in the retrograde direction, opposite to that of the planet's rotation and orbital motion about the Sun.

plane of the planet's equator. The innumerable particles that make up Saturn's rings act as tiny satellites that move in accordance with Kepler's third law, with the inner parts moving at a faster speed than the outer ones.

1.3 What holds the solar system together?

Gravity and motion – a delicate balance

The great English scientist Isaac Newton (1643–1727) showed that the same unchanging physical laws apply to the terrestrial and the celestial. Motions everywhere, whether up above in the heavens or down below on the ground, are described by the same concepts, and all material objects in the cosmos are subject to universal gravitation, with its unlimited capacity to act on matter. So everything in the cosmos moves in predictable and verifiable ways, described by Newton's laws of motion and gravitation. The basic ideas are that a moving body will continue to move in a straight line unless acted upon by an outside force, and that every object in the Universe attracts every other object as the result of universal gravitation.

According to tradition, Newton was sitting under an apple tree when an apple fell next to him on the grass.

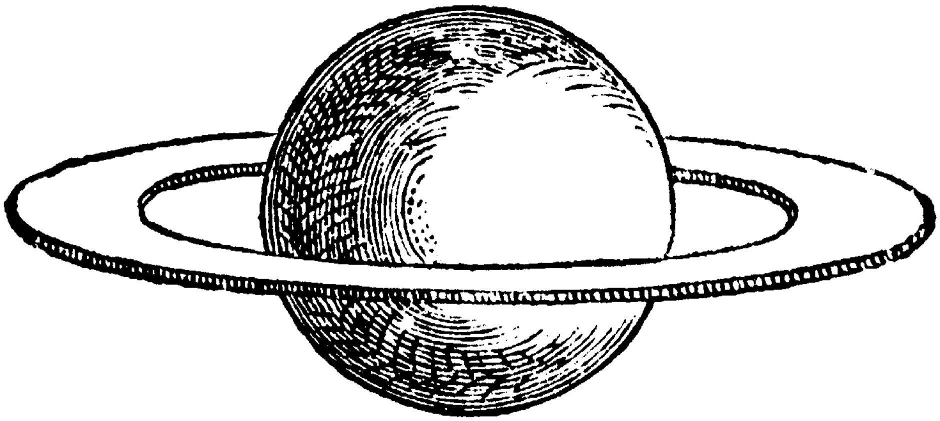


Fig. 1.18 Saturn's ring In 1659 Christiaan Huygens (1629–1695) published this drawing of Saturn and its ring in his monograph *Systemia Saturnium*. Huygens recognized that a detached ring would explain the planet's ever-changing appearance, and announced his discovery in the form of an anagram, a succession of scrambled letters. The drawing shown here was accompanied by the deciphered anagram "Saturn is girdled by a thin flat ring, nowhere touching it, and inclined to the ecliptic."

Table 1.4 The main rings of Saturn

Name	Distance from planet center (Saturn radii[a])	Orbital period[b] (hours)	Width (km)
A ring	2.025 to 2.267	12.1 to 14.2	14 670
Cassini division	1.949 to 2.025	11.4 to 12.1	4 585
B ring	1.525 to 1.949	7.9 to 11.4	25 580
C ring	1.235 to 1.525	5.8 to 7.9	17 490

[a] The equatorial radius of Saturn at the one-bar pressure level is 60 268 kilometers, or 6.0268×10^7 meters, nearly ten Earth radii.
[b] Saturn's rotation period is 10.6562 hours, so the inner B ring and the entire C ring move around the planet at a faster rate than the planet rotates.

This reminded him that the power of gravity, whose pull influences the motion of falling bodies, seems undiminished even at the top of the highest mountains. He therefore argued that the Earth's gravitational force extends to our Moon, and showed that this force can pull the Moon into an orbit. It was as if the Earth's Moon is perpetually falling toward the planet while always keeping the same mean distance from it. The Sun's gravity similarly deflects the moving planets into their curved paths, so they forever revolve around the Sun.

So Newton discovered the cosmic reach of gravity, which keeps our feet on the ground. Gravity has pinned us there, so we rotate with the spinning Earth and stay on it. The air and oceans are similarly held close to the planet by its gravitational pull.

The English genius was a self-isolated intellect, a bit obsessed, famously distracted, and frequently depressed. Newton didn't like interacting with people. He declined most invitations, avoided personal contact, never traveled abroad, and, they say, died a virgin at the age of eighty-five. He was also a rebel against authority, and spent much of his life immersed in experiments in alchemy and theological or mystical speculations, hoping to understand the origin of the elements and the eternal mysteries of health and mortality, examining mystic clues left by God.

It was his friend, the English astronomer Edmond Halley (1656–1742), who persuaded the secretive Newton to write his greatest work, the *Philosophiae naturalis principia mathematica*, or the *Mathematical Principles of Natural Philosophy*, commonly known as the *Principia*. It was presented to the Royal Society of London in 1686, which withdrew from publishing it owing to insufficient funds, so Halley, a wealthy man, paid for the publication the following year.

As Newton wrote in the *Principia*, "the Copernican system of the planets stands revealed as a vast machine working under mechanical laws here understood and explained for the first time". These were his laws of motion and the law of universal gravitation, achievements that resulted in Sir Isaac becoming the first person in England to be knighted, in 1705, for his scientific work.

The enormous reach of gravity can be traced to two causes. In the first place, gravitational force decreases relatively slowly with distance, and this gives gravitation a much greater range than other natural forces, such as those that hold the nuclei of atoms together. In the second place, gravitation has no positive and negative charge as electricity does, or opposite polarities as magnets do. This means that there is no gravitational repulsion between masses.

In contrast, the repulsive and attractive forces among like and unlike electrical charges in an atom cancel each other, shielding it from the electrical forces of any other atom.

The gravitational force is mutual, so any two objects attract each other, and every atom in the Universe feels the gravitational attraction of every other atom. Their attraction is proportional to the product of their masses and inversely proportional to the square of the distance between them. Their mass possesses inertia, the tendency to resist any change in its motion. Mass is incidentally an intrinsic aspect of an object, different from its weight which alters with distance from the main source of gravity. An astronaut weighs less when leaving the Earth, but retains the same mass.

Expressed mathematically, any mass M_1 produces a gravitational force F_{gravity} on another mass M_2, given by the expression:

$$\text{Gravitational force} = F_{\text{gravity}} = \frac{G M_1 M_2}{D^2}$$

where G is the universal gravitational constant, $G = 6.6726 \times 10^{-11}\ \text{m}^3\,\text{kg}^{-1}\,\text{s}^{-2}$, and D is the distance between the centers of the two masses. This is sometimes called an inverse square law, since the force of gravity is inversely proportional to the square of the distance or separation.

Newton used observations of planetary motions to determine just how the force decreases with increasing distance. Contrary to popular belief, Newton did not use his theory to show that planets move in elliptical orbits, but instead used Kepler's third law, connecting a planet's orbital period and distance, to show that the force of gravity must fall off as the inverse square of the distance.

The concept of universal gravitation, and Newton's expression for the gravitational force, can be used to derive Kepler's third law in the form:

$$P_{\text{P}}^2 = \frac{4\pi^2}{G}\,\frac{a_{\text{P}}^3}{(M_{\text{P}} + M_{\odot})} = 5.9 \times 10^{11}\,\frac{a_{\text{P}}^3}{M_{\odot}}\ \text{s}^2$$

where a_{P} is the semi-major axis of the planet's orbital ellipse in meters, P_{P} is the orbital period in seconds, and M_{P} and $M_{\odot}$ respectively denote the mass of the planet and the mass of the Sun in kilograms.

Within the solar system, the dominant mass is that of the Sun, which far surpasses the mass of any other object there. That is why we call it a solar system, governed by the central Sun. The sum $(M_{\text{P}} + M_{\odot})$ is therefore, to the first approximation, a constant equal to the Sun's mass, $M_{\odot}$, regardless of the planet under consideration.

So it is the Sun's gravitational attraction that keeps the planets in their orbits, and holds the solar system together. But why doesn't the immense solar gravity pull the entire planetary system into the Sun? Motion holds the planets up, opposing the relentless pull of the Sun's gravity and keeping the planets from falling into the Sun.

The reason that the planets do not plunge into the Sun is that each planet is also moving in a direction perpendicular to an imaginary line connecting it to the Sun, at exactly the speed required to overcome the Sun's gravitational pull, keeping the planet in perpetual motion. This orbital speed depends only on the Sun's mass and the planet's distance, but it is independent of the planet's mass.

If a planet moved any faster, it would leave the solar system, and if it moved any slower it would be pulled into the Sun. This delicate balance between motion and gravity also explains why the Moon revolves around the Earth, and why Saturn's ring particles remain separate from the planet.

You might say that motion seems to define existence, for there is nothing in the Universe that is completely at rest. Everything that exists, from atoms to planets and stars to galaxies, moves through space, all surely going somewhere. It is this motion that shapes the Universe, giving it form, structure and texture. When you stop moving it is all over, or as Bob Dylan (1941–) sang: "better start swimming, or you'll sink like a stone".

Neptune's discovery, triumph of Newtonian gravitational theory

Neptune's discovery was no accident, in contrast to those of Uranus and the first asteroid. It was a direct consequence of precise mathematical calculations of Uranus' motion. Uranus had been detected by professional astronomers and mistaken for a star on no less than 22 occasions during the century that preceded the realization that it was a planet. These additional observations could be combined with the post-discovery ones to determine Uranus' trajectory and calculate its future position. Before long it was found that the planet was wandering from its predicted path.

A large unknown world, located far beyond Uranus, was evidently producing a gravitational tug on Uranus, causing it to deviate from the expected location. Two astronomer–mathematicians, John Couch Adams (1819–1892) in England and Urbain Jean Joseph Le Verrier (1811–1877) in France, independently located the planet by a mathematical analysis of the wanderings of Uranus.

Adams, a recent graduate from Cambridge University, finished his work first, deriving a precise position of the planet in mid-1845. He left a summary of his results with the then Astronomer Royal, George Biddell Airy (1801–1892), who did not feel compelled to look for the unknown world.

Le Verrier finished his best calculations about a year later, and, unlike Adams, published his results. Both scientists had assumed that the undiscovered planet occupied the next place in the sequence of the Titius–Bode law, and they arrived at nearly identical locations for it.

When Le Verrier's memoir reached Airy, he persuaded James Challis (1803–1882), professor of astronomy at Cambridge University, to make a search for the undiscovered planet. For a variety of reasons, Challis began the investigation slowly, and Le Verrier had in the meantime sent his results to the Berlin Observatory where Johann Gottfried Galle (1812–1910) and his student Heinrich Louis d'Arrest (1822–1875) found the planet. They identified it on the first night of their search, on 23 September 1846, using a 0.23-meter (9-inch) refractor; it was located within a degree of both Adams' and Le Verrier's predicted positions. Only later did Challis realize that he had previously observed the planet twice when beginning his own search.

The discovery of the new planet, named Neptune after the Roman god of the sea, was acclaimed as the ultimate triumph of Newtonian science. It resulted from mathematical calculations, based on Newton's theories, of the effects of an unknown planet whose gravity was pulling Uranus from its predicted place. If proof were needed, this achievement certified the validity of gravitational theory.

Neptune is located at a mean distance of about 30 times the distance between the Earth and the Sun. The remote planet takes about 165 Earth years to travel once around the Sun, so it will not make a full orbit since its discovery until 2011.

1.4 Physical properties of the Sun

What are the distances to the Sun and other nearby stars?

How far away is the Sun from the Earth, and how fast is the Earth moving through space? Kepler's model of planetary motion only provided a scale model for the relative distances of the planets from the Sun, and for a long time no one knew exactly how big the solar system was. Our planet's true distance from the Sun remained unknown for centuries. And since the distance was not reliably known, the velocity of the planet around the Sun could not be determined. The Earth's mean orbital speed is equal to the circumference of the orbit divided by one year, the time for the Earth to complete one trip around the Sun.

The crucial unit of distance for the planets is the mean Earth–Sun distance, known as the astronomical unit (AU). It can be determined by first estimating the distance between Earth and a nearby planet, and using the measurement together with geometry and Kepler's third law to infer the Earth–Sun distance.

The distance to a planet can be estimated by measuring the angular separation of the planet when observed simultaneously from two widely separated locations. This angle is known as parallax, from the Greek *parallaxis*, for the "value of an angle". If both the parallax and the separation between the two observers are known, then the distance of the planet can be determined by triangulation. It is based on the geometric fact that if you know the length of one side of a triangle and the angles of the two corners, all the other dimensions can be calculated.

The parallax technique of estimating a planet's distance is similar to the way your eyes infer how far away things are. To see the effect, hold a finger up in front of your nose, and look at your finger with one eye open and the other closed, and then with the open eye closed and the closed one open. Any background object near to one side of your finger seems to move to the other side, making a parallax shift. When this is repeated with your finger held farther away, the angular shift is smaller. In other words, the more distant an object, the smaller the parallax shift, and vice versa.

Giovanni Domenico Cassini (1625–1712), an Italian astronomer and the first director of the Paris Observatory, obtained an early triangulation of Mars in 1672, combining his observations from Paris with those taken from Cayenne, French Guiana. The planet was then in opposition, at its closest approach to Earth. From the two sets of observations, made 7200 kilometers apart, it was possible to estimate the distance to Mars and to infer a value of 139 million (1.39×10^8) kilometers for the astronomical unit.

Astronomers in the 18th and 19th centuries attempted to improve the measurement accuracy of the Sun's distance during the rare occasions when Venus crossed the face of the Sun, in 1761, 1769, 1874 and 1882. The method also involved comparison of observations from widely separated locations to determine the distance by triangulation. Subsequent determinations of the distance to the nearby minor planet 433 Eros during its closest approaches to the Earth resulted in an estimated 150 million kilometers for the astronomical unit.

Significant improvements in the precision of planetary distances came in the late 1960s by bouncing pulsed radio waves off Venus and timing the echo (Fig. 1.19). The round-trip travel time – about 276 seconds when Venus is closest to the Earth – was measured using accurate atomic clocks, and a precise distance to Venus was then obtained by multiplying half the round-trip time by the speed of light (Focus 1.6). The distance of Venus from the Sun is equal to one half of the difference between the Earth and Venus

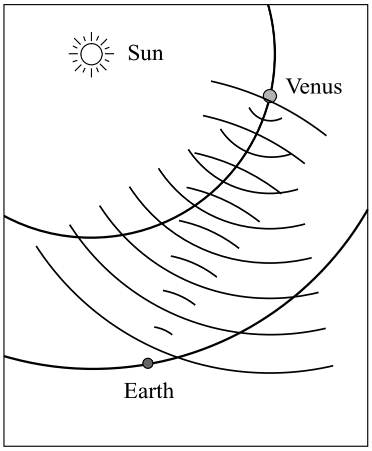


Fig. 1.19 Radar-ranging to Venus Accurate distances to the nearby planets have been determined by sending radio pulses from Earth to the planet, and timing their return several minutes later. The figure shows the emission of a pulse toward Venus; when it bounces from Venus the radiation spreads over the sky and we receive only a small fraction of the original signal, delayed by the round-trip travel time. If T is the round-trip time and c is the speed of light, the total distance traveled is cT and the distance to Venus is cT/2. For Venus, the round-trip time is 4.6 minutes when the planet is nearest Earth and increases to 28.7 minutes when it is furthest away from us.

when it is closest and furthest away from us, on the other side of the Sun. The value of the astronomical unit inferred from the radar determination of the distance of Venus is 149 597 870 kilometers, with an accuracy of about 1 kilometer, or for the accuracy required in most astronomical calculations 1 AU = 149.6 million kilometers.

Once the Venus–Sun distance is known, we can infer the distance of any other planet from the Sun using Kepler's third law, which relates the orbital periods and orbital distances of the planets. The average distances of the planets from the Sun are: Mercury 0.39 AU, Venus 0.72 AU, Earth 1.00 AU, Mars 1.52 AU, Jupiter 5.2 AU, Saturn 9.54 AU, Uranus 19.19 AU and Neptune 30.07 AU.

Nowadays the accuracy of the mean Earth–Sun distance is fixed by the exact value for the speed of light. The time τ_{AU} for light to travel across 1 AU is given as a primary astronomical constant:

$$\begin{aligned}\text{Earth–Sun light travel time} &= \tau_{\mathrm{AU}}\\ &= 499.004\,782 \text{ seconds}\end{aligned}$$

with a derived value for the mean Earth–Sun distance of:

$$1\,\mathrm{AU} = c\tau_{\mathrm{AU}} = 1.495\,978\,70 \times 10^{8} \text{ kilometers}$$

where the speed of light c is 299 792.458 kilometers per second. By way of comparison, one light-year is equal to 63 240 AU.

Once you have an accurate value for the Sun's distance, the Earth's mean orbital velocity can be determined by assuming, to a first approximation, a circular orbit and dividing the Earth's orbital circumference by the Earth's orbital period of $P_{\mathrm{E}} = 1 \text{ year} = 3.1557 \times 10^{7}$ seconds, or:

$$\begin{aligned}&\text{Earth's mean orbital velocity}\\ &\quad = \frac{2\pi \times (1\,\mathrm{AU})}{P_{\mathrm{E}}} = 29.8 \text{ kilometers per second}\end{aligned}$$

which is equivalent to about 170 000 kilometers per hour.

And what about the distances to the other nearby stars? Even the closest stars, other than the Sun, are too far away for us to detect a shift in position from any two points on Earth. To triangulate the distances of these nearest stars, astronomers needed a wider baseline, measuring their parallax from opposite sides of the Earth's annual orbit, or from a separation of twice the astronomical unit. The measurement involves careful scrutiny of two stars that appear close together in the sky, a bright one that is relatively nearby and the other fainter one that is much further away (Fig. 1.20).

The annual parallax of the nearest, brighter star can then be determined by measuring its angular separation from the fainter, distant one for a year or more. During the course of the year, the nearby star will seem to sway to and fro, in a sort of cosmic minuet that mirrors the Earth's orbital motion. Measurements separated by six months, from opposite sides of the Earth's orbit, can reveal a shift in the position of a nearby star with respect to the more distant ones. Half of this angular displacement, known as the annual parallax π, is the ratio of the AU and the star's distance D, or to be precise $\sin \pi = (1\,\mathrm{AU})/D$ radians. The nearer the star, the larger the annual-parallax sways.

The first star whose distance was reliably determined in this way was 61 Cygni. As reported by Friedrich Wilhelm Bessel (1784–1836) in 1838, this star lies at an estimated distance of 10.4 light-years, corresponding to an annual parallax of just 0.314 seconds of arc. The star 61 Cygni had been christened the "Flying Star" in 1792 by Piazzi who discovered its unusually large angular motion across the sky; if all stars moved at the same velocity perpendicular to the line of sight, such a large angular motion would indicate a nearby distance. Modern measurements provide a parallax of 0.287 18 seconds of arc for 61 Cygni, yielding a distance of 11.36 light-years or about 718 000 times further away than the Sun.

Within a century of Bessel's result, the annual parallax of about 2 000 nearby stars had been determined using long exposures on photographic plates, and the number tripled in succeeding decades. The closest star, known as Proxima Centauri, is 4.22 light-years away. And many of the brightest stars are hundreds of light-years away, so you

Focus 1.6 Light, the fastest thing around

It was once thought that light moves instantaneously through space, but we now know that it travels at a very fast, but finite, speed, which was first inferred from observations of Jupiter's moon Io in the 18th century. The King of France had directed Giovanni Domenico Cassini (1625–1712), the Director of the Paris Observatory, to use such observations to improve knowledge of terrestrial longitude and maps of France. While working at the observatory the Danish astronomer Ole Roemer (1644–1710) and Cassini noticed a varying time between eclipses of the satellite by the planet. Although Io's orbital period was about 42 hours, the duration of the orbit seemed to grow shorter when Jupiter was closer to Earth, and larger when it would move away, with a total time difference of about 22 minutes. Both Cassini and Roemer concluded that it was not the orbit itself that varied, but the time it took Jupiter's light to cross the Earth's orbit and reach our planet. Neither astronomer gave a value for the speed of light, which would have been equal to the diameter of the Earth's orbit divided by the time difference, or a velocity of $c = 2$ AU / 22 minutes or about 227 000 kilometers per second, where 1 AU is the mean distance between the Earth and the Sun.

More refined laboratory measurements during succeeding centuries indicated that light is always moving at a constant speed with the precise velocity of $c =$ 299 792.458 kilometers per second. Light emitted by any star will move through empty space for all time, never stopping or slowing down and never coming to rest. Moreover, nothing outruns light; it's the fastest thing around.

The unvarying speed of light was first demonstrated in 1887 by the American physicist Albert A. Michelson (1852–1931), assisted by his friend the chemist Edward W. Morley (1838–1923), when they attempted to precisely measure how the speed of light depends on the Earth's motion through a hypothetical, space-filling medium, the ether, in which light waves were supposed to propagate and vibrate.

As the Earth moves through the stationary medium, an ether wind should blow past the Earth in the direction of its motion, and the speed of light would vary, like a swimmer moving downstream or struggling upstream. But Michelson and Morley found that there was no detectable difference in the speed of light measured in the direction of the Earth's motion or at right angles to it. So the experiment meant that there was no light-carrying ether. It also implied that the velocity of light is constant, exactly the same in all directions and at all seasons, and independent of the motion of the observer.

The speed of light, c, enters into Albert Einstein's (1879–1955) *Special Theory of Relativity* through the factor

$$\gamma = \frac{1}{\sqrt{\left[1 - \left(\frac{v}{c}\right)^2\right]}}$$

for an object moving at a velocity v. We normally regard time as absolute and immutable, with nothing disturbing its relentless, steady tick. But for Einstein, time was relative and variable. In rapid travel, the rate at which time flows decreases, so clocks run slower by the factor γ. Lengths are diminished at high speed, shrinking in the direction of motion by the amount γ. At very high velocities, all is relative, even mass, which increases with the speed by the same infamous γ factor. If any material body reached the velocity of light, it might shrink into an insignificant nothing, perhaps even disappearing, while time might be stretched to infinity. That doesn't happen because an object's mass increases without bound when it moves as fast as light, and there is nothing that can propel it so fast.

can walk outside at night and see stars whose light was emitted before your parents were born.

Distant stars with parallaxes smaller than 0.05 seconds of arc cannot be measured with Earth-based telescopes because of atmospheric distortion that limits their angular resolution. However, instruments aboard the *HIPPARCOS* satellite, which orbited the Earth above its atmosphere in the 1990s, pinpointed the positions of more than 100 000 stars with an astonishing precision of 0.001 seconds of arc, determining the parallax and distance of many of them out to a few hundred light-years.

How big, massive and hot is the Sun?

Once an accurate value for the mean distance between the Earth and the Sun is known, we can use it with the orbital period of the Earth to infer the mass of the Sun, $M_\odot$, from Newton's formulation of Kepler's third law:

$$\text{Sun's mass} = M_\odot = 5.9165 \times 10^{11} \times \frac{(1\,\text{AU})^3}{P_E^2}$$
$$= 1.989 \times 10^{30}\ \text{kilograms}$$

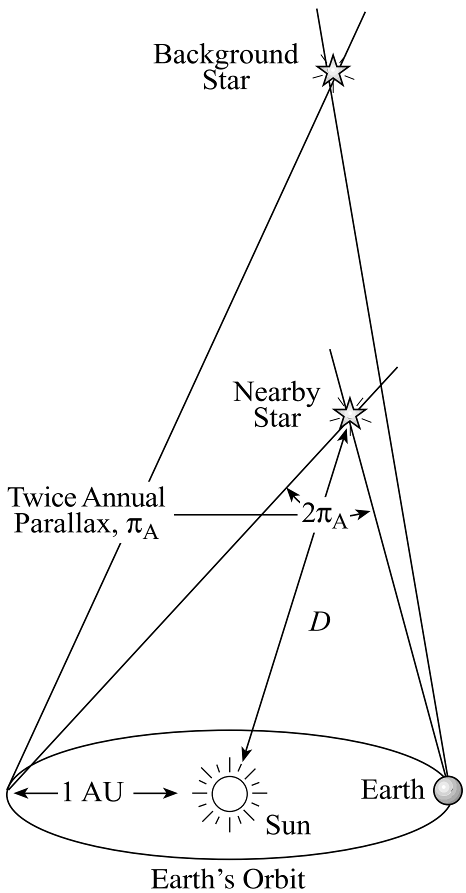


Fig. 1.20 Annual parallax When a distant and nearby star are observed at six-month intervals, on opposite sides of the Earth's orbit around the Sun, astronomers measure the angular displacement between the two stars. It is twice the annual parallax, designated by π_A, which can be used to determine the distance D of the nearby star. The German astronomer Friedrich Wilhelm Bessel (1784–1846) announced the first reliable measurement of the annual parallax of a star in 1838.

using the Sun–Earth distance of $1\,\text{AU} = 1.495\,978\,7 \times 10^{11}$ meters, and the Earth's orbital period of $P_E = 1\,\text{year} = 3.1557 \times 10^7$ seconds.

The linear radius of the Sun, $R_\odot$, can be determined from its angular diameter, θ, using:

$$R_\odot = \theta \times (1\,\text{AU})/2 = 6.955 \times 10^8 \text{ meters}$$

where $\theta = 31.97$ minutes of arc $= 0.0093$ radians, and a full circle subtends 2π radians and 360 degrees.

The mass density of the Sun, $\rho_\odot$, is obtained by dividing this mass by the Sun's volume, $4\pi R_\odot{}^3/3$, where the Sun's radius $R_\odot = 6.955 \times 10^8$ meters. It is $\rho_\odot = 1409$ kilograms per cubic meter, only about one-fourth of the mass density of the Earth, which is 5520 kilograms per cubic meter.

The entire Sun is nothing but a big luminous ball of gas, very hot and concentrated at the center and cooler and more tenuous further out. From the Sun's size and luminous output, we can infer the temperature of its visible disk, 5780 kelvin (Focus 1.7).

The material deep down inside the Sun must become hotter and more densely concentrated to support the overlying weight and to keep the star from collapsing. Calculations show that the temperature reaches 15.6 million (1.56×10^7) kelvin at the center of the Sun (also see Focus 1.7). The center is also extremely compacted with a density of 151 300 kilograms per cubic meter. All of the essential physical properties of the Sun are given in Table 1.5.

What is the Sun made out of?

The ingredients of the Sun can be determined when the intensity of sunlight is spread out into its visible wavelengths or different colors. Such a display is called a spectrum, and the study of spectra is known as spectroscopy.

Each chemical element or compound produces a unique set, or pattern, at certain specific wavelengths in the spectrum, and only at those wavelengths. They resemble a barcode or a fingerprint that can be used to identify the element or compound.

A hot, glowing body like the Sun emits radiation at all wavelengths, with a continuous spectrum. If this radiation passes through a cool, tenuous gas, such as the outer layers of the Sun's visible atmosphere, part of the radiation is absorbed at discrete wavelengths. These spectral features are called absorption lines because they look like a line in the spectrum. If the gas is heated to incandescence, it will emit radiation at the same specific wavelengths, and the spectral features are called emission lines. The patterns of either the absorption or emission lines tell us the atoms or molecules that are present in the gas.

The technique of astronomical spectroscopy was first developed using the bright light of the Sun. When its spectrum is examined carefully, with fine wavelength resolution, numerous fine, dark absorption lines are seen crossing the rainbow-like display (Fig. 1.21). The separate colors of sunlight are somewhat blurred together when coarser spectral resolution is used, and the dark places are no longer found superimposed on its spectrum.

These dark gaps, or absorption lines, were first noticed in 1802 by the English chemist William Hyde Wollaston (1766–1828), and investigated in far greater detail by the Bavarian telescope-maker Joseph Fraunhofer (1787–1826). By 1815, Fraunhofer had catalogued the wavelengths of hundreds of them, assigning Roman letters A, B, C, . . . to the darkest and most prominent of them, starting from

Focus 1.7 Taking the Sun's temperature

Satellites have been used to accurately measure the Sun's total irradiance just outside the Earth's atmosphere, establishing the value of the solar constant:

$$f_\odot = 1361 \text{ joule per second per square meter}$$

where one joule per second is equivalent to one watt. The solar constant is defined as the total amount of radiant solar energy per unit time per unit area reaching the top of the Earth's atmosphere at the Earth's mean distance from the Sun. We can use it to determine the Sun's absolute luminosity, $L_\odot$, from:

$$L_\odot = 4\pi f_\odot\,(1\,\text{AU})^2 = 3.854 \times 10^{26} \text{ joule per second}$$

where the mean distance between the Earth and the Sun is 1 AU $= 1.496 \times 10^{11}$ meters.

The effective temperature, $T_{e\odot}$, of the visible solar disk, called the photosphere, can be determined using the Stefan–Boltzmann law:

$$L_\odot = 4\pi\sigma\ R^2 T_{e\odot}^4$$

where the Stefan–Boltzmann constant is $\sigma = 5.670 \times 10^{-8}$ J m^{-2} K^{-1} s^{-1}, and the Sun's radius is $R_\odot = 6.955 \times 10^8$ meters. Solving for the temperature:

$$T_{e\odot} = [L_\odot/(4\pi\ R_\odot^2]^{1/4} = 5780 \text{ kelvin}$$

Incidentally, the Stefan–Boltzmann law applies to other stars, indicating that at a given temperature giant stars, with greater radii than most stars, have greater luminosity.

The temperature, $T_{c\odot}$, at the center of the Sun can be estimated by assuming that a proton must be hot enough and move fast enough to counteract the gravitational compression it experiences from all the rest of the star. That is:

$$\text{Thermal energy} = \frac{3}{2}\,k\,T_{c\odot} = \frac{G m_P M_\odot}{R_\odot} = \text{Gravitational potential energy}$$

where Boltzmann's constant k is $1.380\,66 \times 10^{-23}$ joule per kelvin, the gravitational constant G is 6.6726×10^{-11} m^3 kg^{-1} s^{-2}, the Sun's mass $M_\odot$ is 1.989×10^{30} kilograms, and the mass of the proton is $m_P = 1.6726 \times 10^{-27}$ kilograms. Solving for the central temperature we obtain:

$$T_{c\odot} = \frac{2 G m_P\ M_\odot}{3k\,R_\odot} = 1.56 \times 10^7 \text{ kelvin}$$

So the temperature at the center of the Sun is 15.6 million kelvin.

Table 1.5 Physical parameters of the Sun[a]

Mass, $M_\odot$	1.989×10^{30} kilograms (332 946 Earth masses)
Radius, $R_\odot$	6.955×10^8 meters (109 Earth radii)
Volume	1.412×10^{27} m^3 (1.3 million Earths)
Density (center)	151 300 kg m^{-3}
(mean), $\rho_\odot$	1409 kg m^{-3}
Pressure (center)	2.334×10^{11} bars
(photosphere)	0.0001 bar
Temperature (center) $T_{c\odot}$	15.6 million kelvin
(photosphere) $T_{e\odot}$	5 780 kelvin
(corona)	2 million to 3 million kelvin
Luminosity, $L_\odot$	3.854×10^{26} J s^{-1}
Solar constant, $f_\odot$	1361 J s^{-1} m^{-2} = 1361 W m^{-2}
Mean distance, AU	$1.495\,978\,7 \times 10^{11}$ m = 1.0 AU
Age	4.55 billion years
Principal chemical constituents (by number of atoms)	
Hydrogen	92.1 percent
Helium	7.8 percent
All others	0.1 percent

[a] Mass density is given in kilograms per cubic meter, or kg m^{-3}; the density of water is 1000 kg m^{-3}. The unit of pressure is bars, where 1.013 bars is the pressure of the Earth's atmosphere at sea level. The unit of luminosity is joule per second; power is often expressed in watts, where 1.0 watt = 1.0 joule per second.

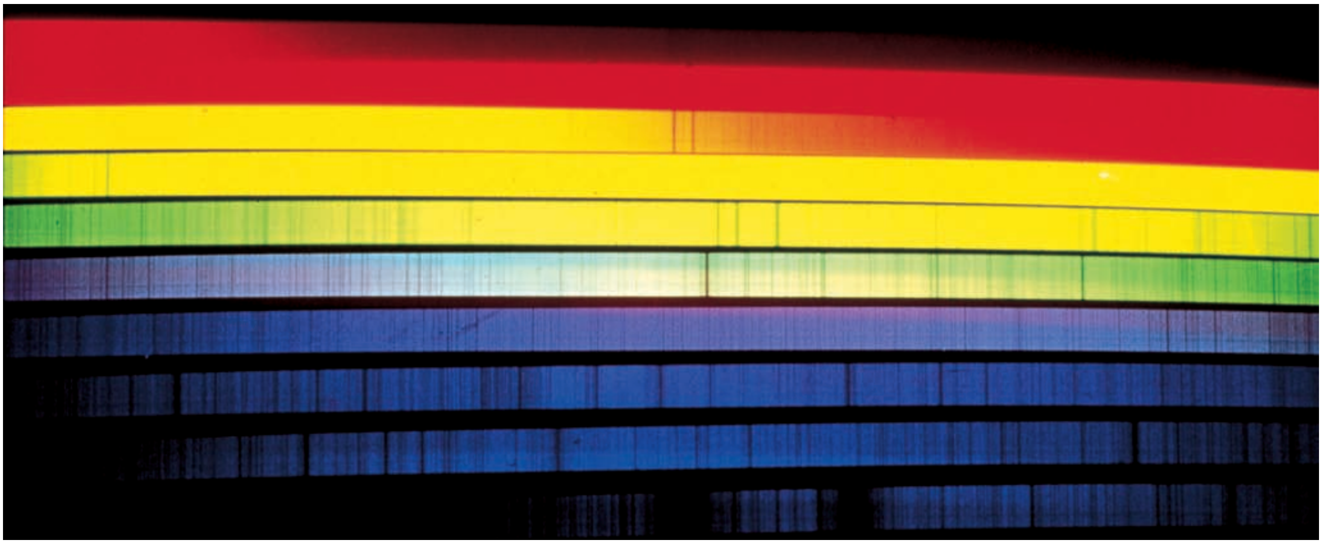


Fig. 1.21 Visible solar spectrum A spectrograph has spread out the visible portion of the Sun's radiation into its spectral components, displaying radiation intensity as a function of wavelength. When we pass from long wavelengths to shortest ones (*left to right* and *top to bottom*), the spectrum ranges from red through orange, yellow, green, blue and violet. Dark gaps in the spectrum, called Fraunhofer absorption lines, represent absorption by atoms or ions in the Sun. The wavelengths of these absorption lines can be used to identify the elements in the Sun, and the relative darkness of the lines helps establish the relative abundance of these elements. (Courtesy of NSO/NOAO.)

the long-wavelength, red end of the visible solar spectrum and progressing to its short-wavelength side.

An explanation for the Sun's absorption lines was provided in the mid 19th century in a laboratory in Heidelberg, Germany, where the chemist Robert Bunsen (1811–1899), inventor of the Bunsen burner, and his physicist colleague Gustav Kirchhoff (1824–1887) unlocked the chemical secrets of the Universe. When they vaporized an individual element in a flame, the hot vapor produced a distinctive pattern of bright emission lines whose unique wavelengths coincided with the wavelengths of some of the dark absorption lines in the Sun's spectrum, identifying that element as an ingredient of the solar gas. Solar lines designated by Fraunhofer by D and E are respectively ascribed to sodium and iron, while both the H and K lines are due to calcium.

The lightest element, hydrogen, was also identified in the solar spectrum, in 1862 by the Swedish physicist Anders Jonas Ångström (1814–1874); it accounts for Fraunhofer's C and D lines. Subsequent investigations of the great strength of these lines indicated that hydrogen is the most abundant element in the visible solar gases.

Since the Sun was most likely chemically homogeneous, a high hydrogen abundance was implied for the entire star. This accounts for the Sun's low mass density. We now know that hydrogen accounts for 92.1 percent of the number of atoms in the Sun, and that hydrogen is the most abundant element in most stars, in interstellar space, and in the entire Universe.

Helium, the second-most abundant element in the Sun, is so rare on Earth that it was first discovered in the Sun. A previously unknown emission line was initially noticed during a solar eclipse on 18 August 1868, at a wavelength of 587.56 nanometers near the two yellow sodium lines. Since this feature had no known Earthly counterpart, it was thought that a new chemical element had been discovered, which Norman Lockyer (1836–1920) named "helium" after the Greek Sun god Helios. Helium was not found on Earth until 1895, when William Ramsay (1852–1919) discovered it as a gaseous emission from a mineral called clevite. Helium accounts for 7.8 percent of the number of atoms in the Sun, and all the heavier elements in the Sun amount to only 0.1 percent (Table 1.6).

Today, helium is used on Earth in a variety of ways, including the inflation of party balloons and in its liquid state to keep sensitive electronic equipment cold. Though plentiful in the Sun, helium is almost non-existent on the Earth. It is so terrestrially rare that we are in danger of running out of helium during this century.

A more complete understanding of spectral lines followed the discovery that atoms are mostly empty space, just as the room you are sitting in is mainly empty. A tiny, heavy, positively charged nucleus lies at the heart of an atom, surrounded by a cloud of relatively minute, negatively charged electrons that occupy most of an atom's space and govern its chemical behavior. The charge of a proton is exactly equal to that of an electron, so the complete atom, in which the number of

Table 1.6 The ten most abundant elements in the Sun

Atomic number[a]	Element	Symbol	Number of atoms (silicon = 10.0)	Date of discovery
1	Hydrogen	H	279 000	1766
2	Helium	He	2 720	1868[b]
6	Carbon	C	101	(ancient)
7	Nitrogen	N	31.3	1772
8	Oxygen	O	23.8	1774
10	Neon	Ne	34.4	1898
12	Magnesium	Mg	10.7	1755
14	Silicon	Si	10.0	1823
16	Sulfur	S	5.15	(ancient)
26	Iron	Fe	9.00	(ancient)

[a] The atomic number is equal to the number of protons in the nucleus of an atom.
[b] Helium was discovered on the Sun in 1868, but it was not found on Earth until 1895.

electrons equals the number of protons, is electrically neutral.

Hydrogen is the simplest atom, consisting of a single electron circling around a single proton. The nucleus of a more complex atom includes neutrons, which were proposed in order to keep the nuclear protons from repelling each other out of the nucleus. The nucleus of helium, for example, contains two neutrons and two protons, and has two electrons in orbit to balance the charge of the two protons. Ernest Rutherford (1871–1939) established the existence of a small, positively charged nucleus at the center of the atom in 1911; James Chadwick (1891–1974) discovered the neutron in 1932.

The sole electron in a hydrogen atom revolves about the central proton according to very specific rules that were proposed to explain observations of regularities in the Sun's hydrogen lines; adjacent lines of the hydrogen atom systematically crowd together and become stronger at shorter wavelengths.

The Swiss mathematics teacher Johann Balmer (1825–1898) published an equation that describes the regular spacing of the wavelengths of the four lines of hydrogen detected in the spectrum of visible sunlight, and they are still known as Balmer lines. The strongest one, with a red color, is also called the hydrogen alpha line.

In the early 20th century, the Danish physicist Niels Bohr (1885–1962) explained Balmer's equation by an atomic model, now known as the Bohr atom (Fig. 1.22), in which the electron in a hydrogen atom revolves about the proton in specific orbits with definite quantized values of energy, which are characterized by a quantum number that takes on integer values of $n = 1, 2, 3, 4, \ldots, \infty$ (infinity).

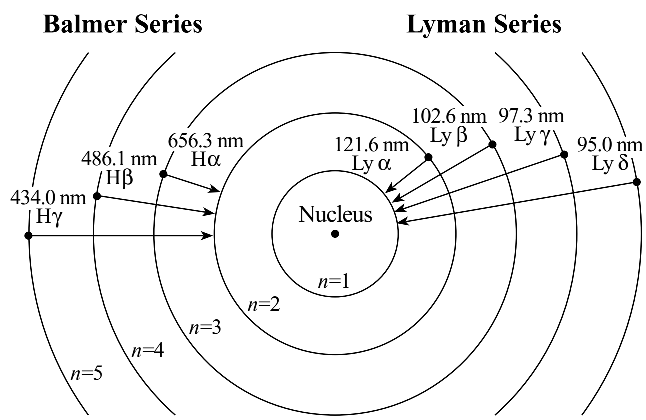


Fig. 1.22 Bohr atom In this model, proposed in the early 20th century by the Danish physicist Niels Bohr (1885–1962), a hydrogen atom's one electron revolves around the hydrogen nucleus, a single proton, in well-defined orbits described by the integer $n = 1, 2, 3, 4, 5, \ldots$ An electron absorbs or emits radiation when it makes a transition between these allowed orbits. The electron can jump upward, to orbits with larger n, by absorption of a photon of exactly the right energy, equal to the energy difference between the orbits; the electron can jump down to lower orbits, of smaller n, with the emission of radiation of that same energy and wavelength. Transitions that begin or end on the $n = 2$ orbit define the Balmer series that is observed at visible wavelengths. They are designated by Hα, Hβ, Hγ, ... The Lyman series, with transitions from the first orbit at $n = 1$, is detected at ultraviolet wavelengths. The orbits are not drawn to scale for the size of the radius increases with the square of the integer n.

Focus 1.8 Satellite motions and planetary mass

Kepler's third law also applies to the motion of a natural satellite, which can be used to determine the planet mass M_P from the expression:

$$M_P = \frac{4\pi^2}{G}\frac{a_S^3}{P_S^2} = 5.9 \times 10^{11}\frac{a_S^3}{P_S^2} \text{ kilograms}$$

for a satellite orbiting the planet at a distance a_S in meters and with period P_S in seconds, with $\pi = 3.141\,59$ and the Newtonian constant of gravitation $G = 6.673 \times 10^{-11}\ \mathrm{m^3\,kg^{-1}\,s^{-2}}$. The Earth's Moon, for example, orbits the Earth at a mean distance of $a_M = 384\,400$ kilometers $= 3.844 \times 10^8$ meters with a period of $P_M = 27.322$ days $= 2.36 \times 10^6$ seconds, giving a mass $M_E = 60 \times 10^{23}$ kilograms for the Earth.

As another example, we can obtain Jupiter's mass from observations of any one of its four large satellites. The satellite Io's period is $P_S = 1.77$ days $= 152\,928$ seconds and Io's distance from Jupiter is $a_S = 4.22 \times 10^6$ meters, yielding a mass $M_J = 4\,\pi^2\,a_S{}^3\,/\,(GP_S{}^2) = 1.9 \times 10^{27}$ kilograms, or about one-thousandth the mass of the Sun.

This expression can also be used to determine the orbital distance of an artificial geosynchronous satellite that is launched to always hover above the same location on the Earth, using:

$$a_S^3 = \frac{G\,M_E\,P_S^2}{4\pi^2} = 1.7 \times 10^{-12}\,M_E\,P_S^2 \text{ meters}$$

This yields $a_E = 4.2 \times 10^7$ meters $= 42\,000$ kilometers from the Earth's center for an orbital period of one day, or $P_S = 86\,400$ seconds and an Earth mass of $M_E = 6 \times 10^{24}$ kilograms. Since the Earth's mean radius is 6 400 kilometers, the geosynchronous satellite orbits at a distance of about 35 600 kilometers above the ground.

The energy of the nth orbit is proportional to the inverse square of n, or to $1/n^2$.

The electron only emits or absorbs radiation when jumping between these allowed orbits, each jump being associated with a specific energy equal to the difference in orbital energies, and to a single wavelength, like one pure note. If an electron jumps from a low-energy orbit to a high-energy one, it absorbs radiation at this wavelength; radiation is emitted at exactly the same wavelength when the electron jumps the opposite way. Bohr was awarded the 1912 Nobel Prize in Physics for his investigation of the structure of atoms and the radiation emanating from them.

1.5 Terrestrial and giant planets

The eight major planets have been divided into two groups, the terrestrial and the giant planets, which differ in size, composition and distance from the Sun. The four planets close to the Sun, Mercury, Venus, Earth and Mars, are known as terrestrial planets because they are similar to the Earth. These inner planets are rocky and relatively compact and dense. In contrast, the four giant planets, Jupiter, Saturn, Uranus and Neptune, which reside in the outer parts of the planetary system, are big, gaseous and have relatively low mean mass densities. Unlike the inner terrestrial planets, rings and numerous satellites also encircle each of the outer giant planets.

Once the distance of a planet is known, its radius can be determined from its angular extent. For instance, Jupiter has an angular diameter of $\theta = 46.86$ seconds of arc when it is closest to Earth, or at a distance of $D = 4.2$ AU from us. That corresponds to a radius of $R = 7.14 \times 10^7$ meters. You can do the arithmetic yourself using $R = \theta D/2$ with 1 AU $= 1.496 \times 10^{11}$ meters and converting the angle to radians with 1 radian $= 2.062\,65 \times 10^5$ seconds of arc. The radius of Jupiter is 11.2 times the Earth's radius, and its volume is more than 1000 times that of the Earth.

The distance to a planet can be combined with the angular separation of one of its satellites to determine the orbital distance of that satellite from its planet, which can then be combined with the satellite's orbital period to establish the planet's mass. The motion of a satellite is governed by the mass of its planet, all in accordance with the inverse square law of gravity and Kepler's third law (Focus 1.8). Newton used it centuries ago to infer the masses of the Earth, Jupiter and Saturn from their satellite motions, obtaining values comparable to modern ones.

The masses of the other giant planets can be inferred from their satellites in the same way. The technique can also be used to obtain the Earth's mass, from the motion of its Moon. If a planet has no satellite, like Mercury and Venus, its mass can be obtained from detailed observations of its gravitational effects on spacecraft that pass or orbit near it.

The mean, or average, mass density of planet can be computed by dividing its mass, in kilograms, by its volume in cubic meters. The volume of a planet of radius R is $4\pi R^3/3$, so the mean mass density of a planet of mass M is $3M/(4\pi R^3)$. For instance, the mean mass density of Jupiter is $1330\ \mathrm{kg\,m^{-3}}$ (kilograms per cubic meter). That is comparable to the Sun's mean mass density of $1409\ \mathrm{kg\,m^{-3}}$. The mean mass density of the major planets, which is also known as their bulk density, is given in Table 1.7 together with their angular size at closest approach, radius, and mass. Here they are divided into two groups: the four

Table 1.7 Angular diameter, radius, mass and bulk density of the major planets

Terrestrial planets	Mercury	Venus	Earth	Mars
Angular diameter (seconds of arc)[a]	10.9	61.0		17.88
Equatorial radius, R_P (km)	2439.7	6051.8	6378.14	3396.19
Equatorial radius, R_P ($R_E = 1.0$)	0.382	0.949	1.000	0.533
Mass, M_P (kg)	3.3010×10^{23}	4.8673×10^{24}	5.9722×10^{24}	6.4169×10^{23}
Mass, M_P ($M_E = 1.0$)	0.0553	0.815	1.000 00	0.107
Bulk density (kg m^{-3})	5429	5243	5513	3934
Giant planets	**Jupiter**	**Saturn**	**Uranus**	**Neptune**
Angular diameter (seconds of arc)[a]	46.86	19.52	3.60	2.12
Equatorial radius, R_P (km)	71 492	60 268	25 559	24 764
Equatorial radius, R_P ($R_E = 1.0$)	11.19	9.46	3.98	3.81
Mass, M_P (kg)	1.8981×10^{27}	5.6832×10^{26}	8.6881×10^{25}	1.0241×10^{26}
Mass, M_P ($M_E = 1.0$)	317.89	95.18	14.54	17.13
Bulk density (kg m^{-3})	1326	687	1270	1638

[a] The largest angular diameter seen from the Earth when the planet is at its closest approach to Earth.

Table 1.8 Distribution of mass in the solar system

Mass of the Sun	$M_\odot = 1.989 \times 10^{30}$ kilograms
Mass of Jupiter	$M_J = 1.899 \times 10^{27}$ kilograms
Mass of the Earth	$M_E = 5.974 \times 10^{24}$ kilograms
Total mass of the planets	2.668×10^{27} kilograms $= 446.6\ M_E$
Total mass of the satellites	6.2×10^{23} kilograms $= 0.104\ M_E$
Total mass of the Kuiper-belt objects	3.0×10^{23} kilograms $= 0.05\ M_E$
Total mass of the asteroids	1.8×10^{21} kilograms $= 0.0003\ M_E$
Total mass of the planetary system	2.669×10^{27} kilograms $= 446.7\ M_E = 0.001\,34\ M_\odot$

rocky, dense and relatively small terrestrial planets, and the four giant, massive, low-density worlds.

When the mass of the Sun and planets are determined, we find that the Sun doesn't just lie at the heart of our solar system; it dominates it. Some 99.866 percent of all the matter between the Sun and halfway to the nearest star is contained in the Sun (Table 1.8). All of the objects that orbit the Sun – the planets and their satellites, the comets and the asteroids – add up to just 0.134 percent of the mass in our solar system. As far as the Sun is concerned, the planets are insignificant specks, left over from its formation and held captive by its massive gravity.

1.6 What is inside the major planets?

Density is a measure of compactness of an object, and all planets have greater density in their compact centers. The temperature in the interiors of the giant planets also increases with depth. To understand this, imagine a hundred mattresses stacked into a pile. The mattresses at the bottom must support those above so they will be squeezed thin. Those at the top have little weight to carry, and they retain their original thickness. The material at the center of a gaseous giant planet is similarly squeezed into a smaller volume by the overlying material, so the central regions become hotter and more densely concentrated. All of the giant planets are hot inside, with temperatures that increase with depth, and three of them still radiate more heat from their cloud tops than absorbed from the Sun.

The rocky terrestrial planets were all so hot in their formative stages, beginning about 4.6 billion years ago, that their interior rock and metal melted and gravity separated them by density. The denser material sank toward the center, while less dense rocky material remained closer to the surface. This process is called differentiation, since the internal layers are composed of different material. The Moon, for example, contains a small, partially molten liquid core, while the Earth, Mars and Mercury have larger ones hidden deep within their surfaces.

The highest-density material, consisting primarily of metals such as iron, resides in the central core of the terrestrial planets. Rocky material of moderate density is found in a thick mantle that surrounds the core. The mantle consists mostly of rocky silicate material containing silicon, oxygen and other elements. A terrestrial planet's thin outer crust contains the lowest density rocks such as granite and volcanic basalt.

Although the terrestrial planets and rocky satellites started off with hot interiors, as the result of their origin by colliding objects and the heat associated with radioactive decay, they have cooled from the outside in as time went on. After all, we now walk across solid rock rather than the originally molten crust.

When all the radioactive elements inside a rocky planet or moon are depleted, or decay into non-radioactive elements, there will be no more sources of internal heat, and these objects should become cooler inside while retaining a high internal pressure. As the remaining heat escapes from the surface, more heat will flow upward to replace it, until their interior is no hotter than its surface.

Since the total amount of internal heat depends on the volume, the time required to lose its heat depends on the ratio of the surface area $4\pi R^2$ to its volume $4\pi R^3/3$, so the rate of heat loss scales as $3/R$ for a planet of radius R. This means that larger planets or satellites with a greater radius will cool more slowly than smaller ones, provided they started off with the same internal temperature. Moreover, bigger objects most likely started off with more internal heat and higher temperatures. These two factors explain why the Moon's interior is so much cooler than the Earth's interior.

The terrestrial planets all now have solid surfaces, and on these surfaces is preserved a long record of their evolution. They have been shaped and modified by geological processes that include impact craters, produced when meteoroids similar to today's asteroids or comets struck them, and the volcanic eruption or flow of molten rock, from the planet's interior onto its surface. They are discussed in the next chapter.

2 The new close-up view from space

- Close examination of the planets and moons began when spacecraft flew past them, providing an initial reconnaissance. Orbiting spacecraft that mapped out the global terrain of the Earth's Moon, Venus and Mars, as well as the realms of Jupiter and Saturn, followed this. Probes have been parachuted down into the atmospheres of Jupiter, Titan and Venus, and landers and rovers have been sent to the surfaces of the Moon and Mars.
- The space-age investigation of the solar system began in a cold-war competition between the Soviet Union, which launched the first artificial satellite, and the United States, which won the race to the Moon.
- The *Voyager 1* and *2* flyby spacecraft transformed our understanding of the four giant planets, Jupiter, Saturn, Uranus and Neptune, and revealed fascinating, unexpected aspects of their moons and rings.
- The *Giotto* spacecraft was the first to provide a close-up view of a comet, showing that its nucleus is a black, city-sized chunk of water ice and dust that emits sunward jets of water when passing near the Sun.
- Orbiting spacecraft have greatly increased the time for study of the planets and moons, revealing ancient water flow on Mars, vast outpourings of lava on Venus, Jupiter's volcanic moon Io, and an ice-covered ocean on its satellite Europa, and Saturn's marvelous rings, water-spewing satellite Enceladus, and haze-shrouded moon Titan.
- Three rovers have explored the surface of Mars and provided evidence for water flow across its surface roughly 4.0 billion years ago.
- The *Huygens Probe* and radar from the orbiting *Cassini* spacecraft have discovered rain, rivers and lakes of liquid methane on Saturn's moon Titan.
- The planets and moons gathered together as the result of the collisions of smaller bodies beginning about 4.6 billion years ago.
- Every solid planet or satellite contains impact craters, but in different amounts that depend on the ages of their surfaces.
- Impact craters on the Moon, Mercury, and Jupiter's icy moon Callisto all record an ancient, intense rain of meteorites, which occurred about 4.0 billion years ago, and a continued cosmic bombardment at lower rates since then.

- Ancient craters on the relatively young surfaces of Earth, Venus and Io have been erased by geologic and volcanic activity.
- The round craters on the Moon were formed by the explosive impact of large meteorites that came from interplanetary space, releasing enormous energy, melting rock and excavating circular craters with raised rims on impact.
- Massive impacts of exceptionally big objects gouged out large impact basins on the Earth's Moon and the planet Mercury. Concentric, ring-like rims as tall as mountains can surround the impact basins, and the basins have been subsequently filled with lava.
- Giant impacts in the early history of the solar system may account for the origin of the Earth's Moon, the removal of Mercury's low-density mantle, the backwards rotation direction of Venus, and the crustal dichotomy between the low-lying northern plains and southern highlands of Mars.
- Upon impact with the surface of Mars, ground water ice can be melted, lubricating the ejected material that flows like mud.
- The material ejected from craters on Venus has been shaped by the planet's hot, thick atmosphere into asymmetric, lobate forms. Small impacting projectiles have been burnt up in the thick atmosphere, so there are no small craters on Venus.
- Internal heat can be produced by radioactive decay of rocks inside a terrestrial planet, or within a satellite as the result of varying gravitational interaction with its planet. The giant planets still retain the heat of their formation.
- Molten rock, or magma, that is localized in underground chambers of a planet can rise to the surface and cause two types of basaltic volcanism – tall shield volcanoes and smooth volcanic flows known as plains.
- Earth has unique underwater volcanoes found in mid-ocean ridges that supply a spreading sea-floor, as well as chains of hot-spot volcanoes, such as the Hawaiian Islands, and volcanoes arising from the downward plunge of moving plates.
- Upwelling of internal magma is cracking part of Africa open, in a great rift valley.
- Extensive lava flows filled large impact basins on the Moon, creating the dark lunar maria between 3.9 and 3.2 billion years ago.
- Ancient, smooth volcanic flows on Mercury have obliterated small craters, filled the interiors of large impact basins, and spread out between large craters, producing about 40 percent of the planet's surface. Most of this volcanic activity occurred after the heavy bombardment about 4.0 billion years ago, but before some craters were formed on the smooth plains.
- Extensive volcanic activity on Venus resurfaced the planet about 750 million years ago.
- Mars has the tallest volcanoes in the solar system, and most of its northern hemisphere is covered with volcanic flows of lava.
- The volcanoes on Jupiter's satellite Io have turned the satellite inside out; it is heated inside by the tidal flexing action of nearby massive Jupiter.
- Liquid water flows out of cracks in the icy surface of Jupiter's moon Europa, and erupts as jets of water ice and water vapor from Saturn's moon Enceladus.

- Volcanoes of ice may have created some of the features now frozen into the bright smooth surface of Neptune's largest moon, Triton; dark geyser-like plumes have been observed in the process of eruption on the satellite.
- Seventy-one percent of the Earth's surface is covered with liquid water, and our bodies are largely composed of water.
- Dark, permanently shadowed regions inside craters in the Moon's polar regions could contain water. The *Clementine*, *Lunar Prospector*, *Chandrayaan-1* and *LCROSS* spacecraft have provided evidence of very small amounts of water on the Moon.
- Strong radar echoes from the highly reflective polar regions of Mercury suggest that thick deposits of water ice reside in the permanently shadowed interiors of craters near the planet's poles.
- Although Venus is now dried out, it may have once contained a small ocean.
- Small amounts of water vapor are found in the atmosphere of Mars, together with clouds and fogs of water ice.
- Vast amounts of frozen water now exist in the polar caps of Mars and beneath the surface of the polar, mid-latitude, and equatorial regions of the red planet.
- It cannot now rain on Mars, and liquid water cannot now exist for any length of time on the planet's surface.
- Catastrophic floods and deep rivers once carved channels on Mars, and an ancient ocean may have once covered the planet's northern lowlands.
- Saturn's rings consist of billions of particles of water ice.
- Jupiter's satellite Europa is covered with bright, smooth water ice, which has cracked due to the contorting tidal effects of Jupiter's strong gravity. The warmth generated by tidal heating may have been sufficient to form an ocean of liquid water below Europa's icy covering.
- Magnetic measurements provide indirect evidence for an ocean of salty, liquid water below the icy crust of Jupiter's satellite Europa.
- Jupiter's satellite Ganymede also probably contains an ocean of liquid water under its ice-covered surface.
- Saturn's satellite Enceladus has a frozen covering of water ice, and the moon emits icy jets, feeding the E ring that encircles Saturn.

2.1 Flybys, orbiters, probes and landers

We live at an incredible time, when all of the major planets and most of their satellites have been viewed close up with the inquisitive eyes of robotic spacecraft, revealing awesome, unanticipated features that cannot be seen in any other way. No two of these fascinating new worlds are exactly the same. Most of them have been investigated many times, with increasingly sophisticated instruments aboard many different spacecraft. At the same time, ground-based telescopes and telescopes aboard Earth-orbiting satellites have provided other new insights about the planets and their moons.

This captivating voyage of discovery began close to home, in 1957, when the Soviet Union launched the first artificial satellite, the beeping *Sputnik*, stunning the American public and initiating the space race with the United States. The subsequent cold-war competition included the Russian *Luna 3* that swung once around the far side of the Earth's Moon, which had never been seen before, and

Fig. 2.1 Lunar rover The battery-powered lunar rovers, used in the last three *Apollo* missions, could carry two astronauts and all their equipment for kilometers across the lunar surface. The astronauts deployed instruments and returned rocks to the *Lunar Module*, which carried them back to their orbiting *Command Module* while the rover remained on the Moon. Because there is no substantial atmosphere, water or weather on the Moon, both the rover and the footprints in the lunar soil may last for millions of years. By that time micrometeorites will have pitted the rover and erased the footprints. (Courtesy of NASA.)

culminated on 20 July 1969 when the *Lunar Module Eagle* carried two *Apollo 11* astronauts to the shores of the Moon's Sea of Tranquility.

Altogether twelve astronauts roamed the surface of the Earth's Moon, during the four-year *Apollo* program, dissipating the Soviet Union's lead in space, tarnishing the image of Soviet competence, and teaching the American people that even the frontier of space can be conquered with resolve and willpower, especially in a democratic nation that stresses individual freedom.

The exploration of our Moon served as a stepping-stone to the planets, and established a blueprint for subsequent planetary missions. For both our Moon and the planets, the initial reconnaissance was provided by spacecraft that flew by them, obtaining just a brief glimpse. In the lunar case, three *Ranger* spacecraft were also sent crashing into the Moon's surface, transmitting high-definition pictures on the way down. This was followed by a more-detailed exploration with orbiting spacecraft, such as *Lunar Orbiter 1, 2, 3, 4* and 5, that can circle a moon or planet many times and map out its global terrain. Like the explorers of new territories on Earth, the orbiters were sent to reconnoiter, to get the lay of the land, and to disclose possible dangers awaiting future visits. The next step involves landers, like *Luna 9* and *13*, and *Lunar Surveyor 1, 3, 5,* 6 and *7,* that explore the surfaces and probes that plunge into the atmospheres.

So far, humans have only visited the Moon. An estimated half billion people watched the televised first visit, on 20 July 1969, when Neil Armstrong (1930–) moved cautiously down the lander's ladder and stood firmly on the fine-grained lunar surface. An ancient dream had come true – man had set foot on another world.

In all, there have been six manned landings on the Moon, beginning with *Apollo 11* in July 1969 and ending with *Apollo 17* in December 1972. The actual landings were performed by the bug-like *Lunar Module* that separated from the main spacecraft while in orbit around the Moon, and returned to it. At first the astronauts traveled on foot, staying near to the *Lunar Module*, but they subsequently moved to more-remote locations in roving vehicles (Fig. 2.1). Altogether, 382 kilograms of rocks were brought back from the Moon for analysis in the terrestrial laboratory, determining the Moon's age, chemical composition, history and probable origin.

Robotic missions then took over, first exploring Earth's nearest planetary neighbors, Venus and Mars, and eventually ranging to the distant giant planets and beyond. The US *Mariner 2* spacecraft had already flown past Venus in 1962, and on the way detected a perpetual wind of

Table 2.1 Important flyby missions in the solar system

Spacecraft[a]	Launch date	Encounter date	Object	Discovery
Luna 3[b]	4 Oct. 1959	7 Oct. 1959	Moon	Photographed backside of Moon
Mariner 2	26 Aug. 1962	14 Dec. 1962	Venus	First successful planetary flyby Measured solar wind
Mariner 4	28 Nov. 1964	14 July 1965	Mars	Craters on Mars
Pioneer 10	3 Mar. 1972	3 Dec. 1973	Jupiter	Passage through asteroid belt
Mariner 10	3 Nov. 1973	29 Mar. 1974	Mercury	Heavily cratered surface
Voyager 1	5 Sept. 1977	5 Mar. 1979	Jupiter	Ring, volcanoes on satellite Io
Voyager 1		12 Nov. 1980	Saturn	Titan's dense atmosphere
Voyager 1		16 Dec. 2004		Termination shock of solar wind at 94 AU
Voyager 2	20 Aug. 1977	9 July 1979	Jupiter	
Voyager 2		25 Aug. 1981	Saturn	
Voyager 2		24 Jan. 1986	Uranus	Rings, magnetic field
Voyager 2		24 Aug. 1989	Neptune	Excess heat, winds, rings, magnetic field
Voyager 2		30 Aug. 2007		Termination shock of solar wind at 84 AU
Giotto[c]	2 July 1985	14 Mar. 1986	Comet Halley	Flyby, image of nucleus
Galileo	18 Oct. 1989	29 Oct. 1991	Asteroid 951 Gaspra	Flyby, image
Galileo		28 Aug. 1993	Asteroid 243 Ida	Ida and its moon Dactyl
NEAR[d]	17 Feb. 1996	27 June 1997	Asteroid 253 Mathilde	Flyby
Deep Space	24 Oct. 1998	28 July 1999	Asteroid 9969 Braille	Flyby
Deep Space	24 Oct. 1998	22 Sept. 2001	Comet Borrelly	Flyby; image of nucleus
Stardust	7 Feb. 1999	2 Jan. 2004	Comet Wild 2	Encounter; sample returned to Earth on 15 Jan. 2006
Hayabusa[e]	9 May 2003	30 Sept. 2005	Asteroid 25143 Itokawa	Encounter, rubble-pile asteroid, landing 19 Nov. 2005
MESSENGER	3 Aug. 2004	14 Jan. 2008	Mercury	Flybys on 14 Jan. 2008, 6 Oct. 2008, and 29 Sep. 2009; Mercury orbit insertion, 18 Mar. 2011
Dawn	27 Sept. 2007	Sept. 2011	Asteroid 4 Vesta	Flyby
Dawn		Feb. 2015	Dwarf-planet 1 Ceres	Flyby
New Horizons		July 2015	Pluto	Flyby

[a] Unless otherwise noted, these spacecraft are missions of the United States National Aeronautics and Space Administration (NASA).
[b] Spacecraft launched by the former Soviet Union.
[c] *Giotto* was a mission of the European Space Agency (ESA).
[d] The acronym *NEAR* stands for *Near Earth Asteroid Rendezvous*.
[e] *Hayabusa*, meaning "falcon" in Japanese, is a mission of the Japan Aerospace Exploration Agency (JAXA).

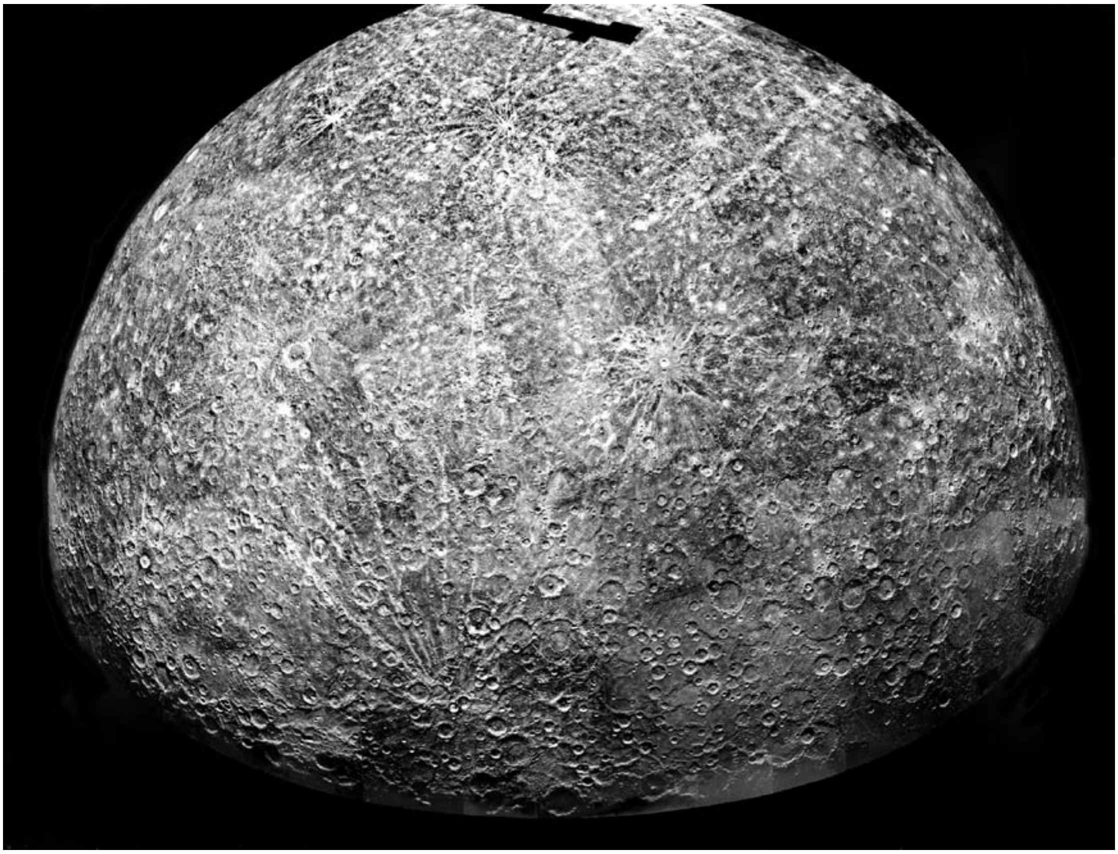


Fig. 2.2 First visit to Mercury A photomosaic of Mercury's southern hemisphere produced from images acquired by *Mariner 10* during its first encounter with the planet in March 1974. Mercury has a heavily cratered surface that resembles the lunar highlands. Bright rayed craters are also present on Mercury, as they are on the Moon. (Courtesy of NASA/JPL.)

charged particles in interplanetary space, emanating from the Sun. In 1970, an unmanned entry probe, the *Venera 7* spacecraft, was sent from the USSR into the thick, carbon-dioxide atmosphere of Venus, measuring the temperature and pressure all the way down to the surface. These data showed that the surface of Venus is hot enough to melt lead and that the atmosphere is ninety times as heavy as our air. This was the first craft to touch down successfully on another planet's surface, and it was followed by other *Veneras*, which had just enough time to send back pictures of the volcanic surface before being wiped out by the intense heat and pressure.

The first spacecraft to be launched on lengthy journeys beyond the Earth's Moon were flyby missions – the *Mariners*, *Pioneers* and *Voyagers* – that passed near the planets and their satellites to give us new vistas, unavailable from the ground, making important discoveries in the process (Table 2.1). *Mariner 10*, for example, obtained the first spacecraft photographs of the atmosphere of Venus in 1974, and traveled on to reveal the heavily cratered surface of Mercury (Fig. 2.2).

In 1972–74, the *Pioneer 10* and *11* missions to Jupiter showed that spacecraft could pass safely through the asteroid belt, blazing a trail for the extraordinarily successful *Voyager 1* and *2* flyby missions to far-distant worlds (Figs. 2.3, 2.4). Their "Grand Tour" included Jupiter (1979; Fig. 2.5), and Saturn (1980, 1981; Fig. 2.6). *Voyager 2* went on to Uranus (1986) and Neptune (1989), and both *Voyager 1* and *2* encountered the termination shock of the solar wind (2004, 2007).

Voyager 1 and *2* vastly improved our understanding of the atmospheres of the giant planets, and discovered unexpected rings, moons and magnetic fields. They also transformed the satellites of the giant planets into unique and distinctive places with diverse surfaces and in some cases atmospheres or magnetic fields.

Comets are so tiny and so far away that you cannot detect them until they come near the Sun, and their centers are then hidden within the brilliant glare of fluorescing gases and reflected sunlight. As a result, no one had ever seen the bare surface of a comet's nucleus until 1986, when the *Giotto* spacecraft peered into the core of comet Halley. It found the nucleus to be a black, oblong chunk of ice and dust, roughly the size of Paris or Manhattan (Fig. 2.7). At the moment of encounter, the comet was spewing out about 25 tons of water every second, propelled

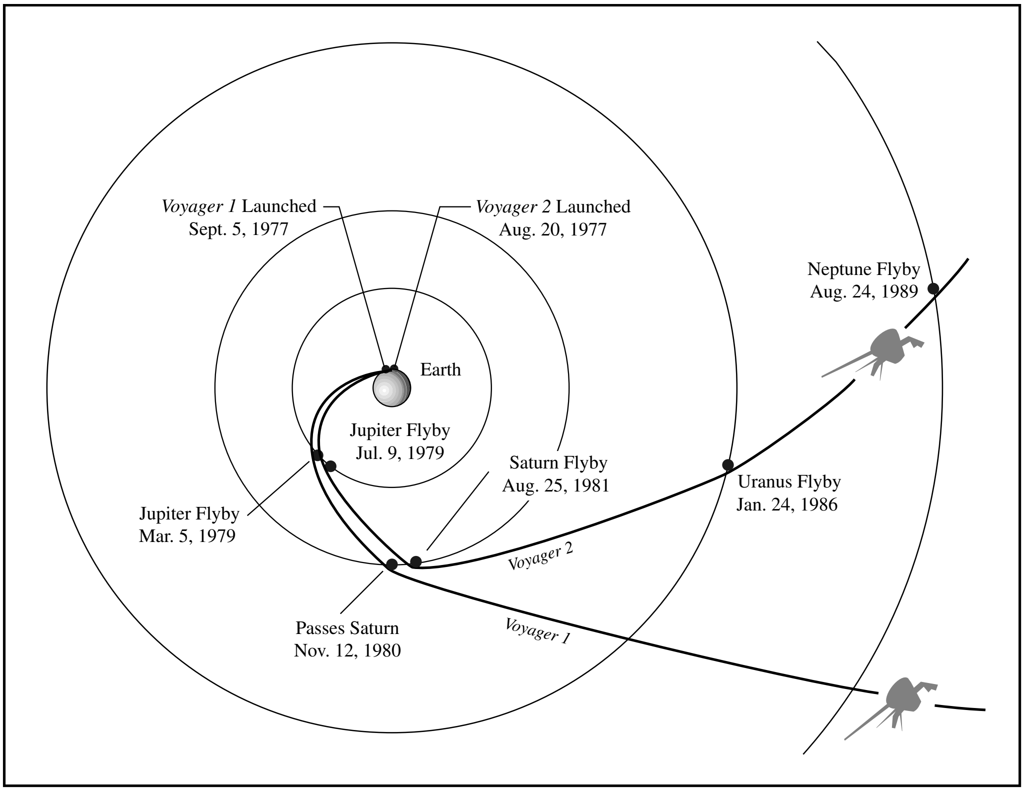


Fig. 2.3 Grand tour of *Voyager 1* and 2 The flights of the two *Voyager* spacecraft through the solar system. Both spacecraft were launched in 1977 and flew past Jupiter in 1979, transmitting remarkable details of the giant planet's weather and the surfaces of its four largest satellites. *Voyager 1* and *2* used the gravity of Jupiter to accelerate them on toward Saturn, providing close-up images of its rings and satellites in 1980–81. Saturn provided another gravity assist to propel *Voyager 2* on to Uranus, in 1986, and then on to Neptune in 1989. *Voyager 1* was targeted differently at Saturn, sacrificing its grand tour for close views of the satellite Titan. Both spacecraft have now crossed the termination shock of the solar wind, marking its outer boundary.

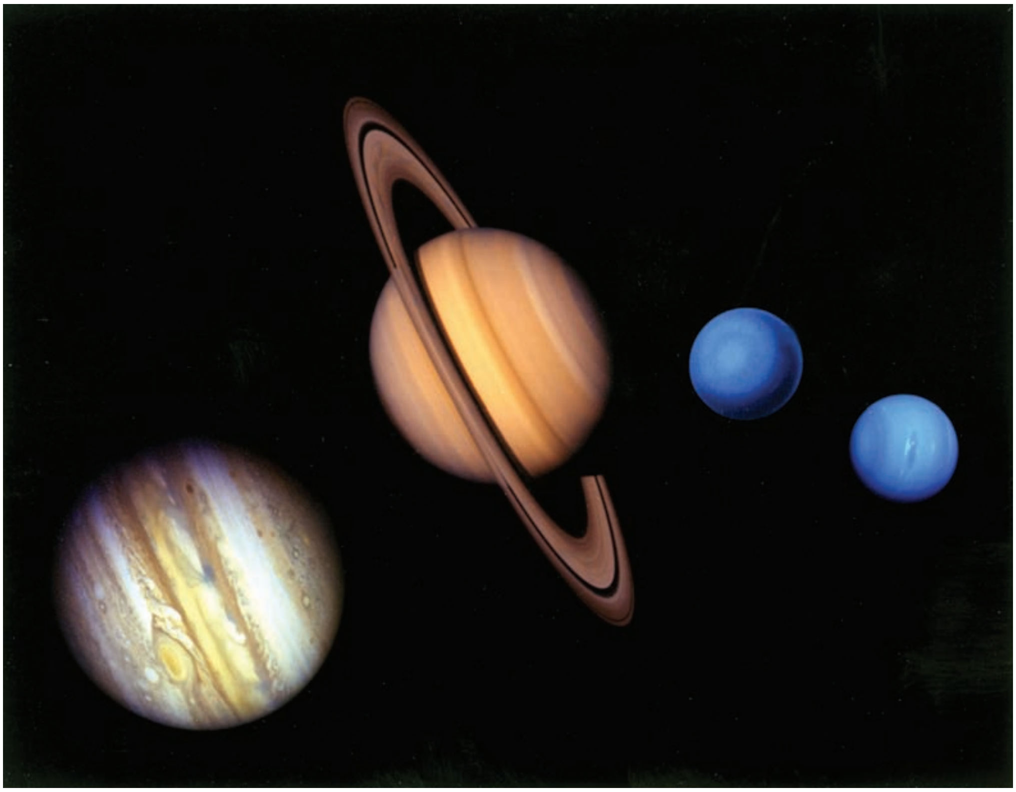


Fig. 2.4 New perspectives of the giant planets This montage of the giant planets was prepared from images taken by the *Voyager 2* spacecraft. They include banded Jupiter, ringed Saturn, and Uranus and Neptune with their blue clouds. (Courtesy of NASA/JPL.)

Fig. 2.5 Giant Jupiter's clouds Jupiter's clouded world with its alternating structure of light zones and dark belts. The two innermost Galilean satellites are also visible. Bright orange Io is seen just above the cloud tops, and icy-white Europa lies to the right. (Courtesy of NASA/JPL.)

into sunward jets by the vaporizing ice. So comets provide evidence that large quantities of water ice can be found in the outer solar system.

Flyby missions are still being used. The *MErcury Surface, Space ENvironment, GEochemistry and Ranging* (*MESSENGER*) spacecraft, for example, investigated roughly half of the previously unmapped surface of Mercury in 2008 and 2009, settling down into Mercury orbit in March 2011. *Dawn* is on its way for a rendezvous with the asteroids Vesta and Ceres, in 2011 and 2015, respectively. *New Horizons* is expected to fly by Pluto in 2015.

Orbiting spacecraft followed the initial planetary explorations using flybys. The orbiters greatly increased the time available for detailed study, often for years at a time (Table 2.2). They revealed many features that previous flyby missions had missed, and forever changed our view of the planets and their satellites.

The *Mariner 9* mission to Mars first demonstrated the extraordinary promise of planetary orbiters in 1971–72. The three previous flyby missions, *Mariner 4, 6* and *7*, had discovered the ancient cratered terrain on Mars, but missed all of the younger geologic features. The orbiting

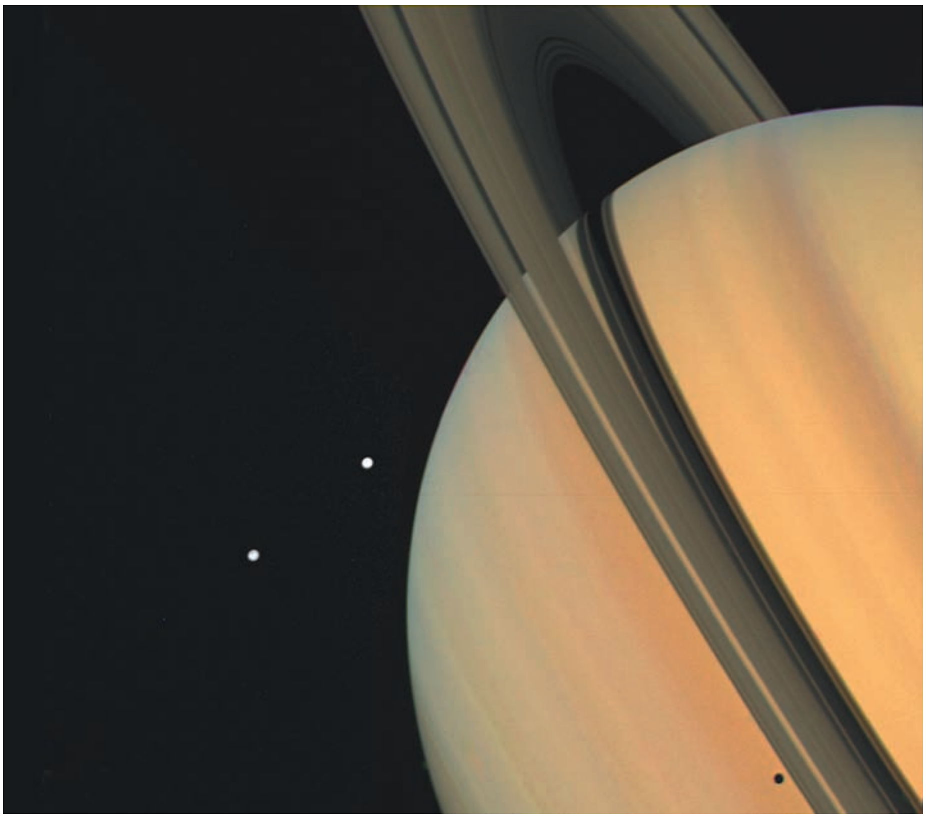


Fig. 2.6 Ringed Saturn with icy moons Saturn's yellow-brown clouds are swept into bands by the planet's rapid rotation. Two of its white moons (*left*), Tethys (*above*) and Dione (*below*), are covered with water ice. The shadows of Saturn's main rings and Tethys are cast onto the cloud tops. The outer A ring is separated from the central B ring by the dark Cassini Division, which is 3500 kilometers wide. This gap is so tenuous that the edge of Saturn can be seen through it. The faintest of Saturn's main rings, the inner C ring or crepe ring, is barely visible against the planet. This image was obtained from the *Voyager 1* spacecraft on 3 November 1980. (Courtesy of NASA/JPL.)

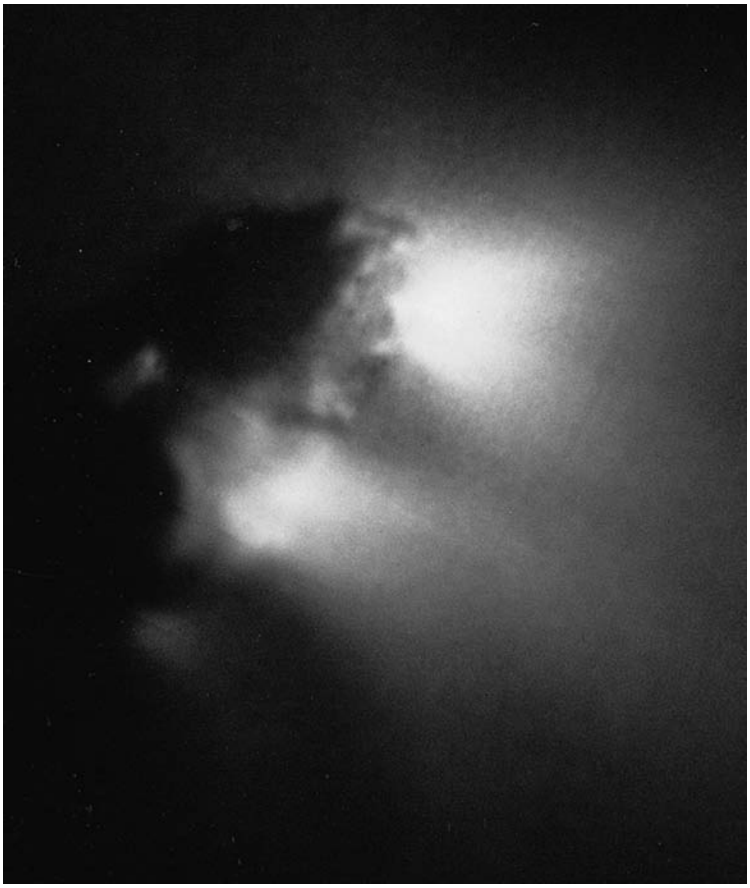


Fig. 2.7 The black heart of comet Halley
The coal-black nucleus of comet Halley is a dirty ball of ice, about the size of Paris or Manhattan. It is silhouetted against bright jets of water and dust that stream sunward (*right*) from at least three places that have been warmed by the Sun's radiation. In this projection, the nucleus measures 14.9 kilometers by 8.2 kilometers. This is a composite of images taken by the European Space Agency's *Giotto* spacecraft near its encounter with the nucleus of comet Halley on 14 March 1986. (Courtesy of ESA.)

Mariner 9 had sufficient time to completely explore the planet, revealing for the first time Mars' great volcanoes, the vast canyon system Valles Marineris, and evidence of ancient stream-beds and water erosion.

The *Viking 1* and *2* orbiters amplified and enhanced this new perspective of Mars, beginning in 1976 (Fig. 2.8). Close-up, high-resolution views of the surface of Mars were next obtained from the *Mars Global Surveyor* at the end of the 20th century and the beginning of the 21st century. The images showed much finer detail than those obtained with the *Viking* orbiters. Beginning in April 2001, the *2001 Mars Odyssey* mapped the amount and distribution of the chemical elements and minerals on the Martian surface, and obtained evidence for subsurface water ice. The mission was named after the movie *2001: A Space Odyssey*. In August 2005 the *Mars Reconnaissance Orbiter* began obtaining new information about the surface, subsurface and atmosphere of Mars, to characterize the planet's climate and geology, to determine if life arose there, and to prepare for eventual exploration of Mars by humans.

The *Pioneer Venus* spacecraft began orbiting Venus in 1978, measuring strong winds in the planet's upper atmosphere and using radar to penetrate the thick atmosphere and map much of the planet's surface with low resolution. Five probes were also launched from *Pioneer Venus*, measuring the properties of the planet's atmosphere as they descended through it.

Five years later, the *Venera 15* and *16* orbiters also used radar to map the surface of Venus, but it wasn't until the 1990s that the *Magellan* orbiter used radar to map the entire planet with a clarity and resolution not available for much of Earth. Since Venus is perpetually shrouded in opaque, sulfurous clouds, this was the only way to establish its global surface terrain. *Magellan*'s radar images have revealed an unearthly world that was resurfaced about 750 million years ago by rivers of outpouring lava, and disclosed numerous volcanoes that now pepper its surface (Fig. 2.9).

The *Galileo* orbiter–probe spacecraft, launched in October 1989, was so massive that no existing rocket had the power to launch it directly to Jupiter, its primary target. Instead, the spacecraft was placed on a looping trajectory that took it past Venus once and Earth twice (Fig. 2.10). The gravity of these planets was used to accelerate and propel the spacecraft in slingshot fashion toward its eventual rendezvous with the giant planet, somewhat like a pitcher winding up to throw a high-velocity strike. While

Table 2.2 Important orbital missions in the solar system

Spacecraft[a]	Launch date	Encounter date	Object	Discovery
Lunar Orbiter 1	10 Aug. 1966	14 Aug. 1966	Moon	Global photographs of lunar surface
Mariner 9	30 May 1971	13 Nov. 1971	Mars	Global image, volcanoes, canyons, outflow channels
Viking 1	20 Aug. 1975	19 June 1976	Mars	Orbiter and lander, surface photographs, life search
Viking 2	9 Sept. 1975	7 Aug. 1976	Mars	Orbiter and lander, surface photographs, life search
Pioneer Venus	20 May 1978	4 Dec. 1978	Venus	Orbiter and multi-probe, global radar images
Magellan	4 May 1989	10 Aug. 1990	Venus	Orbiter, radar maps of surface, volcanic resurfacing
Clementine	25 Jan. 1994	21 Feb. 1994	Moon	Mapped global surface composition and topography of Moon; possible evidence for water ice at lunar poles
Galileo	18 Oct. 1989	7 Dec. 1995	Jupiter	Orbiter and probe, atmosphere and four largest satellites; *Galileo* impacted Jupiter on 21 Sept. 2003
NEAR[g] *Shoemaker*	17 Feb. 1996	14 Feb. 2000	Asteroid	First orbiter of an asteroid, 433 Eros, determining its composition, size, shape, and mass
Mars Global Surveyor	7 Nov. 1996	12 Sept. 1997	Mars	Orbiter, laser altimeter, magnetometer, high-resolution images, mapping began on 4 April 1999, water flow and volcanic activity, ancient magnetism
Lunar Prospector	7 Jan. 1998	15 Jan. 1998	Moon	Maps global elemental abundance, magnetic field and gravity, detection of lunar core, evidence for substantial water ice at lunar poles
2001 Mars Odyssey	7 Apr. 2001	24 Oct. 2001	Mars	Maps the amount and distribution of chemical elements and minerals on the Martian surface, and provides evidence for substantial subsurface water ice; contact lost 14 Nov. 2006
Cassini[c]	15 Oct. 1997	1 July 2004	Saturn	Atmosphere, rings, magnetic environment, methane rain and lakes of methane and ethane on moon Titan, active water jets on moon Enceladus
Mars Reconnissance Orbiter	12 Aug. 2005	1 July 2006	Mars	Information about surface, subsurface and atmosphere of Mars to determine whether life ever arose on Mars, to characterize the climate and geology of Mars, and to prepare for human exploration of Mars
Venus Express[d]	09 Nov. 2005	11 April 2006	Venus	Measurements of atmosphere, clouds, polar vortex

Table 2.2 *(cont.)*

Spacecraft	Launch date	Encounter date	Object	Discovery
Kaguya[e] (Selene)	14 Sept. 2007	3 Oct. 2007	Moon	Lunar topography maps, gravity map of far side of Moon, south pole region, Shackleton crater
Chang'e-1[f]	24 Oct. 2007	5 Nov. 2007	Moon	Composition and images of surface
Chandrayaan-1[g]	22 Oct. 2008	8 Nov. 2008	Moon	Small amounts of water on Moon
Lunar Reconnaissance Orbiter and *LCROSS*	18 June 2009	18 Sept. 2009	Moon	High-resolution mapping of lunar surface in preparation for manned landing; *LCROSS* impacts lunar surface providing evidence for water on Moon

[a] Unless otherwise noted these spacecraft are missions of the United States National Aeronautics and Space Administration (NASA), except the *Cassini* mission, which is a joint ESA and NASA venture.
[b] The acronym *NEAR* stands for Near Earth Asteroid Rendezvous.
[c] The *Cassini–Huygens* mission is a joint ESA and NASA venture, where ESA denotes the European Space Agency.
[d] *Venus Express* is a mission of the European Space Agency.
[e] *Kaguya* is a Japanese mission.
[f] *Chang'e-1* is a mission of the People's Republic of China.
[g] *Chandrayaan-1* is a mission of India's national space agency.

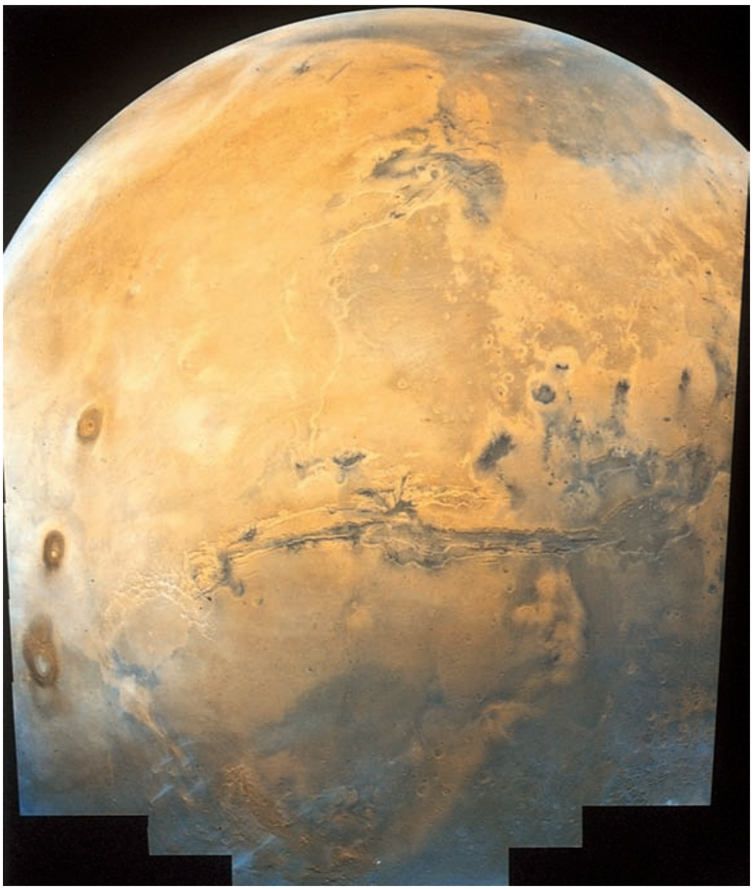


Fig. 2.8 Mosaic of Mars This computer-generated mosaic of *Viking 1* and 2 orbiter images of Mars shows three volcanoes as dark spots to the west (*left*), while the bottom center of the scene shows the entire Valles Marineris canyon system, from Noctis Labyrinthus (*left*) to the chaotic terrain (*right*). Outflow channels are found in the north (*top*), and a variety of clouds and hazes are also visible, especially near the planet's edge. (Courtesy of NASA/USGS.)

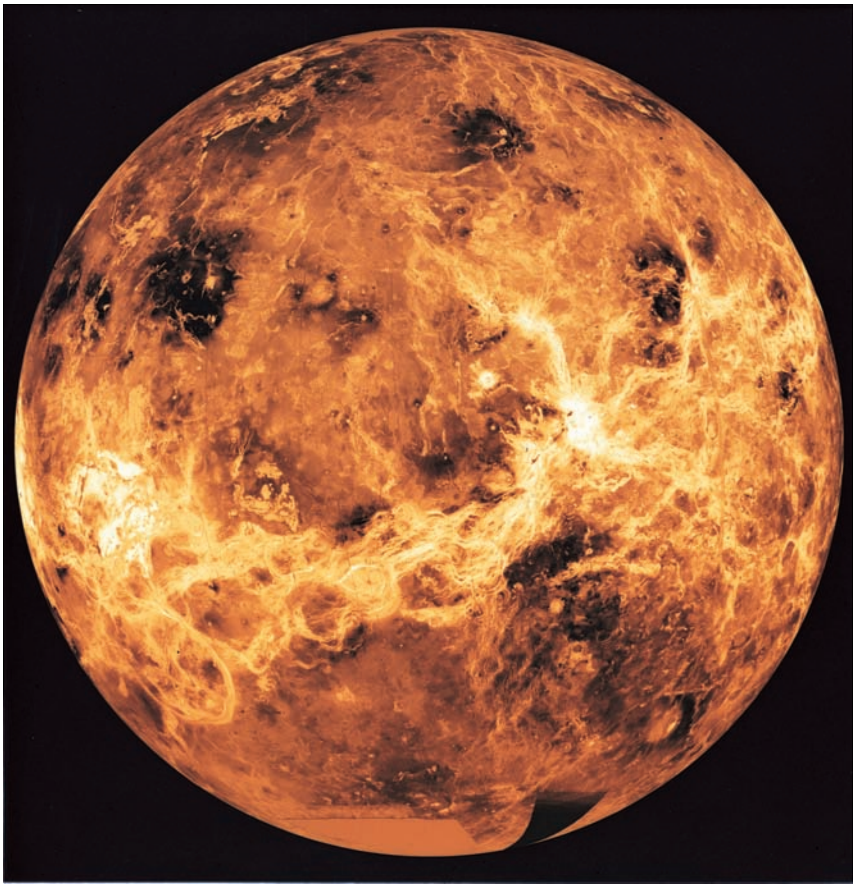


Fig. 2.9 Venus unveiled with radar A cloud-penetrating radar system on board the *Magellan* spacecraft has mapped the global landforms and features on Venus with a resolution of 120 meters, more completely than any planet including Earth. This hemisphere, centered at 180 degrees east longitude, shows the bright, planet-wide, equatorial highlands that contain towering volcanoes, long lava flows and deep faults and fractures. They run from lower left to upper right through Aphrodite Terra (*left of center*), a continent-sized highland, and the bright highland Atla Regio (*just right of center*) to Beta Regio (*far right and north*). Dark areas correspond to terrain that is smooth on the scale of the radar wavelength (0.13 meters); bright areas are rough. The orange tint, based on color images taken by the *Venera 13* and *14* landers, simulates the color of sunlight at ground level after being filtered through the planet's thick atmosphere and clouds. (Courtesy of NASA/JPL.)

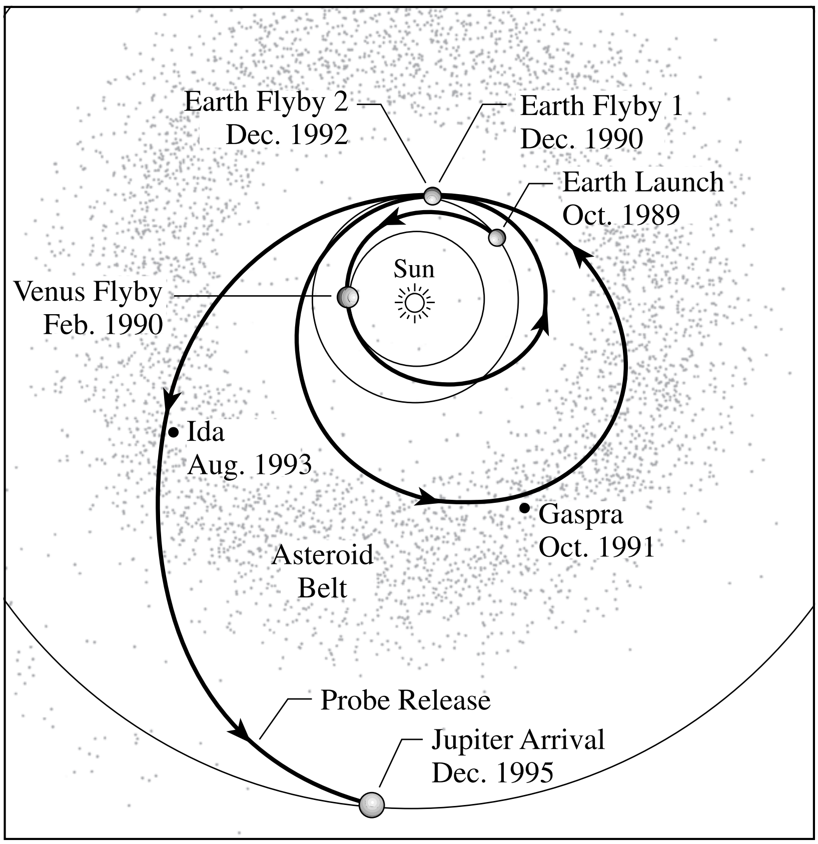


Fig. 2.10 Galileo's long flight to Jupiter After launch in October 1989, the *Galileo* spacecraft used the gravity of the Earth and Venus to accelerate it on to its encounter with Jupiter, six years after launch. In its long, indirect flight path, *Galileo* was able to fly past two asteroids at close range, 951 Gaspra in October 1991 and 243 Ida in August 1993 (see Fig. 2.11). The spacecraft entered into orbit around Jupiter in December 1995, when its descent probe, which had been released five months earlier, dove into the giant planet's atmosphere. The main spacecraft continued to orbit the planet and examine its satellites until 21 September 2003.

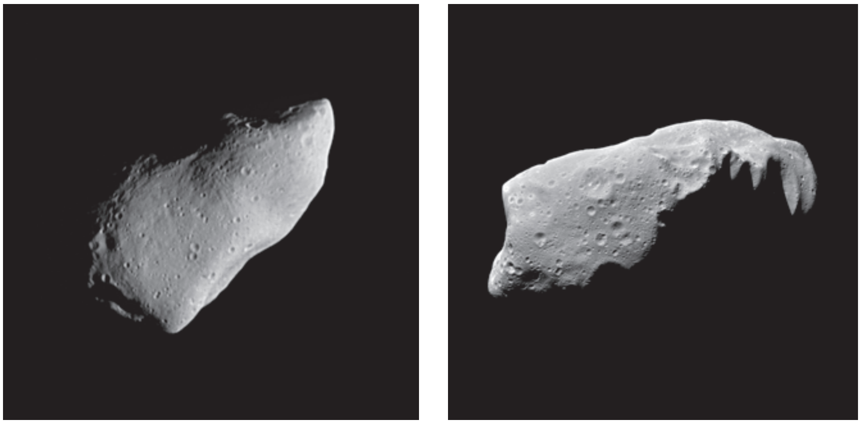


Fig. 2.11 Asteroids close up The Jupiter-bound *Galileo* spacecraft took images of the asteroids 951 Gaspra (*left*) and 243 Ida (*right*) on 29 October 1991 and 28 August 1993, respectively. Both objects have irregular, elongated shapes, suggesting numerous past collisions. Gaspra has dimensions of $19 \times 12 \times 11$ kilometers, and it contains a striking abundance of small craters. Ida is almost three times as long as Gaspra, and a small moon, named Dactyl, accompanies Ida, though the tiny satellite is not visible in this image. Ida is a member of the Koronis family of asteroids, presumed fragments left from the breakup of a larger precursor asteroid in a catastrophic collision. (Courtesy of NASA/JPL.)

the roundabout route took six years, in comparison to the direct 21-month flights of *Pioneer 10* and *11*, it also took *Galileo* on close encounters with two asteroids along the way (Fig. 2.11).

Galileo carried an entry probe that penetrated Jupiter's kaleidoscopic clouds, obtaining the first direct, or in-situ, sampling of a giant planet's atmosphere. The main orbiting spacecraft looped around Jupiter for more than five years, until 2003, obtaining high-resolution images and analysis of the planet's stormy weather, its sparse ring system, and the four large moons. It has provided new insights to these four Galilean satellites (Fig. 2.12), including: volcanic activity on Io; compelling evidence for a global ocean beneath the ice crust of Europa; and the discovery of a global dipolar magnetic field generated within Ganymede, the first such field to be found on a moon.

Seventeen countries and several space agencies, including NASA and ESA, contributed to the *Cassini–Huygens* mission to Saturn and its moons. After a seven-year journey, the spacecraft reached Saturn in July 2004, beginning unparalleled investigations of the planet's atmosphere, rings, satellites and magnetic environment (Fig. 2.13).

Landers and probes are the third category of spacecraft used to investigate the moons and planets (Table 2.3). They have landed on the Moon and Mars, and impacted comet Tempel 1 and the Moon's south polar region, while entry probes have parachuted into the atmospheres of Venus, Jupiter and Saturn's satellite Titan.

The robotic exploration of Mars began with the two *Viking* landers, each a one-ton laboratory that safely landed on the planet's surface in 1976. They obtained beautiful panoramas of the Martian surface and measured the properties of the thin, freezing atmosphere. The *Viking* landers were also sent to search for extant life on Mars, but the results were inconclusive.

On 4 July 1997 the *Mars Pathfinder* landed near the mouth of a canyon system carved by massive floods more than 3.5 billion years ago (Fig. 2.14). Unlike the *Viking* landers, which were shackled to one location, *Pathfinder* contained the small mobile *Sojourner Rover*, which could roam across the surrounding terrain. It was about the size of a small microwave oven and equipped with six-wheel drive. *Sojourner* explored about 250 square meters of the Martian surface, measuring the chemical makeup of the rocks and surface. The distribution of nearby rocks, dust and pebbles were consistent with the downstream deposit of flowing water from an outflow channel long ago (Fig. 2.15).

Sojourner paved the way for the two *Mars Exploration Rovers*, named *Spirit* and *Opportunity*, which landed on opposite sides of Mars in 2003 and continued to explore the surface for more than six years. Each the size of a small dune buggy, they contained cameras and instruments that demonstrated the existence of past flowing water at both locations.

The *Cassini* spacecraft carried the *Huygens Probe* that descended into the hazy, dense atmosphere of Saturn's moon Titan, landing on its surface on 14 January 2005. *Huygens* determined the properties of Titan's Earth-like atmosphere and its mysterious surface below, providing evidence for methane rain and rivers of liquid methane

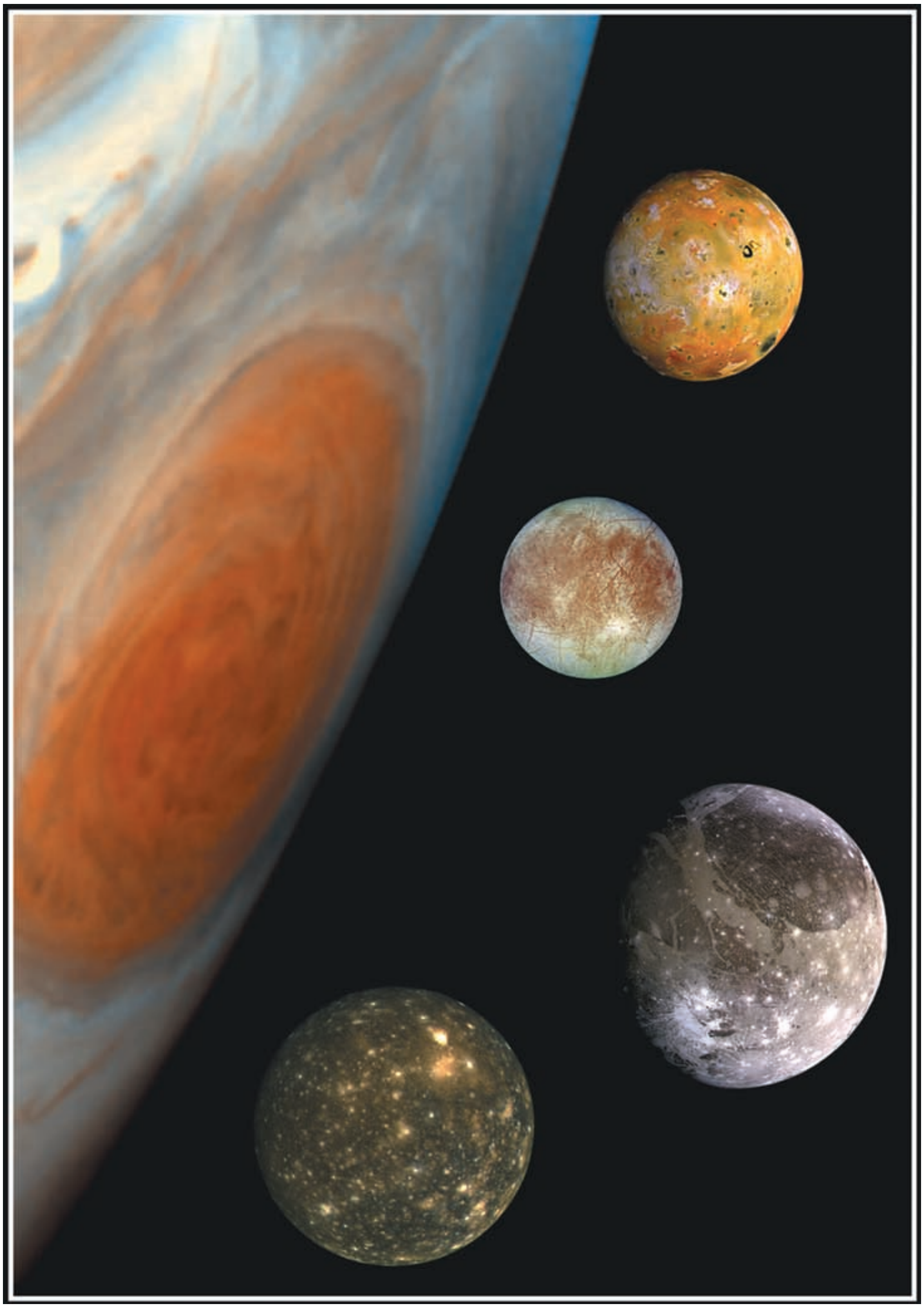


Fig. 2.12 Giant red spot and Galilean satellites The edge of Jupiter with its Great Red Spot and on the same scale the planet's four largest moons, known as the Galilean satellites. From top to bottom the moons are Io, Europa, Ganymede and Callisto. Winds blow counter-clockwise around the Great Red Spot, which has been observed for more than 300 years and is larger than one Earth diameter. Europa is about the size of Earth's Moon, and Ganymede, the largest moon in the solar system, is bigger than the planet Mercury. (Courtesy of NASA/JPL.)

on Titan's surface (Fig. 2.16). Radar from *Cassini* revealed lakes of liquid methane and ethane.

In the early 21st century, we can reflect in amazement at the incredible new worlds that have been discovered by the flybys, orbiters, landers and probes. Future spacecraft are now poised to continue the exploration in greater detail, focusing on issues such as the search for life outside the Earth, and the origin and discovery of planetary systems around the Sun and other stars. Scientists will, for example, ultimately return samples of the surface of Mars for study in our Earth-bound laboratories, to examine them for fossil or recent evidence of life.

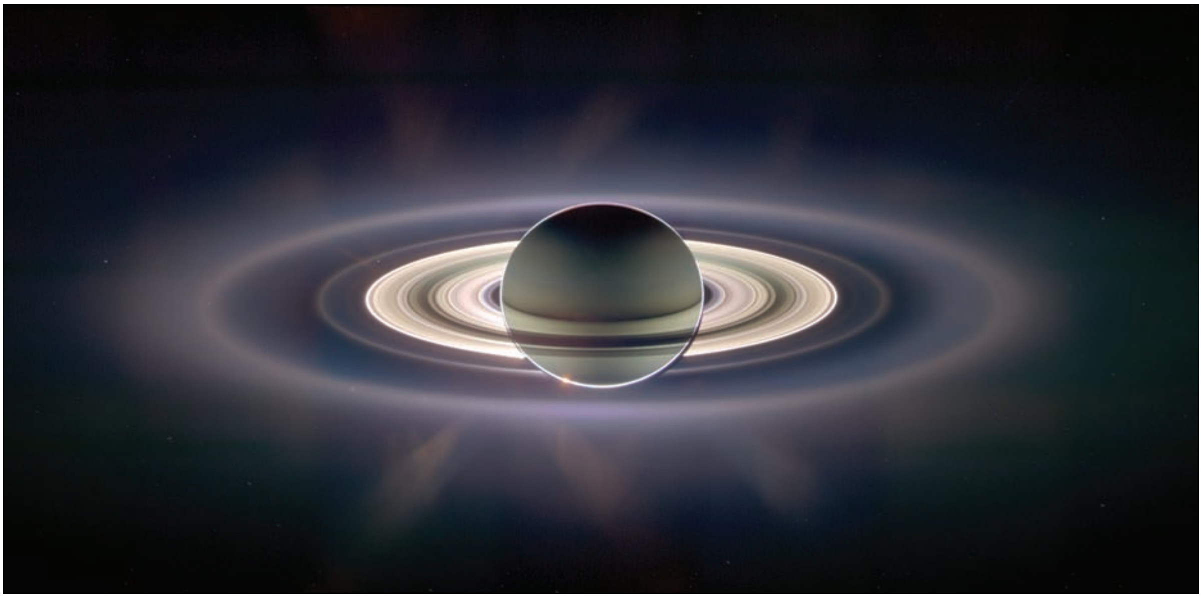


Fig. 2.13 In Saturn's shadow This marvelous panoramic view of Saturn and its rings was created from 165 images taken by a wide-angle camera aboard the *Cassini* spacecraft on 15 September 2006, as it drifted into the darkness of Saturn's shadow and permitted observations of the planet's tiny ring particles. The narrowly confined G-ring resides just outside the bright, inner main rings. The outer, wide E-ring encircles the entire system; icy plumes feed the E-ring from Saturn's satellite Enceladus. The exaggerated color contrast in this mosaic view can be used to infer processes that are sorting the ring particles according to their size. (Courtesy of NASA/JPL/SSI.)

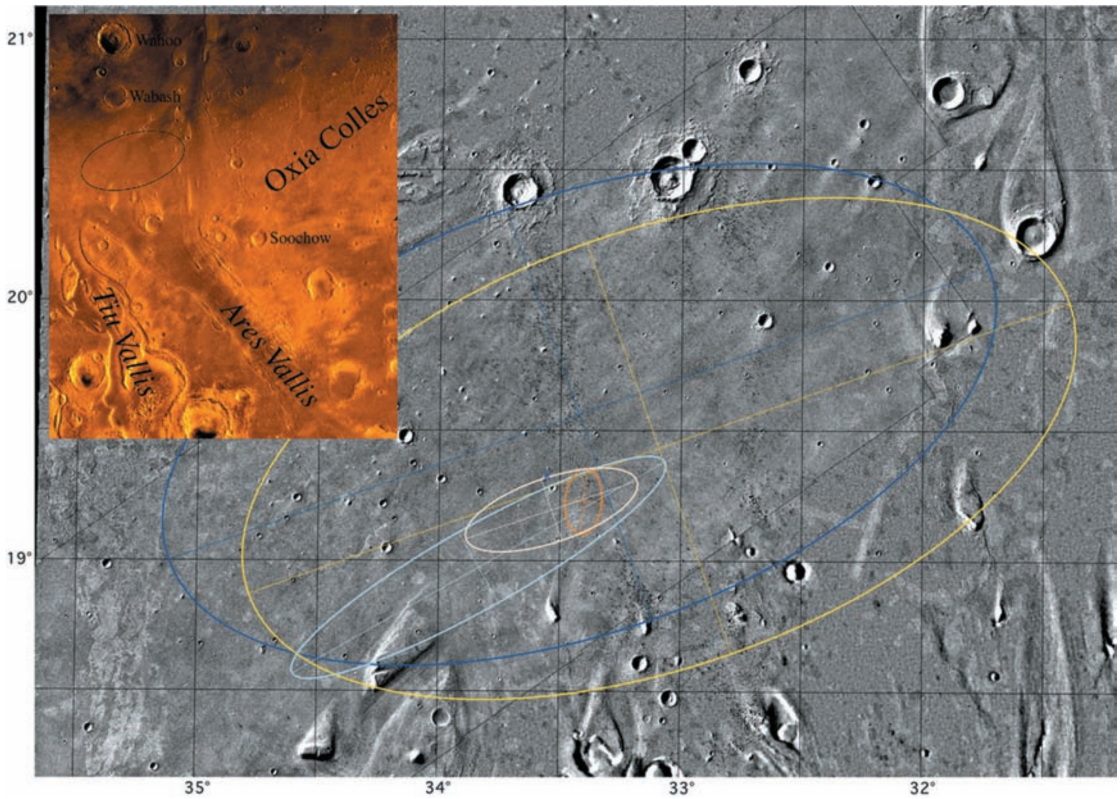


Fig. 2.14 *Mars Pathfinder* lands on an ancient flood plain About 3.7 billion years ago, great floods rushed out of the outflow channel, Ares Vallis, and emptied into the Chryse Planitia, or Plains of Gold, region of Mars (*color inset*). The flowing water carved out streamlined islands around craters (*top right*). This area was chosen as the *Mars Pathfinder* landing site for three reasons: it seemed safe, with no steep slopes or rough surfaces; it had a low elevation, which provided enough air density above the surface for a parachute to work; and it appeared to offer a variety of rock types deposited by the floods. The ellipses mark the area targeted for landing of *Mars Pathfinder*, as refined several times during the final approach to Mars. An X within the smallest ellipse marks the location of the lander at 19.33 degrees north and 33.55 degrees west. The site is about 850 kilometers southeast of the location of *Viking 1* lander. (Courtesy of NASA/JPL.)

Fig. 2.15 Flood debris on Mars A panorama of the surface of Mars taken from the *Mars Pathfinder* spacecraft soon after its landing on 4 July 1997. The scene is littered with boulders and rocks, the debris of catastrophic floods early in the planet's history. The flood residue is found between a few meters away from the lander to the two modest hills, the Twin Peaks, which are about 30 meters tall and located at a distance of about one kilometer. Between and partially covering the rocks is rust-red, iron-oxide dust, the result of chemical weathering of exposed rock surfaces here and elsewhere. (Courtesy of NASA/JPL.)

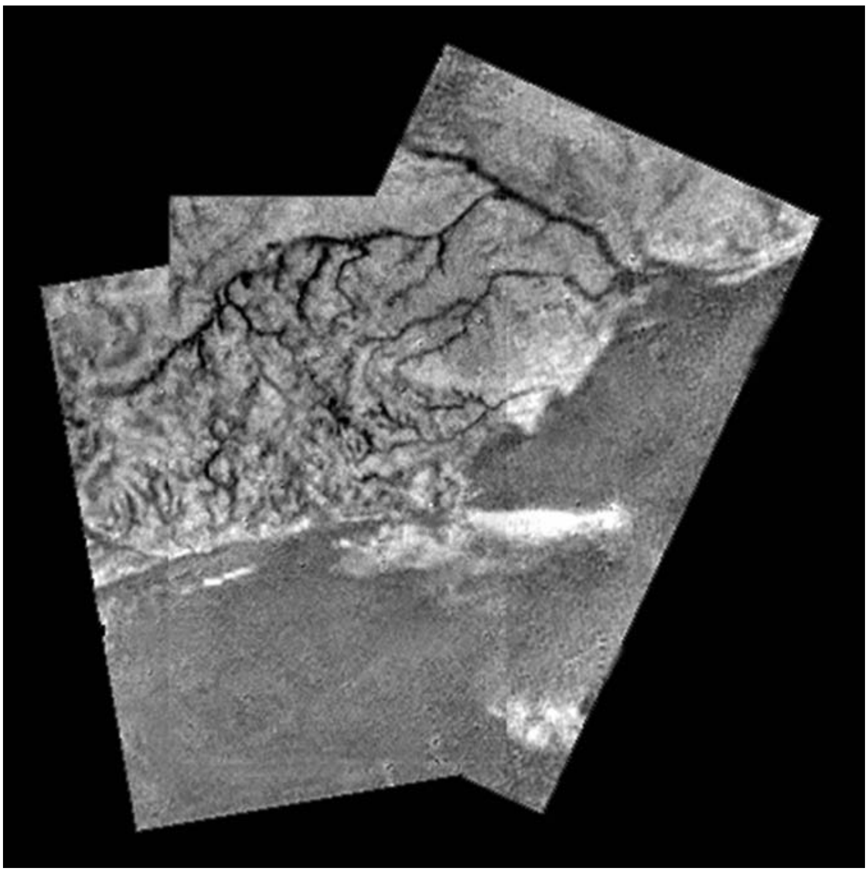


Fig. 2.16 Methane flows on Titan This image was taken form the *Huygens Probe* just before it landed on the surface of Saturn's large satellite Titan on 14 January 2005. It shows flows down a high ridge into a major river channel from different sources. The feature has been attributed to liquid methane fed by the fall of methane rain. (Courtesy of NASA/JPL/ESA/U. Arizona.)

Table 2.3 Landers and probes

Spacecraft[a]	Launch date	Encounter date	Object	Discovery
Luna 9[b]	31 Jan. 1966	3 Feb. 1966	Moon	Soft landing, Oceanus Procellarum
Surveyor 1	30 May 1966	2 June 1966	Moon	Soft landing near Flamsteed
Apollo 11	16 July 1969	20 July 1969	Moon	First humans on Moon, Mare Tranquillitatis, sample return
Venera 7[b]	17 Aug. 1970	15 Dec. 1970	Venus	High surface temperature and pressure
Venera 9[b]	8 June 1975	22 Oct. 1975	Venus	Surface photograph, volcanic rocks
Viking 1	20 Aug. 1975	20 July 1976	Mars	Chryse Planitia, negative life search, monitor environment for more than one Martian year
Viking 2	9 Sept. 1975	3 Sept. 1976	Mars	Utopia Planitia, negative live search, monitor environment for more than one Martian year
Galileo	18 Oct. 1989	7 Dec. 1995	Jupiter	Atmosphere winds, ingredients
Mars Pathfinder	4 Dec. 1996	4 July 1997	Mars	Chryse Planitia, surface rover *Sojourner* examines dust, pebbles and rocks near mouth of outflow channel
Cassini–Huygens[c]	15 Oct. 1997	14 Jan. 2005	Titan	Huygens probe, methane rain, vast lakes of liquid methane
Spirit	10 June 2003	3 Jan. 2004	Mars	Gusev crater, rover, past water flow
Opportunity	7 July 2003	14 Jan. 2004	Mars	Meridiani Planum, rover, past water flow
Deep Impact	12 Jan. 2005	4 Jul. 2005	Comet	Water ice on comet Tempel 1
Phoenix	3 Aug. 2007	25 May 2008	Mars	Water ice in Martian subsurface at north polar plains of Mars
LCROSS[d]	18 June 2009	9 Oct. 2009	Moon	Evidence for water ice by impact in permanently shadowed lunar crater, south polar region
Rosetta	2 Mar. 2004	Nov. 2014	Comet	Landing on Comet 67/Churyumov-Gerasimenko. The spacecraft passed by two asteroids, 2867 Steins in Sept. 2008 and 21 Lutetia in July 2010.

[a] Unless otherwise noted the spacecraft are missions of the United States National Aeronautics and Space Administration, abbreviated NASA.

[b] Spacecraft launched by the USSR.

[c] The *Cassini–Huygens* mission is a joint ESA and NASA venture, where ESA denotes the European Space Agency.

[d] The acronym LCROSS designates *Lunar CRater Observatory and Sensing Satellite*.

The space-age investigations of the moons and planets have shown us that each moon or planet is unique, the result of different combinations of physical, chemical and dynamical processes that have formed and shaped it. Yet there are fascinating similarities between the major planets and some moons, despite the individual differences that make each of them stand apart. They all exhibit common properties and similar processes, such as impact craters, volcanoes, water and atmospheres, reminding us of the basic elements in ancient Greek philosophy — earth, fire, water and air.

2.2 Impact craters

Ubiquitous craters, ancient records of formation

Impacts have played an important role in the early history of the planets and their subsequent evolution. The collisions of small bodies to create larger bodies resulted in the formation of the planets and their satellites, which gradually cleaned out interplanetary space. Even about 4.0 billion years ago, the impact rate was still sufficient to

Fig. 2.17 Craters on Jupiter's moon Callisto The ancient surface of Callisto shows one of the highest densities of impact craters in the solar system. The satellite's icy surface is as rigid as steel, permitting it to record the bright scars of a heavy bombardment by meteorites roughly 4.0 billion years ago. As shown in this image, taken from the *Galileo* spacecraft in May 2001, Callisto's surface seems uniformly cratered, but it is not uniform in color or brightness. Scientists believe the brighter areas are mainly water ice, while the darker areas are highly eroded, ice-poor material. (Courtesy of NASA/JPL/DLR.)

produce large impact basins with diameters measured in hundreds to thousands of kilometers. Impacts by smaller bodies have always been more frequent, since there are more small objects in space than large ones, and they also contributed to the creation of the ancient, heavily cratered terrain on the moons and planets.

The similarity of the highland crusts of the Moon, Mars and Mercury, despite their differing masses and locations, indicates that impacting objects were spread throughout the inner solar system during its early days. The ubiquitous craters found on the icy satellites of Jupiter, such as Callisto (Fig. 2.17), suggests that they were also subject to a heavy bombardment in their early formative stages. All the planets and satellites most likely experienced the most intense hail of impacting projectiles at about the same time as the Earth's Moon, for scientists think that the entire solar system, with its Sun, planets and their satellites, gathered together beginning 4.6 billion years ago.

Using impact craters to estimate surface age

Radioactive dating of the rocks that *Apollo* astronauts have brought back from the Earth's Moon has permitted a reconstruction of the rate of impact during most of the Moon's history. Some of these rocks are the oldest ones ever found, and they show that the Moon accumulated by the aggregation of rocky projectiles about 4.6 billion years ago. When the Moon was very young, its outer layers were probably molten, but the crust cooled and became solid. The battered lunar surface that we see today remains a museum of impact scars that were created back then. It records an intense, heavy bombardment of leftover formation material that created the large impact basins and most of the lunar craters roughly 4.0 billion years ago.

When the measured ages of lunar rocks are combined with crater counts, scientists obtain a record of the cratering rate on the Moon and its variation with time (Fig. 2.18). The earliest pace of bombardment declined very rapidly for the first 800 million years following the start of planetary formation 4.6 billion years ago. The late heavy bombardment, which ended about 3.9 billion years ago on the Moon, is thought to account for the most heavily cratered regions on Mercury and Mars, as well as the lunar highlands, all saturated with craters upon craters.

Gradually the hail of impacting meteorites decreased, as most of the interplanetary meteoroids were swept up and pulled in, and the rate of cratering slowed during

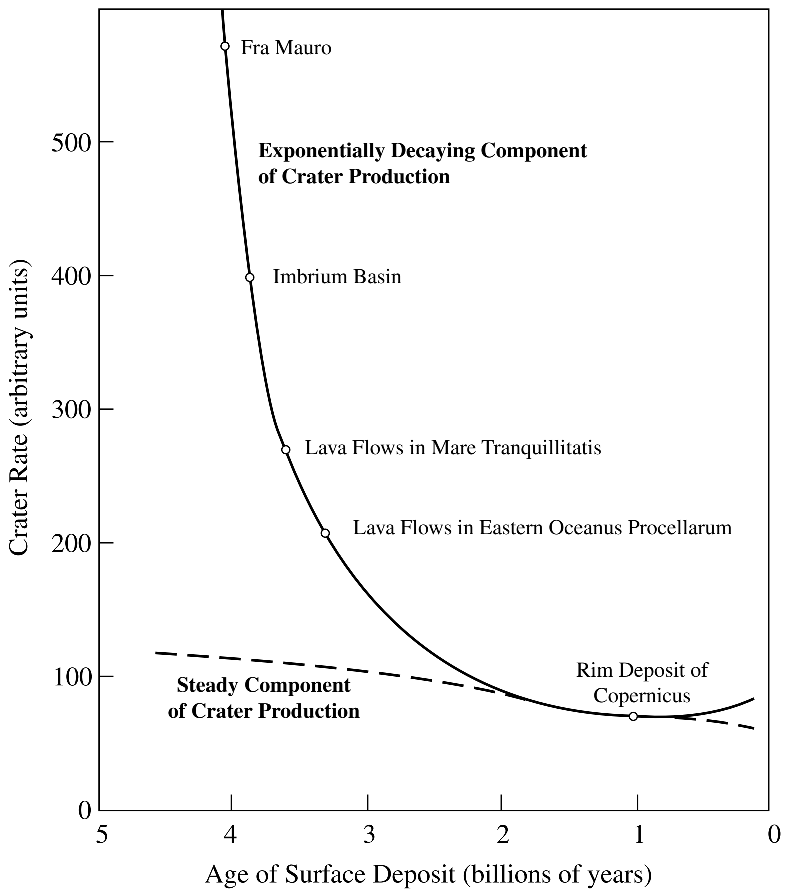


Fig. 2.18 Varying crater rate on the Earth's Moon The rate of forming craters on the lunar surface is plotted against time. The circles denote the crater rate and rock ages at various *Apollo* landing sites. The crater rate was very high during an intense bombardment that occurred 3.9 billion years ago. The rate dropped rapidly during the subsequent billion years, giving way to the lower steady rate of crater production that has persisted for the last 3 billion years. With such a curve, we can obtain approximate surface ages just by counting the number of craters in different parts of the Moon. Estimates of the ages of planetary surfaces can similarly be obtained from the density of craters on them.

the subsequent billion years. A much lower, steady rate of crater production has persisted for the last 3.0 billion years, so relatively young craters are hard to find on the Moon. White rays that splash across the lunar surface distinguish young craters. The rays of older craters are darkened by eons of continued impact by small meteorites.

Similar declining rates of impact crater formation are thought to have occurred in the final stages of planetary and satellite formation throughout the solar system. Detailed counts of the numbers of craters at different part of the ancient surfaces of these bodies can be used to estimate their ages to within a few million years, particularly in locations on Mars or Mercury that date back to the tail end of the heavy bombardment.

In contrast, the surfaces of the Earth, Venus and Jupiter's satellite Io have relatively few craters, for they have been erased by subsequent geological or volcanic activity. Erosion by wind and water, the deposition of sediments, the collisions of continents, and the internal churning of its rocks have erased any record of an ancient intense bombardment of the Earth. Massive volcanic outpourings covered the surface of Venus with lava about 750 million years ago, and Io is so volcanically active that its surface is continually renewed and contains no impact craters whatsoever.

How to make a crater

Comparatively recent craters still exhibit the details of the impact that created them (Fig. 2.19); older craters have been worn away by small particles that continuously bombard the moons and planets.

There are several lines of evidence that the lunar craters were formed by the explosive impact of interplanetary projectiles:

(i) The amount of material piled on a crater's raised rim is nearly equal to the material excavated from the interior, so if the rim was pushed back into the crater its depressed floor would rise to the level of the neighboring surface.

(ii) Nearly all of the Moon's craters are round. The explosive force of a large impacting object, or meteorite, will produce round craters despite the fact that the projectiles that produced them must have arrived in a variety of directions – some nearly vertically, others at a glancing angle.

(iii) The rocks returned from the heavily cratered regions on the Moon consist of fragments of pre-existing rocks that have been welded together by the enormous pressures of impact.

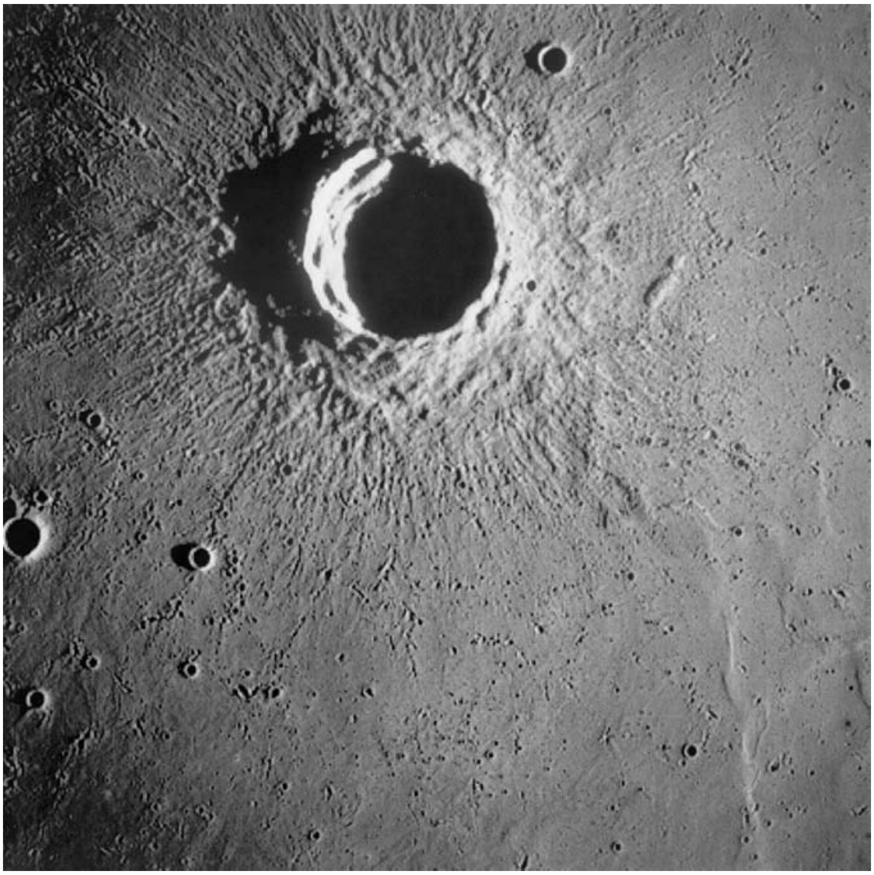


Fig. 2.19 Lunar crater Timocharis
Astronauts on board the *Apollo 15* mission took this image of the medium-sized crater Timocharis, about 34 kilometers across, in August 1971. The deposits and ejected material have been thrown outward in the radial direction by the meteorite impact that created the primary crater with its circular rim. Smaller secondary craters are located outside this material *(lower left)*. (Courtesy of NASA.)

So, solid, rocky objects, named meteoroids, which came from interplanetary space and hit the Moon, must have created the lunar craters, as well as those found on the terrestrial planets and some other moons. Rocky meteoroids were strewn throughout the solar system in its youth, orbiting the Sun for millions and even billions of years until they happened to collide with a moon or planet. This leftover residue from planet formation was more abundant in the early history of the solar system, for it was gradually swept up through collisions with larger bodies.

When the meteoroids strike the surface of a planet or satellite they are called *meteorites*. Meteorites of all sizes have hit the Moon, and its crust records the impact of more small meteorites than large ones.

Although the projectile vaporizes on impact, the explosion excavates material and hurls the pulverized debris outward, creating a raised circular rim 10 to 20 times as wide as the impacting meteorite (Fig. 2.20). Many of the ejected rocks were large enough to create their own craters in turn; they are known as secondary craters. The depth of the crater hole is about one-tenth of its diameter, a relationship that holds for simple craters on the Moon, Mars, Mercury, Earth and Callisto.

The interplanetary bodies, the meteoroids, travel at rapid velocities of tens of kilometers per second, and carry a high kinetic energy. When they impact the surface of a planet or satellite, about half of this kinetic energy is transferred to the target when it stops the moving object. Its energy of motion is suddenly transformed to shocks which engulf and vaporize the impacting object and heat the surface. The high temperature and pressure melt material at the point of impact, and push it downward, while the shocks excavate the crater cavity and throw up a rim of pulverized and melted rock around it. This material is carried out in all directions from the point of impact, creating circular craters and radial ejecta.

The rim is tossed out almost nonchalantly, something like flicking a particle off the end of a whip, but the excavated material is still many times as massive as the impacting projectile and about 10 000 times the volume. This is essentially due to the energy released by the impacting object, which can be equivalent to the explosion of tens of thousands of hydrogen bombs. Violent rebound of the crater floor, from the greater energy and shock of larger meteorites, gives rise to a central peak or peaks. In addition, many of the larger craters have terraced walls caused by rim material slumping in toward the crater center.

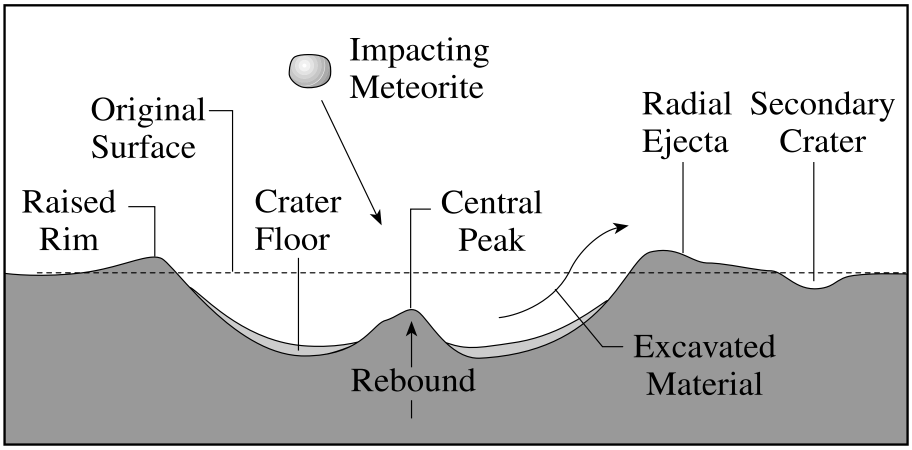


Fig. 2.20 Cross-sectional anatomy of a crater An impacting meteorite excavates a circular crater that is about 20 times the diameter of the meteorite. The depth of the crater is roughly one-tenth its diameter, and the crater floor is depressed below the surrounding terrain. The explosion gouges out a circular hole, depositing material around its rim and ejecting debris outward in the radial direction. The surface rebounds from the impacting force of a large meteorite, creating a central peak in the floor of the biggest craters.

Giant impacts

The most distinctive features on the Moon are the circular craters that closely pepper its surface, and the largest lunar craters are the impact basins. A typical one is the Imbrium Basin, with a diameter of 1500 kilometers. Its outline can be seen with the unaided eye, forming an "eye socket" of the face of the "Man on the Moon". Its outer rim is defined by prominent mountain ranges, such as the Apennine Mountains (Fig. 2.21). Such basins were created early in the Moon's history, and they were subsequently flooded and nearly filled with dark molten lava from the interior.

Ejecta from the Imbrium Basin gouged out radial ridges and valleys that went a quarter of the way around the Moon, scattering a thick blanket of debris over most of the near side of the Moon. The energy of impact was so great that the floor of the crater rebounded, surging up and down and creating multiple, concentric, ring-like rims as the lunar surface vibrated like the head of a drum.

As on the Moon, there are numerous small bowl-shaped craters on Mercury, smaller than 100 meters in diameter, and large impact craters up to a thousand kilometers across. Both worlds also contain a few young craters with bright rays as well as many older craters without rays, and both the Moon and Mercury have no atmosphere or weather to erode their surface.

Mercury's surface also contains multi-ringed impact basins, such as the Caloris Basin (Fig. 2.22). It has been named *Caloris*, the Latin name for "heat", because it is located at a place on Mercury that faces the Sun when the planet is at the point in its orbit that is closest to the Sun. The rim of mountains that marks the outer boundary of the Caloris Basin is about 1550 kilometers in diameter.

The cataclysmic impact that created the Caloris Basin occurred an estimated 3.85 billion years ago when a meteorite roughly 150 kilometers across hit Mercury, like a cosmic bomb with an energy of a trillion 1-megaton hydrogen bombs. The violent explosion reverberated through the young planet, sending strong seismic waves along the surface and through the deep interior (Fig. 2.23). These waves converged to a focus on the side of Mercury opposite to the Caloris Basin, producing a huge region of cracks, faults, hills and valleys.

The Earth's Moon probably formed from the remains of a giant impact with our planet. A Mars-sized object apparently struck the young Earth, melting the surface material at the point of impact and sending debris into orbit that eventually congealed to become the Moon.

A similar massive impact may have sent Venus spinning in the opposite direction to that of the rotation and orbital motion of all the other major planets. It has also been proposed that a giant impact during the early stages in the history of Mars gouged out most of the Martian crust in the planet's northern hemisphere, resulting in a global crustal dichotomy between the low-lying north and the southern highlands. An original low-density mantle of young Mercury might have been similarly blasted off, leaving its dense iron core behind. But all of these ideas are speculations, hypotheses that have not been fully confirmed.

Unique craters on Mars and Venus

The shape of virtually all impact craters is a circular depression with an upraised rim. The details differ according to

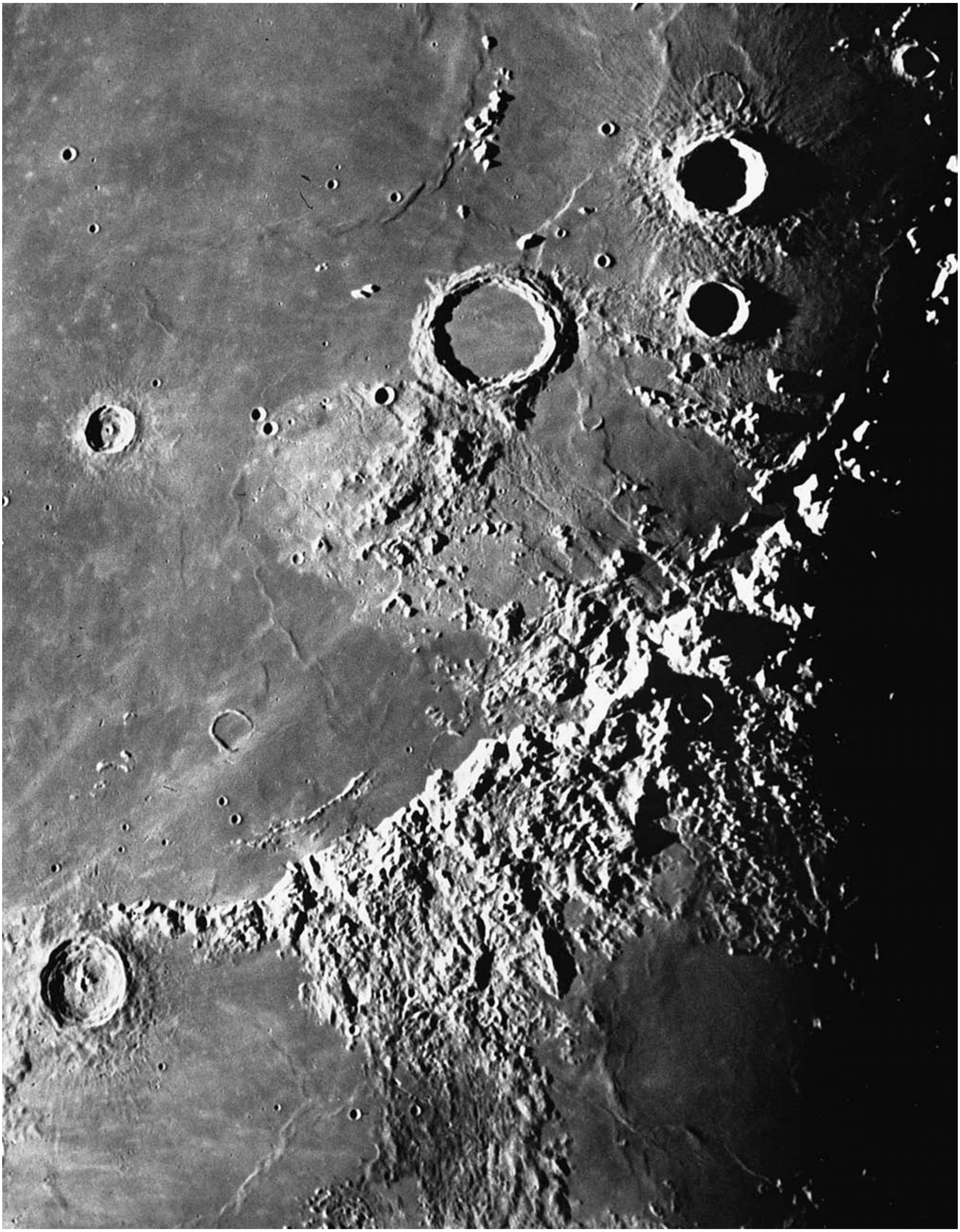


Fig. 2.21 The Moon's Apennine mountains The radial structure and steep inner slopes of these mountains (*lower right*) mark a section of the outer rim of the Imbrium Basin. The huge excavation was subsequently filled with lava to form the smooth Mare Imbrium and partially submerge the inner ring of mountains (*upper left*). The smaller circular craters include Timocharis, which is also shown in Fig. 2.19 and is about 34 kilometers in diameter (*center left*) and the largest round structure Archimedes (*upper center*) with a diameter of 83 kilometers. (Photo courtesy of UCO/Lick Observatory.)

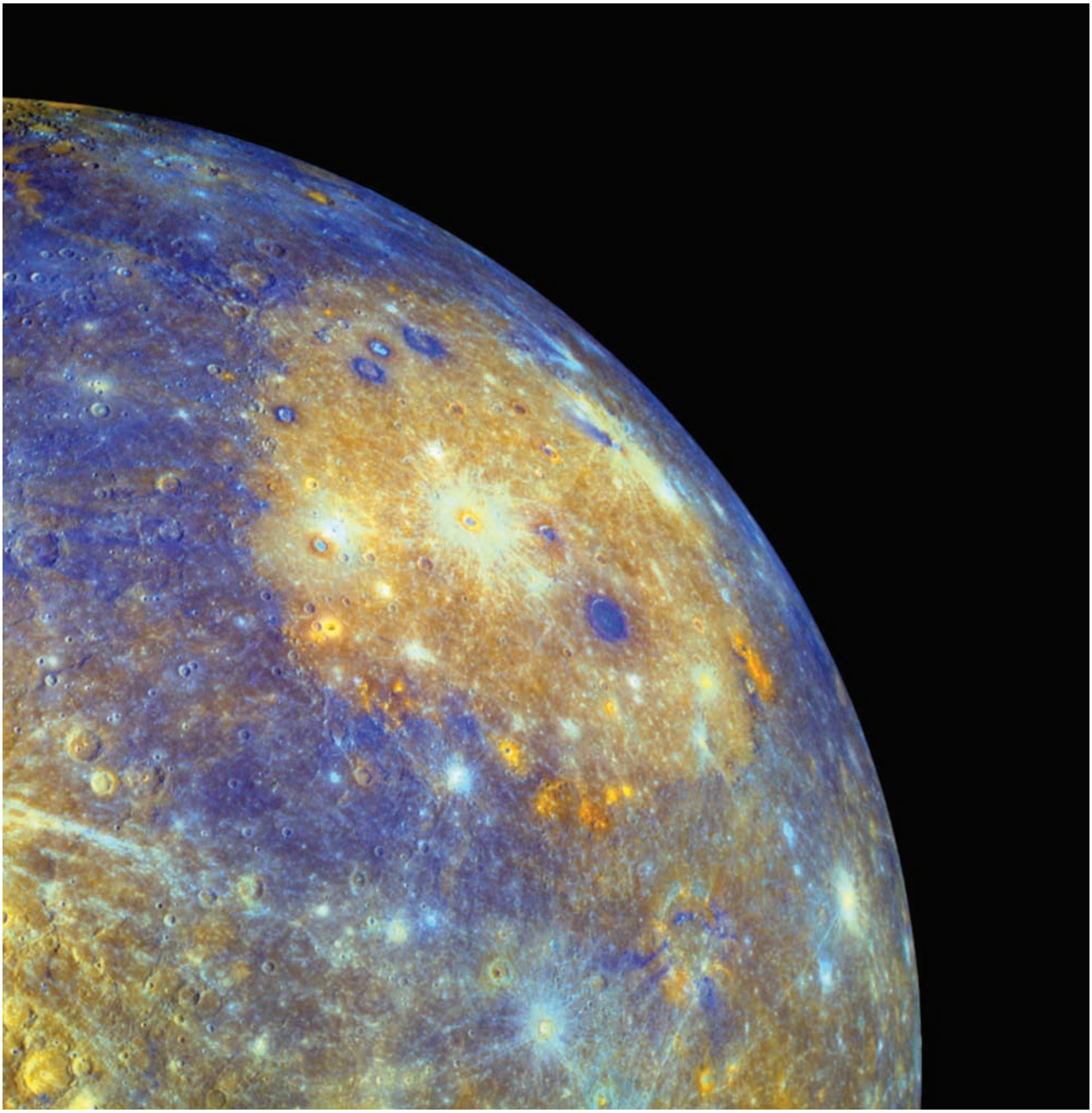


Fig. 2.22 Mercury's Caloris impact basin This false-color image shows Mercury's great Caloris impact basin as a large, circular orange feature in the center of the image. It was acquired on 14 January 2008 from the *MESSENGER* spacecraft. The smaller, bright orange spots just inside the rim of Caloris basin are thought to mark the location of volcanic features. The color variations in the surrounding plains indicate Mercury's variable surface composition. (Courtesy of NASA/JHUAPL/ASU/CIW/*Science*/AAAS.)

the varying size of the impacting object, from small, simple bowl-shaped holes to larger complex craters with central peaks and internally terraced rims and the largest multi-ring impact basins. But given these variations, the craters that have been excavated from the Moon and terrestrial planets are quite similar.

In contrast, there are noticeable variations in the patterns of ejected material. The round, fresh craters on the Moon are surrounded by secondary craters and bright rays thrown out in ballistic trajectories and undisturbed by any atmosphere. Many craters on Mars display mud-like ejecta, most likely associated with impact melting of subsurface ice. Asymmetric ejecta that surround craters on Venus are attributed to the oblique impacts of meteoroids interacting with the planet's thick atmosphere.

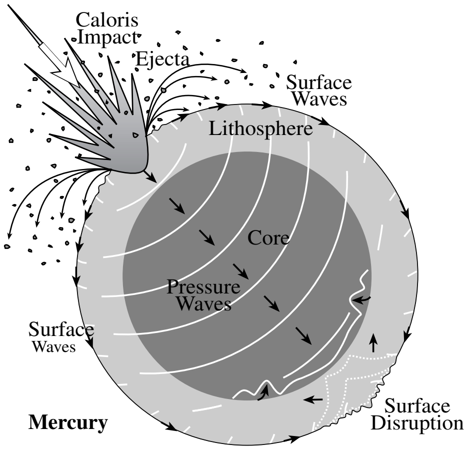


Fig. 2.23 Explosive impact on Mercury When an exceptionally large meteorite hit Mercury an estimated 3.85 billion years ago, it sent intense waves around the planet and through its core. They came to a focus on the opposite side of Mercury, disrupting the surface and producing hilly and lineated terrain there. The Caloris Basin was excavated at the impact site (also see Fig. 2.22), and it now exhibits concentric waves that froze in place after the impact.

Fig. 2.24 Material ejected from impacts on Mars Some Martian craters are surrounded by discrete lobes of fresh-appearing flows, each surrounded by a low ridge or rampart and sometimes layered. The unique pattern of ejected material can be attributed to melting of water ice by the heat of impact. An instrument aboard the *2001 Mars Odyssey* orbiter took this image. (Courtesy NASA/JPL/ASU.)

Upon heating of ground ice by impact on Mars, liquid water is most likely incorporated into the ejecta, lubricating the material that flows along the ground after ejection (Fig. 2.24). As a result, some craters on Mars resemble those produced by impacts into mud.

The largest impact craters on Mars are also shallower than their lunar counterparts, with more subdued rims and flatter floors, and there are fewer smaller craters on Mars than there are on the Moon. These differences might be explained by enhanced erosion that modified the worn, old-looking craters and wiped out many of the existing small craters during the planet's early history, when the majority of craters were still forming.

Following impact, large objects left craters on Venus that at first sight resemble those on the Moon, with central peaks, flat floors and distinct circular rims. But the dense atmosphere on Venus affected both the incoming projectile and its ejected debris, creating features that are unlike any other craters in the solar system. The bright apron of debris that surrounds large craters on Venus often has a lobate, petal-like appearance with an unexpected asymmetry (Fig. 2.25). Material that was ejected from the crater became entrained in the hot, thick atmosphere, transforming it into a turbulent, fluid-like substance. The material flowed and spread out from the crater, creating patterns that resemble flowers or butterflies, rather than hurtling away from it to great distances.

Moreover, when the impact on Venus was oblique, the atmospheric wake of the incoming object prevented the ejecta from scattering back in the direction from which the impacting meteoroid came, so ejecta are missing in this region. Small incoming projectiles never made it to the ground, for they were burned up in the thick atmosphere. There are consequently no very small craters on Venus.

Large impact craters on Venus are relatively scarce when compared with the closely spaced, overlapping lunar craters. At one time Venus was probably as heavily cratered as the Earth's Moon, but the relatively small number and wide spacing of the craters now on Venus indicate that the surface we now see is much younger. When the Moon's cratering rate is scaled to Venus, the relative paucity of craters on Venus's surface indicates an average surface age of about 750 million years, but the planet originated about 4.6 billion years ago. The relatively few craters we now see are due to meteorite impact since rivers of outpouring lava resurfaced the entire planet about 750 million years ago.

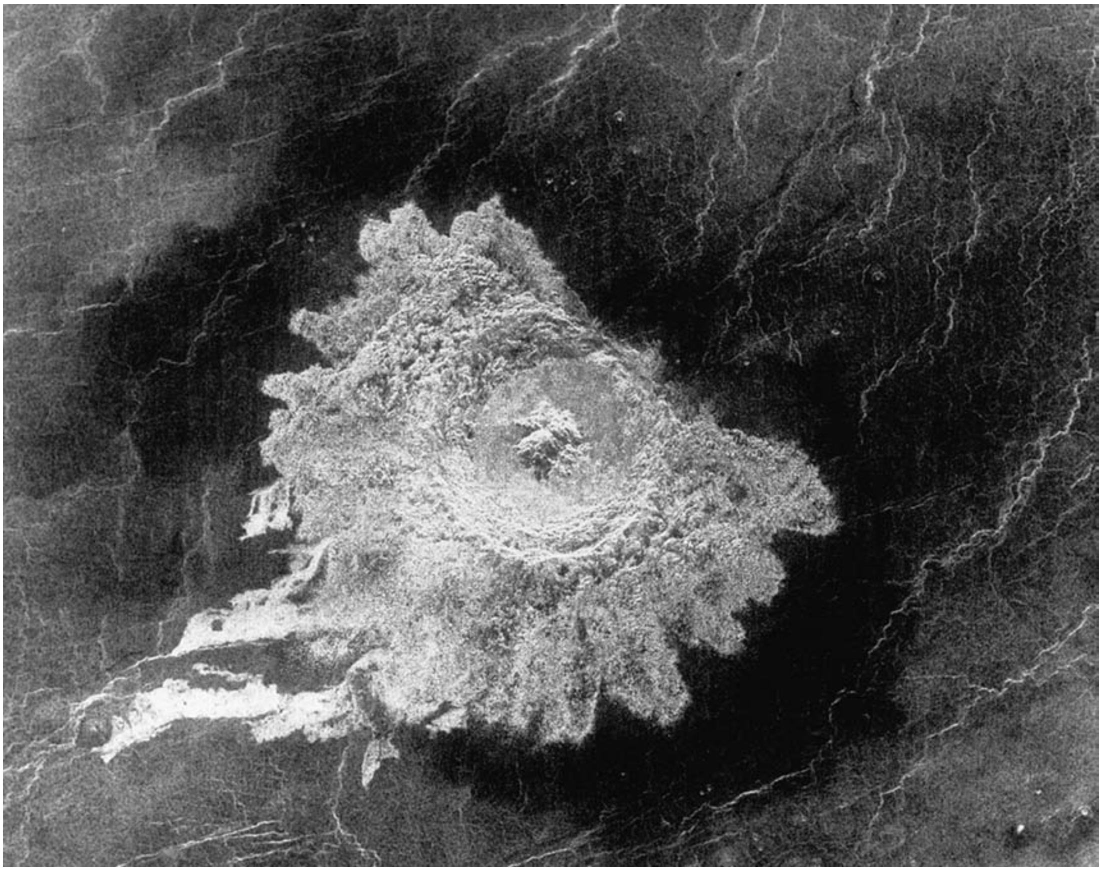


Fig. 2.25 Aurelia impact crater on Venus The unusual crater shapes on Venus are illustrated in this *Magellan* radar image of the Aurelia crater. Like the large impact craters on the Moon, it contains a circular rim, terraced walls and a central peak. But unlike lunar craters, flows emanate from the crater, and a sector of the flow is missing, apparently due to an oblique impact from the upper-right. Interaction with the dense, thick atmosphere on Venus caused the ejected debris to act like a fluid, producing the lacy, rounded lobes. Crater Aurelia, which is 32 kilometers in diameter, has been named in honor of the mother of Julius Caesar; apparently, Aurelia is also the name of Arnold Schwarzenegger's mother. (Courtesy of NASA/JPL.)

2.3 Volcanism

Trapped heat

Volcanoes, another common aspect of the solar system, are driven by internal heat. For a large rocky planet, internal heat is continuously generated by the slow decay of radioactive material. Because of their large size and rocky composition, both Earth and Venus have internal heat powered by radioactive decay. Satellites can be heated by varying gravitational interaction with their planet; Io and Europa are both heated inside through tidal flexing by Jupiter's immense gravitational forces. Heat was also provided when the planets and satellites originated, as the result of high-speed collisions between smaller bodies, creating an ocean of melted rock on the newly formed surfaces of the Earth's Moon and the terrestrial planets.

Volcanic activity has transformed about three-quarters of the surface of the Earth and Venus, extensive parts of the surfaces of Mercury, Mars and the Earth's own Moon, and all of the surface of Jupiter's satellite Io. The volcanism occurs when heat produced in the interior of a planet or satellite rises to the surface, either as tall volcanoes or flatter, surface flows. Internal heat also generates liquid water within the large ice-covered satellites of the giant planets; the water works its way out to erupt as volcanoes of ice.

The underground molten rock, known as magma, is trapped inside a terrestrial planet and the pent-up heat wants to rise up and get out. The liquid magma is swollen by the heat, becoming lower in density than solid rock, and tends to rise through cooler, higher-density material. The surrounding solid rock can also squeeze the molten rock within it, helping to drive the magma upward under pressure. Gases locked within the molten rock expand as it rises, providing another thrust to the upward rise.

The interior of a satellite or planet is not full of molten rock. Within the Earth, for example, the magma is bottled up in chambers that are surrounded by solid rock, and the partially molten material is localized in pockets below the solid mantle and crust. Eventually some of the hot magma rises beneath the surface, pushes up, and sometimes melts and punches a hole in the crust, like a welder's torch. The lava then erupts as a volcano in a relatively small area of the surface. In contrast to the outer parts of a planet's interior, only the internal core is fully molten, but it lies so far down that the core material never makes its way to the surface.

Shield volcanoes and volcanic plains

The two most common types of volcanism are known as shield volcanoes and volcanic plains, and they are both found on Earth, Mars, Venus and Io.

The lava emitted in shield volcanoes solidifies relatively quickly and builds up tall features with gentle slopes after hundreds and even thousands of eruptions. Collapsed depressions called calderas are found at or near the summits of many shield volcanoes, arising when the source of magma drains back into its underground chamber, awaiting the next eruption.

Smooth volcanic plains are created when flowing lava spreads out along the surface, instead of rising far upward, and flattens out before solidifying.

Both the shield volcanoes and lava plains are made of a mixture of minerals called basalt, which arise from metal-rich and silica- and volatile-poor magmas.

Earth's unique volcanoes

Long lines of volcanoes are found under the Earth's oceans, forming volcanic ridges near the centers of the water-filled basins. The underwater volcanoes are fed by magma rising from partially melted mantle rock at the tops of internal convection cells.

Basalt lava has been flowing out from both sides of the mid-ocean ridges for the past 300 million years, producing spreading sea-floors and helping to propel drifting continents across the globe. They are related to the theory of plate tectonics discussed in greater detail in Chapter 4.

Terrestrial volcanoes also rise up from "hot spots", giving rise to volcanism that is not related to the mid-ocean ridges. The Hawaiian Islands are an example. The islands form a chain of shield volcanoes created as a moving plate crosses an underlying hot spot. Mauna Kea and Mauna Loa, on the big island of Hawaii, together form a mountain of basaltic lava that is much broader than it is tall and has gentle slopes; it is more than 120 kilometers across at its base and rises 9 kilometers above the ocean floor. Mauna Loa is still erupting and growing, with repeated surges of lava that flow down its flanks. The smaller Hawaiian Islands mark out a string of extinct volcanoes (Fig. 2.26).

When a moving plate bends downward into the Earth's interior, and is at least partly re-melted, it produces another type of unique terrestrial volcano whose magma is rich in silicas and volatiles like water and carbon dioxide. These volcanoes often erupt explosively and they are found in a ring of fire that encircles the Pacific Ocean.

Cone-shaped terrestrial volcanoes with steep slopes are formed when the lava is propelled out by hot gas. Examples are Vesuvius in Italy and Mount Fuji in Japan. The eruptions that formed these steep-walled mountains often expelled large clouds of volcanic ash.

Terrestrial volcanoes can have direct, unanticipated effects on our lives. In April 2010, for example, a volcano in Iceland, which had been dormant since 1821, suddenly burst through its glacial covering and filled the skies with volcanic ash. Fearing that the metallic particles would damage jet engines, and perhaps cause the airplanes to plunge to the sea or ground, European air traffic was stopped for days. The recent Iceland eruption was nevertheless relatively modest as far as terrestrial volcanoes go. In June 1991 Mount Pinatubo spewed out enough material to shade and cool much of the planet, altering its climate. About a century earlier, on 27 August 1883, the volcano-island of Krakatoa exploded, producing an immense tsunami that killed tens of thousands of people. Vesuvius, the volcano that buried the Roman city of Pompeii in 79 AD, is a disaster waiting to happen. The Italian government has emergency evacuation plans for more than half a million people that now live in the densely populated area near the volcano.

The upwelling of pent-up heat and magma also forms rift valleys on Earth, with steep sides, sunken floors, and copious outpouring of lava. An example is the Great Rift Valley in Africa (Fig. 2.27), a long forking gash that crosses 4500 kilometers of the continent. It extends from Mozambique in the south to Ethiopia in the north, branching out through the Red Sea in one direction and diverging through the Gulf of Aden in another.

Lava flows on the Earth's Moon

Lunar rocks, which were brought to Earth by the *Apollo* astronauts in the 1970s, are of two main types. There are relatively bright rocks from the heavily cratered highlands of the Moon, and dark rocks from the lunar maria. The highlands were cooled long ago from a formerly melted magma ocean, and were extensively modified by the heavy bombardment that formed the craters upon craters between 4.3 and 3.8 billion years ago. The largest craters and impact basins were subsequently flooded episodically

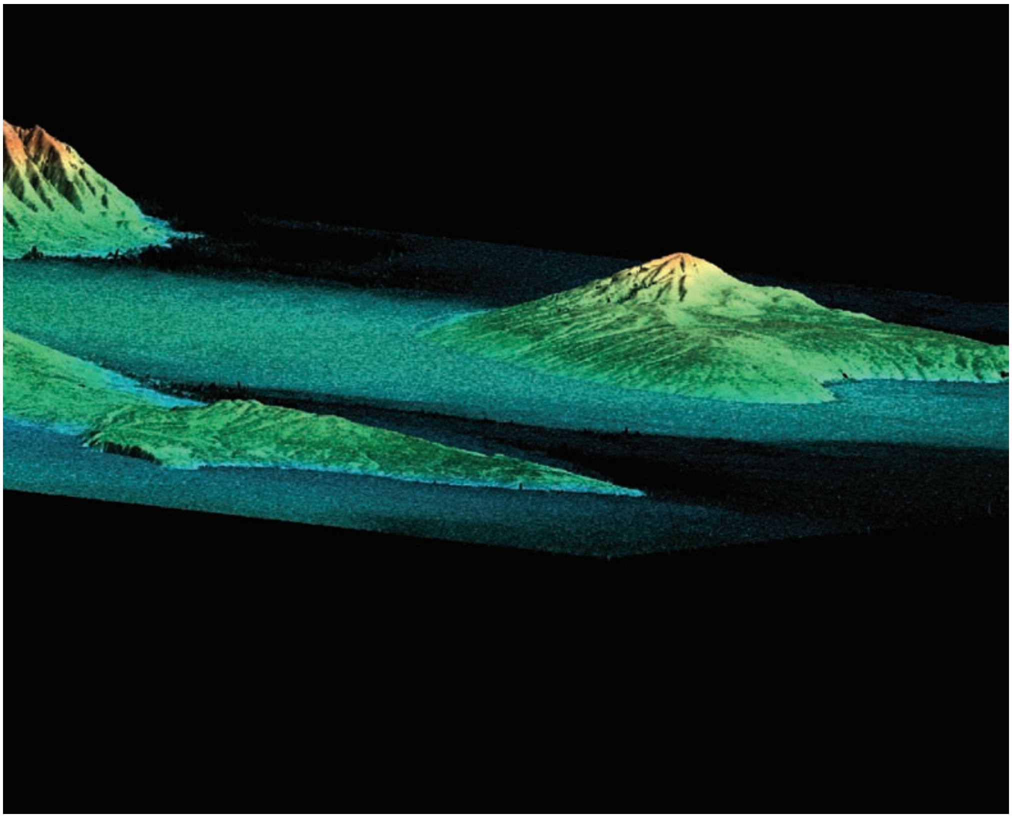


Fig. 2.26 Hawaiian Islands Molokai (*left*), Lanai (*right*), and the northwest tip of Maui (*upper left*) are shown in this radar image obtained from the *Space Shuttle*. These islands are now extinct volcanoes, destined to eventually erode away into sunken islands known as seamounts. (Courtesy of NASA/JPL/NIMA.)

Fig. 2.27 Nyiragongo volcano flow in Africa The continent of Africa is being split apart by the pent-up pressure of hot, rising magma in numerous underlying hot spots along the Great Rift Valley. Volcanic outpourings like Nyiragongo fill the valley with lava as the rift slowly widens. (Courtesy of Bruce Coleman.)

Fig. 2.28 Lava flows in a lunar maria Volcanism on the Earth's Moon is seen frozen into place on the Sea of Serenity in this *Apollo 17* image taken in December 1972. Craters (*bottom*) are superposed on the lava, but the lunar maria contain relatively few craters when compared with the lunar highlands. The maria formed a secondary crust on the Moon when lava filled the giant impact basins over a period of several hundred million years ending around 3.2 billion years ago. The fluid spread rapidly, creating thin extensive sheets rather than piling up to form volcanoes. (Courtesy of NASA.)

by extensive lava flows, forming the dark maria between 3.9 and 3.2 billion years ago. They have been filled with exceptionally fluid basaltic lava, forming relatively crater-free surfaces (Fig. 2.28).

Substantial amounts of volatile elements like water and carbon dioxide have not been detected in the rocks returned from the Moon, presumably being vaporized away during its origin by a giant impact of a Mars-sized body with the young Earth. Comets may have subsequently deposited water on the cooled lunar surface, perhaps accounting for the small amounts of water that have been detected by remote sensing of the lunar surface and in recent examination of the lunar rocks.

Volcanic activity on Mercury

In the heavily cratered terrain on Mercury, there is a paucity of craters of decreasing size when compared to those of the lunar highlands. The smaller craters were probably obliterated by vast volcanic flows that occurred during the period of late heavy bombardment about 4.0 billion years ago. *Mariner 10* scientists designated this terrain as intercrater plains, since they are found between large craters and have relatively smooth surfaces.

Images from the *Mariner 10* and *MESSENGER* spacecraft have shown that large craters and impact basins on Mercury have an internal smoothness (Fig. 2.29). They resemble the maria on Earth's Moon, and are similarly attributed to ancient volcanic flow that is younger than the basins they occupy. These smooth plains cover approximately 40 percent of the surface of Mercury, and are evidence for the volcanic origin of a large part of the planet's crust, after the heavy bombardment that excavated the older craters on the planet but before the smaller, younger craters that are superposed on the smooth plains.

Some source regions of volcanism on Mercury have also been discovered in images taken during *MESSENGER*'s flybys. They show bright areas surrounding irregular-shaped depressions that have been identified as volcanic vents.

Extensive volcanism on Venus

Tens of thousands of shield volcanoes have been identified on the face of Venus, by their round shapes and gentle slopes. They range in size from major, Hawaii-sized edifices that are several hundred kilometers across (Fig. 2.30) to more numerous, smaller domes that pop up everywhere on the surface. These shield volcanoes have probably been built up from runny basaltic lava that spreads out over large distances, with the ease of spilt olive oil. The relatively low number of superimposed impact craters indicates that the extensive lava flows resurfaced the planet about 750 million years ago.

A smaller number of volcanic flows on Venus appear to be built from lava that is as stiff and thick as batter. In places, the sluggish lava has oozed onto the hot, flat surface of Venus, forming volcanic domes as round and flat as pancakes (Fig. 2.31). Each one has a dark feature almost precisely at the center, suggesting a vent from which the pasty lava flowed, like pancake batter on a hot griddle. Some of them even have little craters or pits on them that resemble bubbles that have burst in the batter. So, depending on the internal conditions when the magma formed in Venus, the resulting lava has the consistency and viscosity of either motor oil or toothpaste, and this helps determine the size and shape of the resulting volcanic formations.

Shields and plains on Mars

When *Mariner 9* neared Mars in 1971, the planet was engulfed in a dust storm. The eyes of the spacecraft – its cameras – could only peer at a disappointing, featureless ball, but as the dust storms began to settle four dark, round spots – the Tharsis volcanoes – poked out of the gloom. Even the thick blanket of dust could not cover these towering volcanic mountains. Thus, although Mars has just half the radius of the Earth, the red planet is still large

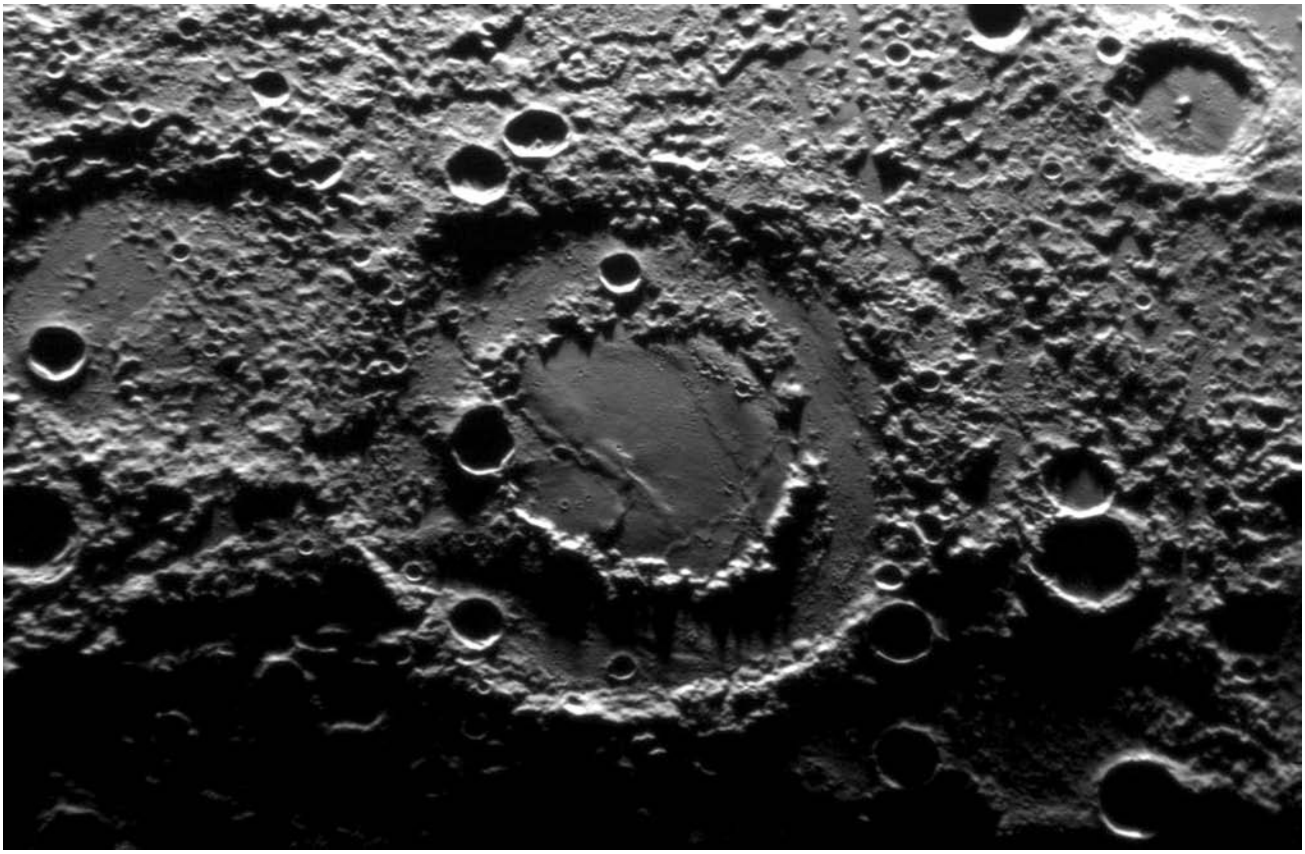


Fig. 2.29 Volcanic activity on Mercury This double-ringed impact basin on Mercury has a smooth inner floor attributed to lava flows that partially flooded the basin some time after impact. The basin is approximately 60 kilometers in diameter. This image was acquired from the *MESSENGER* spacecraft on 29 September 2009. (Courtesy of NASA/JHUAPL/CIW.)

enough to retain significant amounts of internal heat and to sustain long periods of volcanic activity.

The large volcanoes on Mars have the gentle slopes and rounded profiles of shield volcanoes on Earth, but the volcanoes on Mars stand higher. A striking example is Olympus Mons (Fig. 2.32). Highly fluid lava has flowed out of Olympus Mons for 100 million years or longer. The caldera at its peak was formed after the most recent volcanic episode ceased and the roof of the emptied magma chamber collapsed.

Olympus Mons is very much larger than any volcano on Earth. The major volcanic edifice is about 600 kilometers across at its base, stands more than 27 kilometers above the mean surface level of Mars, and is rimmed by a cliff that is 6 kilometers high in some places. For comparison, the diameter of the base of Hawaii's Mauna Loa is just one fifth that of Olympus Mons, and the height of the Hawaiian volcano is a third the height of the Martian one.

The impressive size of the volcanoes on Mars is attributed to the planet's thick outer shell, which does not move sideways and remains fixed over the internal sources of magma. This gives the volcanoes on Mars a long time to grow, sometimes for billions of years. In contrast, the Earth's thinner crust is broken into pieces and moves over an internal magma chamber, limiting the growth of individual terrestrial volcanoes and producing chains of smaller ones, such as the Hawaiian Islands.

Images of the Martian surface suggest that volcanic activity might have persisted from the planet's youth into relatively recent times. Like the Moon and Mercury, the red planet bears the scars of a steady rain of meteorites, and the relative ages of volcanoes and lava flows can be determined from the density of impact craters on them. While the most recent lava flows on Olympus Mons may be only a few million years old, and the average lava age is about 30 million years, lava could have been flowing out of this volcano for a long time before that, as long as 3 billion years ago.

The bulk of the volcanism on Mars is found in the flat, low-lying plains that cover most of the northern hemisphere and about 40 percent of the planet's surface. These lava flows are covered by wind-blown dust, as well as rock debris sent down into them by ancient episodes of water

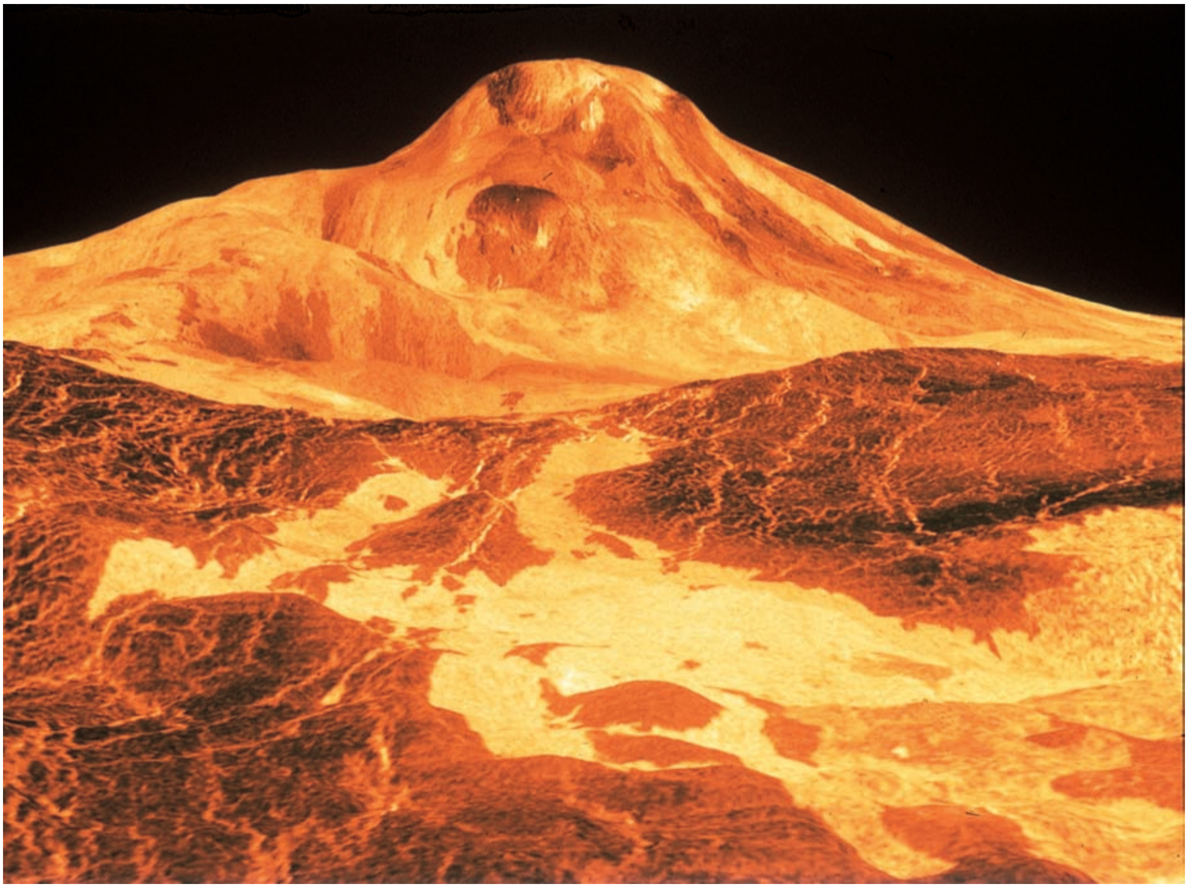


Fig. 2.30 Maat Mons, a volcano on Venus This three-dimensional perspective of Maat Mons on Venus was obtained from radar data taken from the *Magellan* spacecraft in October 1991. The volcano is 8 kilometers high, the second highest peak on the planet. Fresh, dark lava extends for hundreds of kilometers in the foreground, perhaps flowing from a relatively recent eruption. Maat Mons is a giant shield volcano similar in size and shape to the big island of Hawaii. *Maat* is the name of the ancient Egyptian goddess of truth and justice, and *Mons* is the Latin term for "mountain". The orange tint simulates the color of sunlight at ground level after filtering by the dense, thick atmosphere of Venus. (Courtesy of NASA/JPL.)

flow from the southern highlands. The low impact-crater densities on some volcanic flows in the Martian plains suggest relatively young ages, as low as 100 million years or less, but lava most likely also flowed across the surface of Mars billions of years ago.

Volcanoes on Jupiter's moon Io

There is one place in the solar system that is now more volcanically active than any other place; it is Jupiter's innermost satellite Io. It is so hot inside that you can see it melting before your eyes. Io is now spewing out 100 times more lava than all the volcanoes on the Earth. This is a totally unexpected discovery, made by the inquisitive camera eyes of *Voyager 1* in 1979, and confirmed by the *Galileo* spacecraft in 1999–2004 (Figs. 2.33, 2.34). Volcanoes are literally turning the satellite inside out, so parts of Io's surface are younger than your backyard. Because of the satellite's low gravity and lack of substantial atmosphere, the volcanic plumes spread out in graceful fountain-like trajectories, depositing circular rings of material up to 1400 kilometers in diameter.

What drives Io's continuous volcanism? The satellite is too small to now retain internal heat created during its formative years or to be significantly heated by radioactive rocks. The heat released during the satellite's formation and subsequent heating of its interior should have been lost to space long ago. After all, Io is about the same size as the Earth's Moon, which shows no signs of volcanism other than that associated with its earliest history between 3 and 4 billion years ago. Io's internal heat is instead generated by tides that massive Jupiter raises in the solid body of the satellite.

The gravitational force of the nearby giant planet decreases with increasing distance, so Jupiter pulls hardest on the side of Io facing it, and least on the opposite

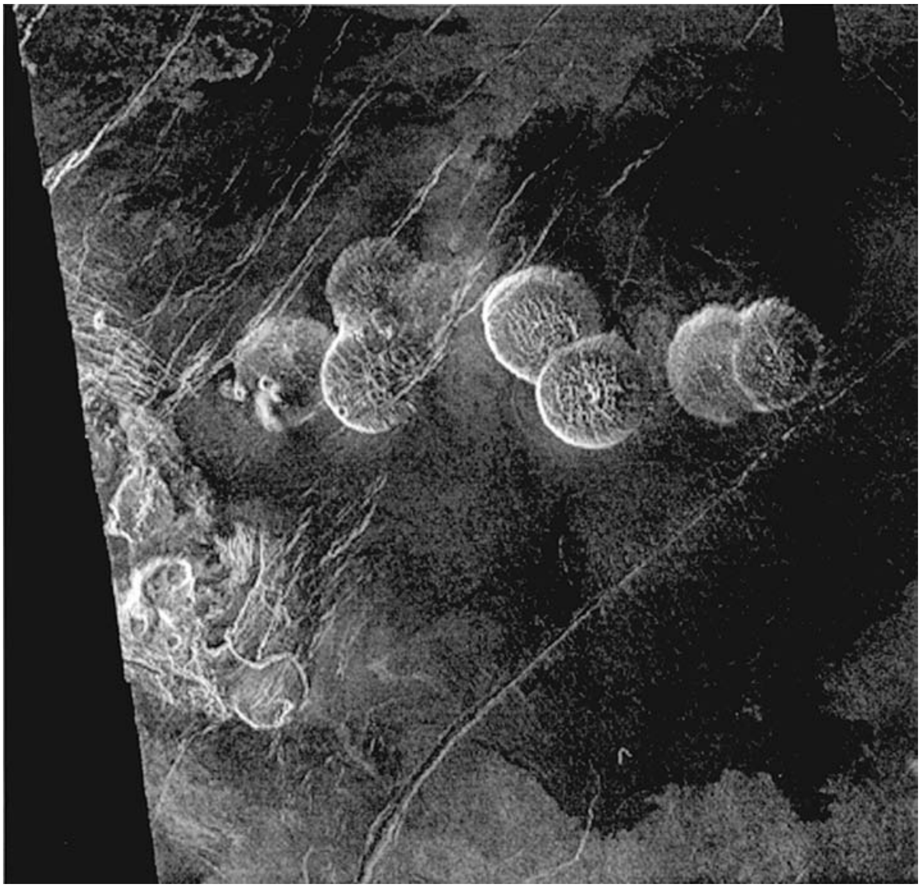


Fig. 2.31 Pancake domes on Venus These seven volcanic domes were discovered in radar images of Venus taken from the *Magellan* orbiter. They all have round shapes that are about 25 kilometers across, and steep sides that are less than 750 meters high. Their central vents may be lined up along a crack in the surface. These domes are attributed to very thick, stiff and sluggish lava flows, rather than the fluid and runny type. Eruptions of the pasty, viscous lava, coming from a central vent on a relatively level surface, would form the circular, flattened shapes that resemble giant pancakes. Since there is little or no erosion by wind or water on Venus, newer pancakes look much the same as the ones on which they are superimposed. (Courtesy of NASA/JPL.)

side; the center of Io is pulled with an intermediate force. These differences in the gravitational attraction of Jupiter on opposite sides of Io produce two tidal bulges in the solid rock inside the satellite – one facing Jupiter and one facing away. The giant planet thus effectively squeezes Io into the shape of an egg.

If Io remained in a circular orbit with the same face toward Jupiter, its tidal bulges would not change in height and no heat would be generated; but its orbit is not perfectly circular. Io's orbit has a forced eccentricity due to the combined gravitational interaction of Io with Jupiter and the other satellites of the giant planet. When the elongated orbit carries Io closest to Jupiter, the shape of Io is distorted more than when the satellite is further away. The resultant variation in the tides flex Io's surface, bending it in and out by as much as 100 meters during each orbit. Friction associated with this tidal flexing heats Io inside, melting its rocks and producing volcanoes at its surface.

Instruments on *Galileo* measured the temperatures of the volcanoes, showing that the lava is at 1700 to 2000 kelvin, up to twice the temperature of volcanoes on Earth. The high-temperature eruptions emit gaseous sulfur and sulfur dioxide; the bright surface flows are attributed to sulfur and the white surface deposits to sulfur dioxide. The very high temperatures apparently rule out liquid sulfur as a dominant volcanic fluid, and they have certainly driven off any water that might have been on Io.

Volcanoes of ice

Some of the large icy satellites of the giant planets exhibit fountains, geysers and smooth surface flows of water ice. This is a sort of ice volcanism, in which the generation of liquid water inside the satellite mimics the partial melting of rocks within the terrestrial planets and Io. The melted ice works its way out from inside the

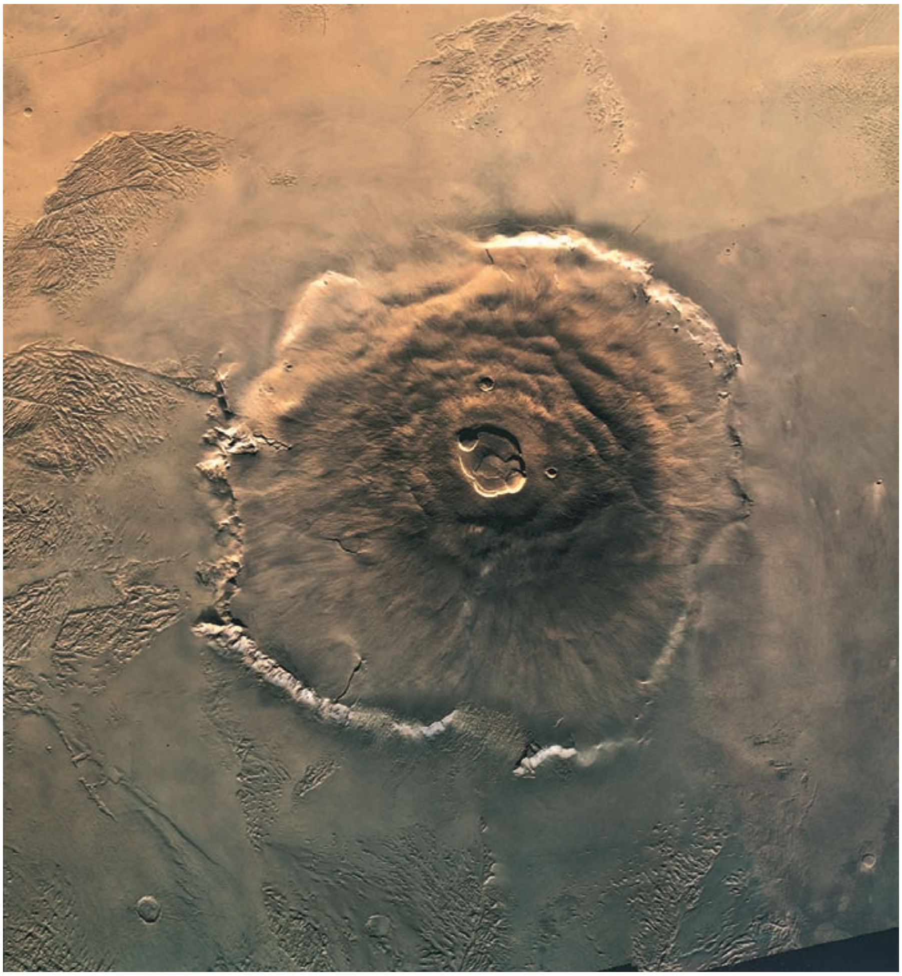


Fig. 2.32 Olympus Mons, a volcano on Mars A mosaic of the towering Martian volcano Olympus Mons, using data obtained from the *Viking 1* orbiter in the late 1970s. It is the largest known volcano in the solar system, rising about 27 kilometers and spreading over 600 kilometers at its base. Counts of impact craters suggest that the lava flows on the gentle slopes of this volcano are relatively young, averaging only about 30 million years old. The summit caldera, or central depression, is a composite of as many as seven roughly circular depressions that formed by recurrent collapse when magma was withdrawn from within the volcano. The caldera is almost 3 kilometers deep and up to 70 kilometers across. The volcano is surrounded by a well-defined scarp, or cliff, that is up to 6 kilometers high. Many of the plains surrounding the volcano are covered by terrain containing ridges and grooves; it is called an *aureole*, the Latin term for "circle of light". *Mons* is the Latin term for "mountain". Mount Olympus, the highest mountain in Greece, is the home of the gods in Greek mythology. (Courtesy of NASA/JPL/USGS.)

satellite and either erupts from it or flows across the surface.

Jupiter's satellite Europa, for example, is almost perfectly smooth and exceptionally bright, with no mountains and valleys in sight (Fig. 2.35). Very few impact craters are present on its face, indicating that the smooth surface was formed relatively recently, geologically speaking. Some process must be keeping it young on timescales of a few hundred million years or less. Liquid water or slush apparently oozes out within cracks in the ice, resurfacing the globe (Fig. 2.36). Some cracks in the icy moon are as long as the distance from Los Angeles to New York, and when you look at them you might see water rising.

Saturn's moon Enceladus has a bright, smooth, icy surface that also contains cracks and grooves (Fig. 2.37), suggesting the release of water from below the surface. This would be consistent with the satellite's low mean mass density of just 1240 kilograms per cubic meter, suggesting that it is just a big ball of water ice.

The satellite is embedded within a ring of water ice, dubbed the E ring, which is derived from the satellite. Instruments aboard the *Cassini* spacecraft have revealed ice jets erupting from fractures in Enceladus' surface (Fig. 2.38). Geysers of water, carbon dioxide and organic molecules spray far out from the moon at high speeds. Some of the fountains and plumes of water ice erupt at

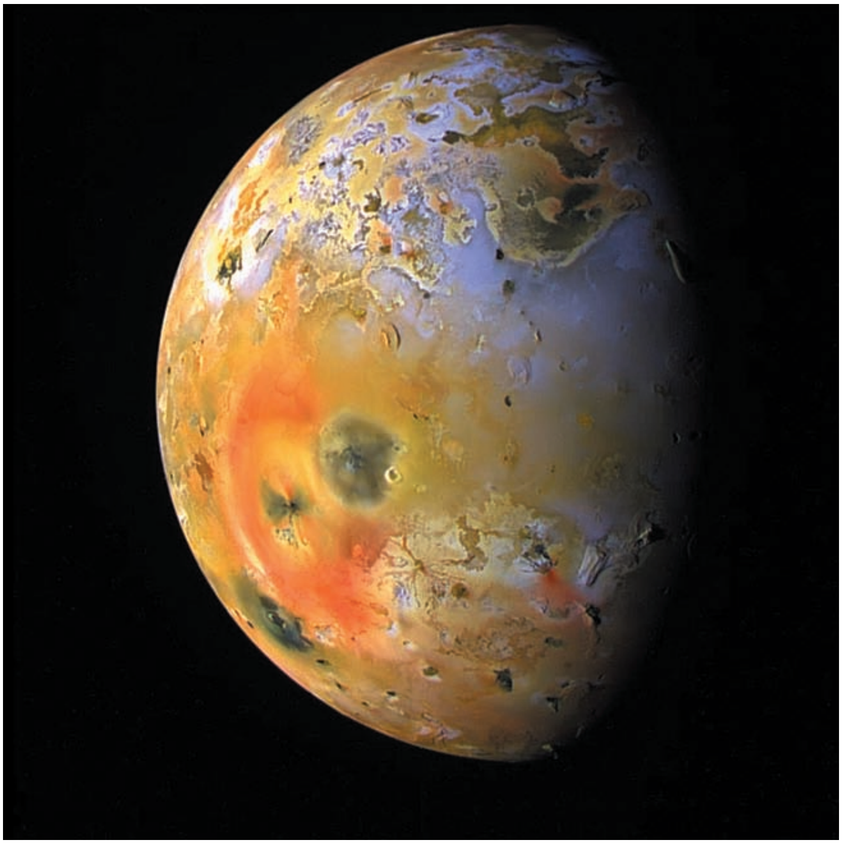


Fig. 2.33 Volcanic activity on Io Massive eruptions continuously disfigure the surface of Jupiter's satellite Io, the most volcanically active body in the solar system. As shown in this color-enhanced image, taken from the *Galileo* spacecraft on 19 September 1997, Io's surface is continuously being covered by lava flowing from its volcanoes, erasing any impact craters. A bright red ring surrounds the volcano Pele, marking the site of sulfur compounds deposited by its volcanic plumes. A dark circular area, about 400 kilometers in diameter, intersects the upper-right part of the red ring and surrounds another volcanic center named Pillan Patera. Deposits of sulfur dioxide frost appear white and gray in this image, while other sulfurous materials probably cause the yellow and brown shades. Pele is the Hawaiian goddess of the volcano, and Pillan Patera is named for the Araucanian thunder, fire and volcano god. (Courtesy of NASA/JPL/U. Arizona.)

Fig. 2.34 Lava flows on Jupiter's satellite Io Numerous volcano calderas and lava flows were discovered on Jupiter's innermost large moon, Io, in 1979 using an instrument aboard the *Voyager 1* spacecraft. This *Voyager 1* image of Ra Patera, a large shield volcano, shows flows up to 300 kilometers long emanating from a dark volcanic vent. The diffuse reddish and orange colorations are probably surface deposits of sulfur compounds. The bright whitish patches probably consist of freshly deposited sulfur dioxide frost. Ra is an Egyptian Sun god. (Courtesy of NASA/JPL.)

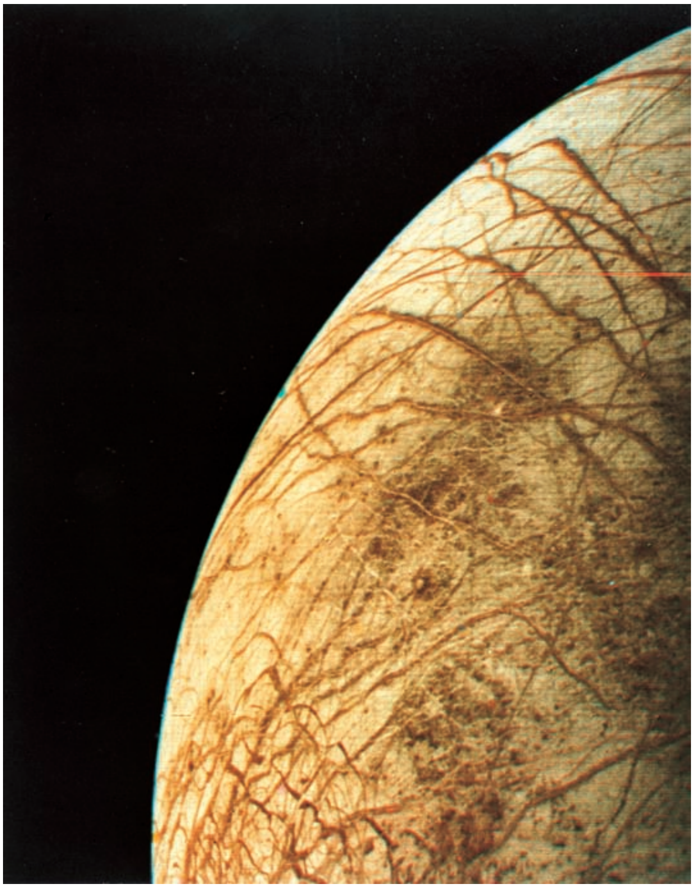


Fig. 2.35 Jupiter's satellite Europa Dark streaks mark Europa's smooth surface, forming a spidery, veined network in this *Voyager* 2 image taken on 9 July 1979. In contrast to Jupiter's satellite Callisto (see Fig. 2.17), Europa has very few impact craters; the absence of craters suggests that the ice crust is relatively young. Internal stresses have apparently fractured Europa's icy mantle, producing intersecting cracks that extend 2000 kilometers but reach depths of less than 100 meters. The fractures may have been filled by liquid water gushing out from a global ocean in the satellite's interior, warmed by tidal heating. (Courtesy of NASA/JPL.)

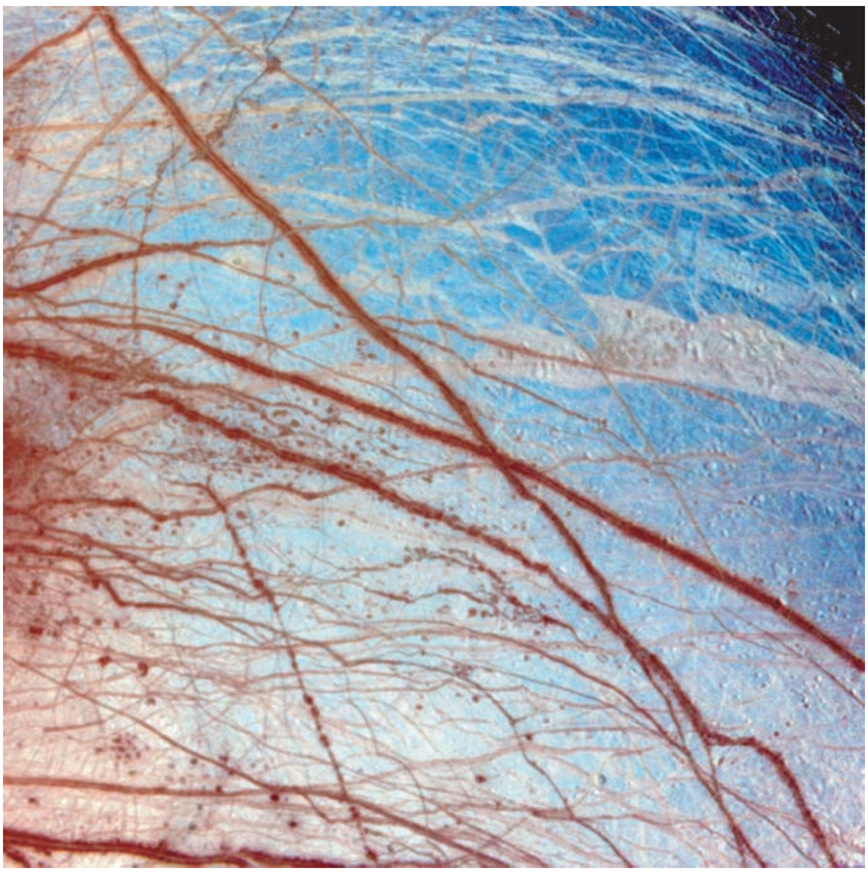


Fig. 2.36 Water oozes out from Jupiter's moon Europa A composite, color-enhanced image of the Minos Linea region of Jupiter's moon Europa, taken on 28 June 1996 by imaging cameras on *Galileo*. The icy plains, shown here in bluish hues, reflect different amounts of light, probably as the result of differences in the sizes of the ice grains. The long red cracks in the ice could mark the sites of liquid water oozing out from the warm interior of Europa. The area covered in this image is about 1.26 kilometers across. In Greek mythology, Minos is the son of Zeus and the king of Crete, who kept a monster named Minotaur in a labyrinth. *Linea* is a "dark or bright elongate marking". (Courtesy of NASA/JPL/U. Arizona.)

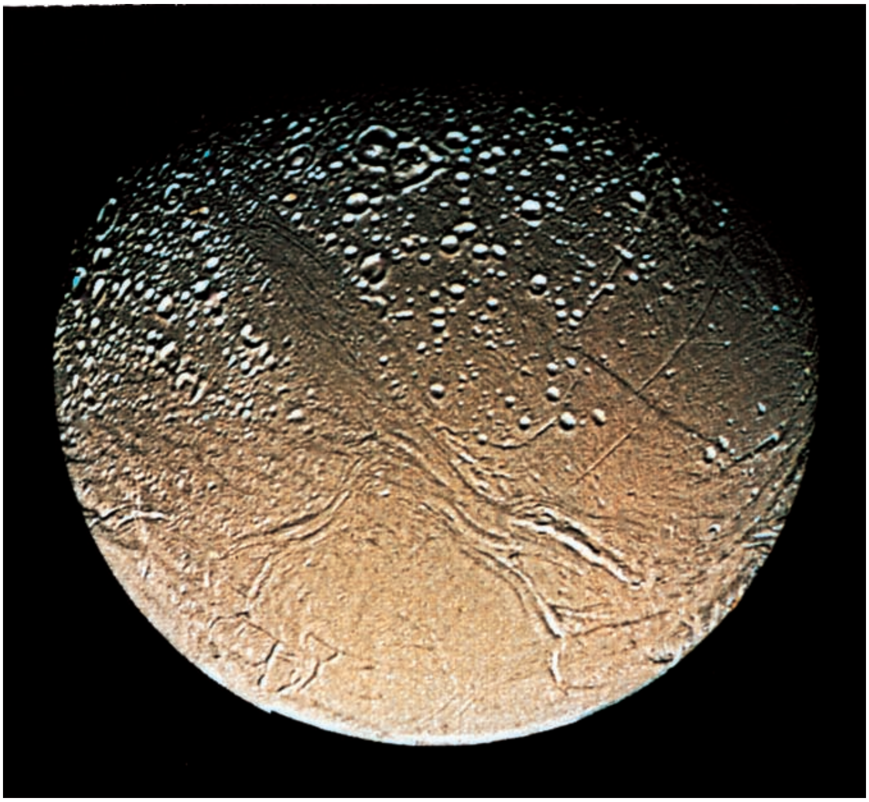


Fig. 2.37 Saturn's satellite Enceladus The bright, smooth surface of Enceladus, shown in this *Voyager 2* image obtained on 25 August 1981, reflects almost 100 percent of the incident sunlight, making it one of the most reflective objects in the solar system. When viewed up close, part of its surface is scarred with impact craters. Other parts of the surface contain cracks and grooves, suggesting that internal stresses may have discharged water that froze into smooth ice. (Courtesy of NASA/JPL.)

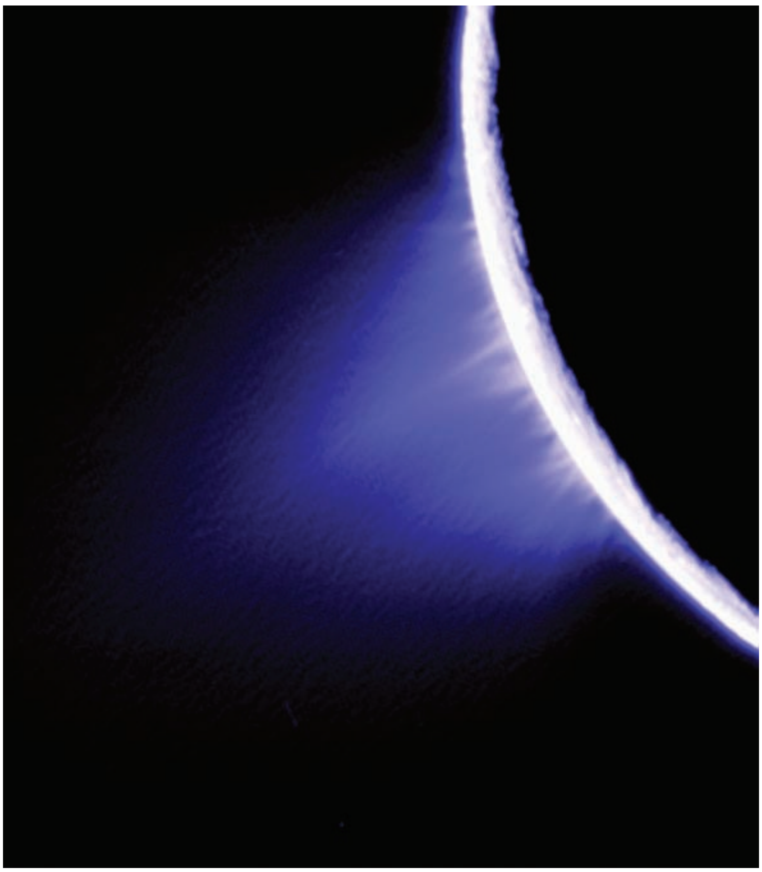


Fig. 2.38 Ice plumes on Saturn's moon Enceladus Enormous jets and fountains of ice are erupting on Saturn's moon Enceladus, feeding the planet's E-ring. This image, taken from the *Cassini* spacecraft on 27 November 2005, exhibits several geyser-like jets, which vent and spurt plumes of ice particles, water vapor and trace amounts of organic compounds. Eight source locations were identified in this image, all on the prominent tiger stripe features, or sulci, in the moon's south polar region. These features were under close scrutiny from *Cassini* for years after their discovery in 2005. (Courtesy of NASA/JPL/SSI.)

fast enough speeds to escape Enceladus and feed the E ring with ice particles.

Enceladus is caught in a gravitational tug-of-war between Saturn and its satellite, Dione, whose orbital period is about twice that of Enceladus. Dione's repeated gravitational tug produces Enceladus' eccentric orbit, and causes Saturn's recurrent tidal flexing of Enceladus, which warms the moon's interior and apparently maintains a liquid ocean beneath its ice.

Neptune's largest satellite, Triton, is the coldest moon ever recorded, with a temperature of just 38 kelvin, approaching absolute zero where all motion stops. It is so cold because Neptune is so far away from the Sun, therefore receiving little sunlight, and also because Triton reflects more of the incident sunlight than most satellites – only Enceladus and Europa are comparable. Yet the frozen moon is a dynamic world, set in motion and molded by volcanic eruptions.

Triton's surface has a smooth, youthful appearance, with no large impact craters and few small ones. Global resurfacing by volcanoes of ice might have wiped out pre-existing craters on Triton, perhaps about a billion years ago when tidal flexing may have heated the satellite's insides. The deep internal heat may have turned the ice into liquid that rose to the surface, like a squeezed slush cone, filling the vast frozen basins that are now found there. These frozen lakes of ice look like inactive volcanic calderas; complete with smooth filled centers, successive terraced flows and vents.

Numerous dark plumes and streaks, found in the midst of the bright southern cap of Triton, suggest a different kind of volcanic activity, propelled by relatively recent eruptions of nitrogen gas (Fig. 2.39). Nitrogen boils at very low temperatures, at just 77 kelvin on Earth, and when it boils it expands, producing enormous pressures that can shoot gas and other material high into Triton's thin nitrogen atmosphere. Thus, geyser-like eruptions may have lofted the dark material outward from beneath the surface. The prevailing winds would then carry it across the satellite, depositing it on the ice as dark streaks.

Four active plumes were observed during the *Voyager 2* encounter with Triton. They rose in narrow, straight columns to an altitude of 8 kilometers, where dark clouds of material were left suspended and carried downwind horizontally for over 100 kilometers, like smoke wafted away from the top of a chimney. Most of the dark streaks are probably remnants of such plumes.

Since the active plumes occur where the Sun is overhead, it is possible that sunlight produces the weak subterranean heat required to make the nitrogen boil and break through the overlying layer of ice. The sunlight would pass through the translucent ice and become absorbed by

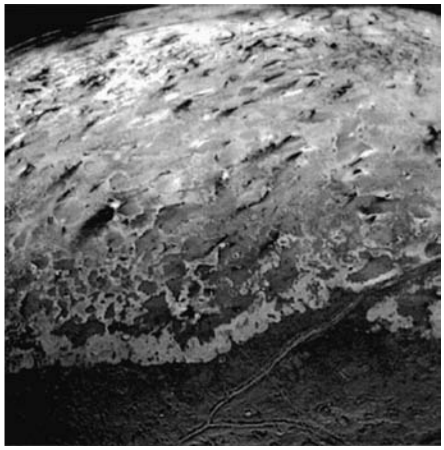


Fig. 2.39 Triton's dark plumes and streaks This image of the south polar terrain on Triton reveals about 50 elongated dark plumes, or "wind streaks", on the moon's highly reflective surface. The plumes originate at very dark spots, generally several kilometers across, probably marking vents where nitrogen gas was driven outward in geyser-like eruptions from beneath the surface. Winds in Triton's thin nitrogen atmosphere may have carried the dark erupted material along, depositing it in the elongated streaks. This image was taken on 25 August 1989 from *Voyager 2*. (Courtesy of NASA/JPL.)

darker material encased beneath. The overlying nitrogen ice would trap the solar heat, for it is opaque to infrared heat radiation, producing a solid-state greenhouse effect. Nitrogen gas, pressurized by the subsurface heat, might then explosively blast off the iced-over vents or lids, launching volcanic plumes of gaseous nitrogen and ice-entrained darker material into the atmosphere, just as the water in an overheated car radiator is explosively released when the radiator cap is removed. The nitrogen geysers also resemble water-driven geysers on Earth, like Yellowstone's Old Faithful, but water boils at a much higher temperature of 373 kelvin.

This discussion of ice volcanism on Europa, Enceladus and Triton brings us to water, another common property of many planets and satellites.

2.4 Water

Earth, the water planet

From outside, our home planet Earth looks like a tiny, fragile oasis in space, a glistening blue and turquoise ball of water, flecked with delicate white clouds and capped with glaciers of ice (Fig. 2.40). Seventy-one percent of the Earth's surface is now covered with water. The oceans contain so much water that if the Earth were perfectly smooth the oceans would cover the entire globe to a depth

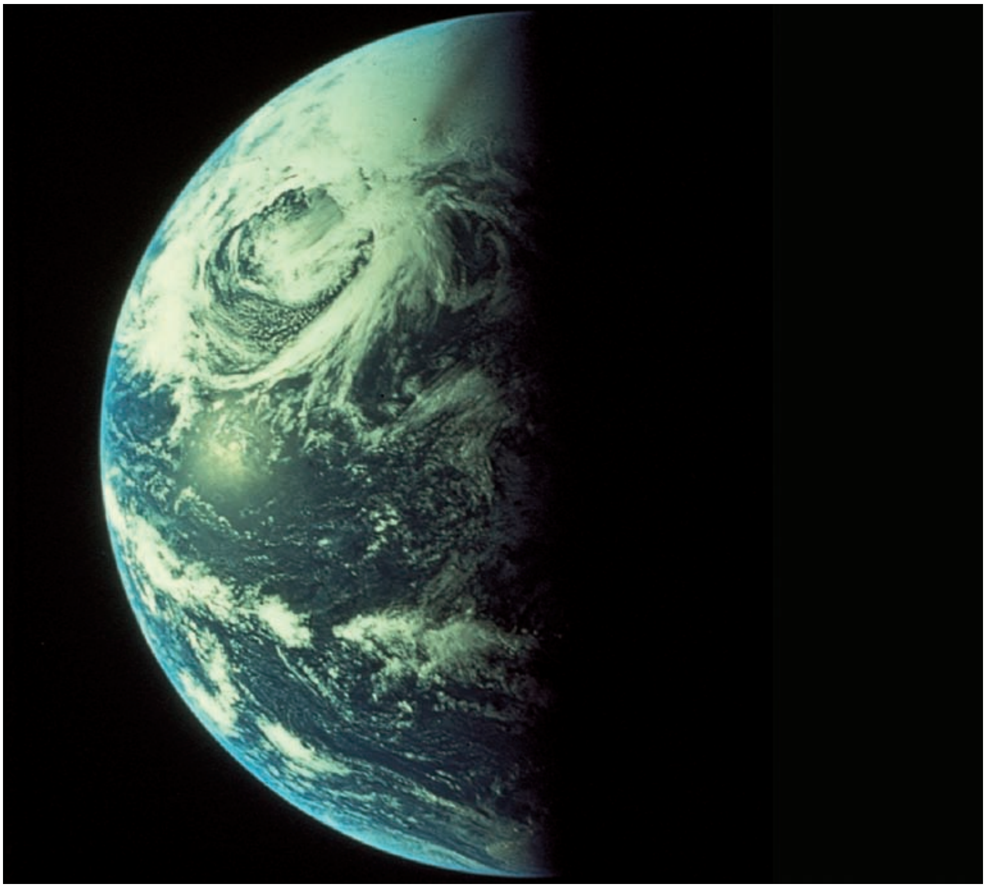


Fig. 2.40 Earth, the water planet Almost three-quarters of the Earth's surface is covered by water, as suggested by this view of the North Pacific Ocean. Earth is the only planet in the solar system where substantial amounts of water exist in all three possible forms – gas (water vapor), liquid and solid (water ice). Here white clouds of water ice swirl near Alaska. The predominantly white ground area, consisting of snow and ice, is the Kamchatka Peninsula of Siberia. Japan appears near the horizon. From this orientation in space, we also see both the day and night sides of our home planet. (Courtesy of NASA.)

of 2.8 kilometers. They contain about one billion trillion (10^{21}) kilograms of water, and provide the home for about half of all species on Earth, including the one-celled plants called phytoplankton, which supply roughly half of the oxygen to our air.

The Earth's oceans absorb heat in hot months, and release it in cold months. Global ocean currents such as the Gulf Stream circulate heat and cold around the world. The oceans also absorb carbon dioxide from the atmosphere, removing roughly half of the heat-trapping gas that is added to our air.

Water is a marvelous substance. As a liquid, it will dissolve almost anything to some extent, and it can hold and release very large quantities of heat. When liquid water freezes, it expands and becomes less dense, in contrast to most substances. As a result, ice floats on the surface of lakes and oceans, so they freeze from the top down.

Water is crucial to life here on Earth, for the chemical reactions that sustain life must take place in water. We ourselves are largely water. Just about anywhere there is water here on Earth, there is some sort of life — even deep inside the Earth. It follows that if liquid water was found on another planet, or on one of its satellites, that place might also be hospitable to life.

Water is made of the two most abundant, chemically reactive elements in the Universe, hydrogen and oxygen, and so ought to be very common. Yet the Earth is the only place in the solar system where substantial quantities of water exist in all three possible states — as a gas (water vapor), liquid, and solid (ice). Ours is the only planet whose surface temperature matches the temperature of liquid water, between water's freezing and boiling temperatures of 273 and 373 kelvin, respectively. The narrow range of distances from the Sun, or other star, at which temperatures from stellar radiation allow liquid water to exist is known as the habitable zone; but the greenhouse warming of a planet's atmosphere can raise the temperature above that due to starlight alone.

When we look at our nearest neighbors, we see that Venus is too hot and Mars is too cold for significant amounts of liquid water to exist on their surfaces. Any water on Venus would now be in the form of steam, and water on Mars is now mainly locked beneath the surface in the form of ice. The terrestrial planets nevertheless had different surface temperatures in the past and these temperatures will change in the future.

It was too hot for liquid water to exist on any of the rocky inner planets in their very early history. Because of the energy released by the colliding rocks that merged to form these planets, they probably began with molten surfaces, too hot for any water to accumulate on them. The heat of impacts would have vaporized any water, and sterilized the young planet. Only when the initial bombardment slowed, and the growing planet cooled, could liquid surface water collect into lakes or oceans, which might have sustained life.

This water might have been liberated from the planetary interiors by volcanoes or carried to the planets by icy comets or water-rich asteroids. The water vapor released from the hot internal materials would cool and condense, falling as rain. Other water may have arrived by the direct impact of comets or asteroids. On Earth, the water fed rivers that flowed across the planet's surface, dissolving salty minerals that settled into the early oceans.

Some astronomers argue that the Earth's oceans were supplied from outside the planet after it first came together. The water was supposed to be carried to our planet by small bodies of rock or ice, similar to today's asteroids or comets, coming in from the cold outer parts of the solar system. Some geologists reason that the ancient oceans were instead steamed out of the Earth's interior by erupting volcanoes, supplying water vapor to the primitive atmosphere. No one knows for sure whether external impact or internal volcanism resulted in most of our water. But since the Earth formed by the accumulation of colliding objects, the water expelled by the volcanoes also had to be originally supplied by cosmic impact.

Water on the Moon

The rocks returned from the Earth's Moon by astronauts of the *Apollo* missions are dryer than a terrestrial desert. They resemble an Earth rock that has had all the water boiled out of it. For this reason, most scientists have assumed that there is no water on the Moon, and that there never was any.

There has nevertheless been speculation since the 1960s that water ice might be found in the polar regions of the Moon, which the astronauts did not visit. This possibility hinges on the fact that the Sun never rises more than a scant 2 degrees above the horizon as seen from the lunar poles. This means that the deepest polar craters have regions inside them that remain permanently in shadow, eternally dark and cold with temperatures that never exceed about 100 kelvin. Consequently, any water that chanced to enter the craters would be frozen solid, remaining permanently frozen, never melting or vaporizing and escaping to space.

If there were water on the Moon, it would probably have been deposited by impacting comets, which are comprised largely of water ice, or by the impact of water-rich asteroids. They have both been bombarding the moons and planets for billions of years, ever since their formation. Liquid water and water vapor, formed by the heat of impact on the Moon, would normally be evaporated into space, but over the eons of impact some of the water could have collected in the shadowed craters near the lunar poles. If the water settled down to the crater floors, it would freeze in their "cold trap". Or perhaps hydrogen from the Sun's winds could combine with oxygen atoms in the Moon's dust and rock, forming water molecules that might enter the cold traps.

For more than a decade, successive observations have either strengthened evidence or dampened hopes for water on the Moon. Early evidence for the water began in 1994 when the US Department of Defense and NASA launched the *Clementine* spacecraft into polar orbit around the Moon, obtaining the first global perspective of its surface composition and topography. Bright radar echoes obtained from an instrument on *Clementine* suggested deposits of water ice in a deep, cold, shadowed crater near the lunar south pole. Indirect supporting evidence was obtained from the *Lunar Prospector* spacecraft, launched in 1998. Its neutron and gamma ray spectrometers discovered enhanced signals from hydrogen atoms that were being struck by energetic cosmic rays. If the hydrogen, designated H, was taken as evidence for water, or H_2O, with O for oxygen, then there seemed to be plenty of water ice near both lunar poles.

Astronomers hoped that more direct evidence of water would be obtained by crashing *Lunar Prospector* into a crater near the Moon's south pole, at the end of the spacecraft's lifetime. The impact might release icy rock and dust, with indications of water vapor. The *Hubble Space Telescope* and numerous ground-based telescopes were focused on the impact, on 31 July 1999, but no debris or spectral signatures of the impact were detected. There were several explanations, including the possibility that the spacecraft missed its target. It is also possible that many of the hydrogen nuclei that have been detected were delivered to the Moon by the Sun's winds, and that they are not a component of water.

Hopes for water on the Moon were further dampened in 2004–06 when the world's most powerful radar transmitter, at the Arecibo Observatory in Puerto Rico, was beamed into the permanently shadowed crater floors at the Moon's polar regions. The absence of strong radar echoes indicated that there are no thick ice deposits at the lunar poles, but thin ones with relatively low concentrations of water were possible.

Then, in September 2009, a trio of satellites obtained evidence for water molecules on the Moon. A spectral mapping instrument carried aboard India's *Chandrayaan-1* spacecraft detected infrared absorption patterns attributed to water molecules and hydoxyl, or OH. Spectrometers aboard the *Cassini* and *Deep Impact* spacecraft contributed to confirmation of this finding.

Within a month, the *Lunar CRater Observation and Sensing Satellite* (*LCROSS*) sent two spacecraft in to bomb the Moon. Although the craft successfully struck their target, in the Cabeus crater near the lunar south pole, no billowing clouds of dust and ice were detected by telescopes trained on the impact site and the televised event was quite a disappointment. Careful scrutiny of the *LCROSS* camera images nevertheless revealed plumes of material ejected from the bottom of the crater on impact, and the satellite's spectrometers showed the spectral signatures of water vapor.

We therefore no longer consider the Earth's Moon to be a completely dry and desolate place. But there aren't any lakes or even puddles of water on the Moon, and there isn't even enough water to drink. The lunar water molecules are bound to other molecules in rock and dust, and exist in only trace amounts.

So why is so much attention being given to the search for water on the Moon? Significant reservoirs of water could reduce the payloads needed to maintain a future human outpost on the Moon. The lunar water could be purified to drink, or it could be chemically split into hydrogen, to burn as a rocket propellant, and oxygen to breathe. This would make it easier to support a colony on the Moon, or to build a fueling station on it for interplanetary spacecraft.

Radar evidence for water ice near Mercury's poles

Since Mercury is so close to the Sun, with boiling temperatures on its sunlit side, it has long been assumed that the planet retains no water on its surface. But a strong tidal lock with the Sun makes Mercury's rotation axis point straight up, without any tilt, so its equator points directly at the Sun at all times, and its polar regions are always in shadow. This means that crater interiors near the poles are never exposed to direct sunlight. The permanently shadowed spots may have remained colder than 120 kelvin for eons, permitting substantial quantities of water ice to accumulate at the frigid crater floors near the poles.

When Mercury's poles tip toward the Earth, while never deviating from the north–south direction and remaining hidden from the Sun, astronomers have beamed radio signals at them and examined the echoes. These are return signals from pulses of radio radiation sent from both the Arecibo and the linked Goldstone – Very Large Array radar facilities. Unusually strong radar echoes coming from the polar regions show prominent, radar-reflective material that is plausibly attributed to water ice. They suggest that the planet's north and south polar regions may contain substantial deposits of water ice, at least a couple of meters thick. The similar radar-scattering properties of Mercury polar craters and those of Jupiter's ice-covered satellites supports the suggestion that the craters contain water ice.

Unlike the weak radar echoes obtained from the Moon, the powerful signals returned to Earth from Mercury's highly reflective polar regions suggest relatively thick sheets of water ice at the planet's poles. They appear to be concentrated only in the young, fresh craters seen in *Mariner 10* images, rather than the ancient, degraded ones, which do not exhibit strong radar reflectivity. This is most likely because the older low-rimmed, shallow craters do not contain permanently shadowed floors that act as cold traps for the water.

A former ocean on Venus

Today the surface and atmosphere of Venus are exceptionally dry, which is what you would expect for a planet whose surface temperature is now a scorching 735 kelvin. The planet may nevertheless once have had liquid oceans, perhaps 4 billion years ago, until a runaway greenhouse effect boiled it all away.

Models of planet formation predict, for example, that the Earth and Venus were once endowed with roughly equal amounts of water. When the strong greenhouse effect of Venus's thick atmosphere raised the planet's surface temperature, most of its water evaporated and was lost to space, while the Earth, with a much thinner atmosphere and weaker greenhouse effect, remained cool enough to keep its oceans. And if the very small quantities of water vapor now found in the atmosphere of Venus are a remnant of an ancient reservoir, then Venus has lost the equivalent of a very large lake or a small ocean.

Evidence that Venus once had an ocean is found in an excess of deuterium now in its atmosphere. Deuterium is an atom chemically identical to hydrogen but heavier and therefore more likely to be retained in the atmosphere. On Earth, it is found in heavy water, which comprises only about 0.016 percent (0.000 16) of the oceans. The natural explanation of the atmospheric deuterium on Venus is that the planet once had vast quantities of normal water, containing light hydrogen, and heavy water, containing deuterium. When these liquids were subsequently boiled away by the intense heat, the lighter hydrogen easily escaped from the planet, but some of the heavier deuterium remained behind as a residue. The amount of remaining deuterium suggests that Venus once had enough liquid water to uniformly cover the planet's surface with a global lake at least 4 meters deep, or just 0.12 percent of a full terrestrial ocean.

Water on Mars

Mars orbits the Sun at the outer edge of the solar system's habitable zone, where water might be either liquid or frozen solid. As expected from its thinner atmosphere and greater distance from the Sun, Mars is a much colder planet than the Earth, with surface temperatures that are usually below the freezing point of water. Instruments aboard orbiting spacecraft have shown that large quantities of water ice exist in the polar caps of Mars and as subsurface ice in many other locations on the planet. Spectroscopic instruments aboard the *2001 Mars Odyssey* spacecraft have, for example, found evidence of subsurface water ice on Mars in large regions surrounding both of the planet's polar regions, with lesser amounts of subsurface ice at mid-latitudes and equatorial regions. The concentration of ice in the upper meter of the ground in the polar plains is surprisingly high – one-fifth to one-third by weight and more than 50 percent water ice by volume. So if you heated one full bucket of this polar material it would be more than half a bucket of water.

Ninety-five percent of the Martian atmosphere is gaseous carbon dioxide; and it contains only small quantities of water ice crystals, in the form of clouds and haze. Such small amounts of water vapor indicate that the Martian atmosphere is drier than the driest of the Earth's deserts. If the water vapor were collected and condensed, it would amount to no more than a good-sized lake. Yet, despite the small amount of water vapor, the cold, thin atmosphere is close to saturation. It is about as wet as it can be. Consequently, the formation of clouds and fogs of water ice is a common feature of Martian weather. Such clouds or fogs have been observed along the flanks of volcanoes, above the polar caps, and in low-lying areas such as canyon floors.

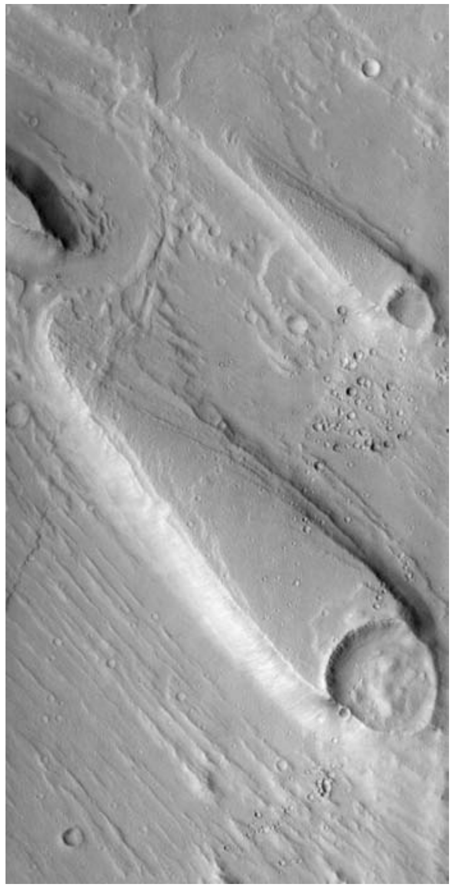


Fig. 2.41 Streamlined islands on Mars An image of the Mangala Vallis region on Mars, taken with an instrument aboard the *2001 Mars Odyssey* orbiter. The scoured floors and teardrop-shaped islands were probably created by powerful, ancient flows of liquid water. The flowing water ran from the heavily cratered southern highlands to the northern lowland plains. The name Mangala is the word for Mars in Sanskrit. (Courtesy of NASA/JPL/ASU.)

Liquid water cannot now exist for any length of time on the surface of Mars. It would immediately begin to boil, evaporate and freeze — all at the same time. Because of the low pressure and freezing temperatures of the thin Martian atmosphere, any liquid water would quickly vaporize or freeze into ice.

Nevertheless, Mars almost certainly contained liquid rivers, lakes and possibly oceans in the distant past, roughly 4 billion years ago. Huge, dry river beds and flood channels, imaged by the *Mariner 9*, *Viking 1* and *2*, *Mars Global Surveyor* and *Mars Reconnaissance Orbiter* spacecraft, provide unmistakable signs of former torrents of flowing water that cascaded across the surface of Mars (Figs. 2.41, 2.42). The flow channels that have been carved and etched into the surface of Mars are immense by terrestrial standards, as much as 100 kilometers wide and 2000 kilometers in length. The amount of water required

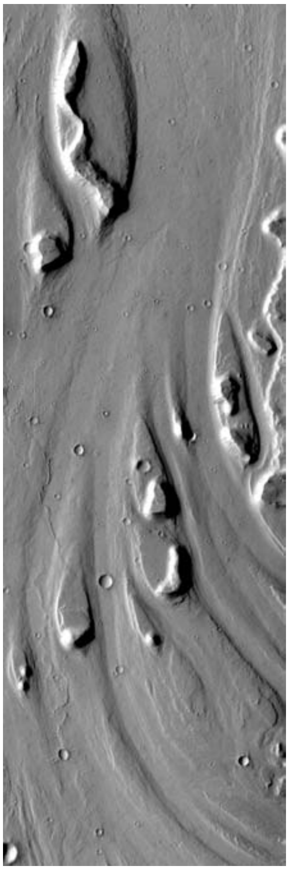


Fig. 2.42 Ancient water flow in Mars' Ares Vallis As shown in this image, obtained from an instrument aboard the *2001 Mars Odyssey* orbiter, large quantities of water were diverted around pre-existing craters in Ares Vallis. As the water made its way downstream, the interference with the flow was reduced, and the water flow reformed at the narrow ends of the islands. The orientation of the islands therefore indicates the direction of flow, with the narrow end of the island pointing downstream. In this case, the flow is from the lower right to upper left. Ares Vallis is an outflow channel opening on to the Chryse Planitia, landing site of *Mars Pathfinder* (see Figs. 2.14 and 2.15). Ares is the Greek god of war, and possibly connected with the Roman god Mars. (Courtesy of NASA/JPL/ASU.)

to gouge out these river-like outflow channels is enormous, requiring catastrophic floods containing million of tons, or billions of kilograms of liquid water.

The outflow channels must have drained and emptied into the vast low-lying plains in the northern hemisphere of Mars. The *Mars Exploration Rover*, *Opportunity*, has found sedimentary layers and round, mineral "blueberries" that must have formed when water flowed across the northern plains about 3.7 billion years ago. Instruments aboard orbiting spacecraft have also identified clay minerals and chloride salt deposits formed at diverse watery environments in the distant past on Mars.

As everyone knows, water flows downhill, which is toward the low-lying northern hemisphere of Mars (Fig. 2.43). Meandering river-like valleys found in the southern highlands of Mars apparently flow down from higher elevations (Fig. 2.44). This suggests that the floodwaters that cut the outflow channels also drained northward and pooled in the vast northern lowlands at the ends of the channels. When topographical maps are combined with measurements of the red planet's gravity, buried subsurface canyons are found located beneath the sediment, emanating from the visible outflow channels. The transport of water therefore continued far into some parts of the northern plains. It has even been argued that the northern lowlands were once the sites of an ancient ocean, covering up to one-third of the surface area of the planet and up to 1.6 kilometers deep.

Most geologists agree that vast amounts of water flowed across the Martian landscape long ago, perhaps 3 to 4 billion years in the past. In one interpretation, the red planet was warmer and wetter in its early history, and was therefore much different from the cold, arid Mars we see today. A thicker, warmer atmosphere would have permitted flowing water that generated the outflow channels and filled low-lying areas with pools and lakes of liquid water. Another explanation involves frequent comet impacts in the early history of the planet.

Water ice in the outer solar system

Dense rocky substances dominate the four terrestrial planets (Mercury, Venus, Earth and Mars) and the Earth's Moon, which are nearest the Sun, while the lighter gaseous and icy substances dominate the outer giant planets (Jupiter, Saturn, Uranus and Neptune) and their moons. These compositional differences appear to result from the fact that the terrestrial planets formed close to the hot, bright, young Sun, and they suggest that water ice might be common in the colder, outer parts of the planetary system. That is, the oxygen, O, and hydrogen, H, in the frigid outer precincts of the young planetary system could have combined to make water ice, or frozen H_2O.

In the inner regions of the solar nebula, the higher temperatures would vaporize water ice so it would not condense, leaving only high-density, rocky substances to coalesce and merge together to form the terrestrial planets. Further out, in the colder realm of the giant planets, large amounts of water ice could survive and not evaporate. The asteroid belt, that forms the great divide between the orbits of Mars and Jupiter, is thus known as the water line of the primeval solar system; it is the first place that a 1-kilometer chunk of ice could form.

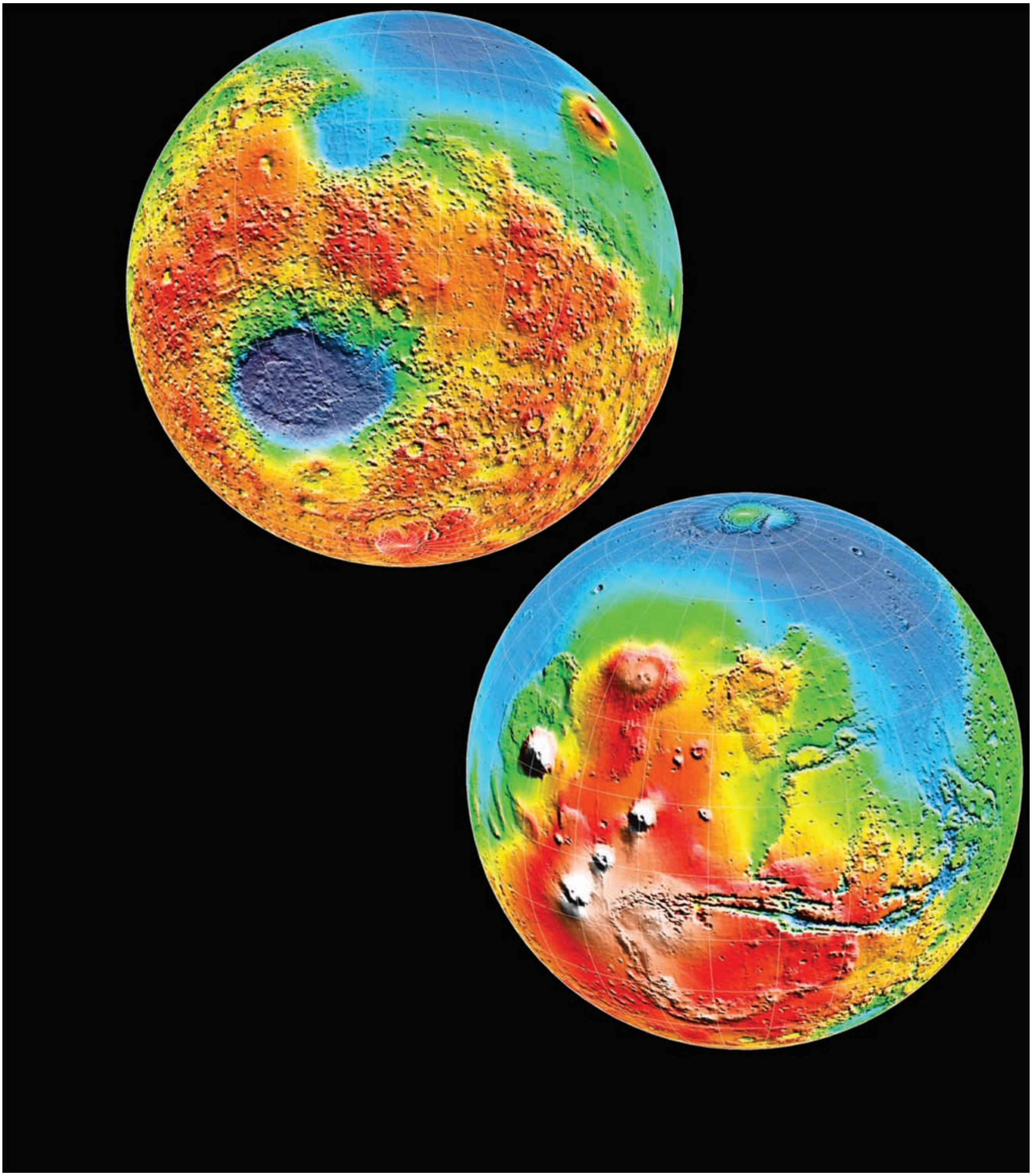


Fig. 2.43 Highs and lows of Mars' topography Global topographic maps of two sides of the planet Mars, based on detailed laser altimeter measurements from the *Mars Global Surveyor* spacecraft. In these images, color represents height above (or below) the mean planetary radius, ranging from dark blue (−8 kilometers) through green, yellow and red (+4 kilometers) to white (over 8 kilometers in altitude). Flowing water would run downhill, collecting in the low-lying blue regions including the Hellas impact basin (*upper image*), about 2800 kilometers across, the Valles Marineris (*lower image*), shown as a horizontal gash beside the Tharsis volcanoes (*pink*), and the extensive lowland plains in the north (*top of both images*). (Courtesy of NASA/JPL/ GSFC.)

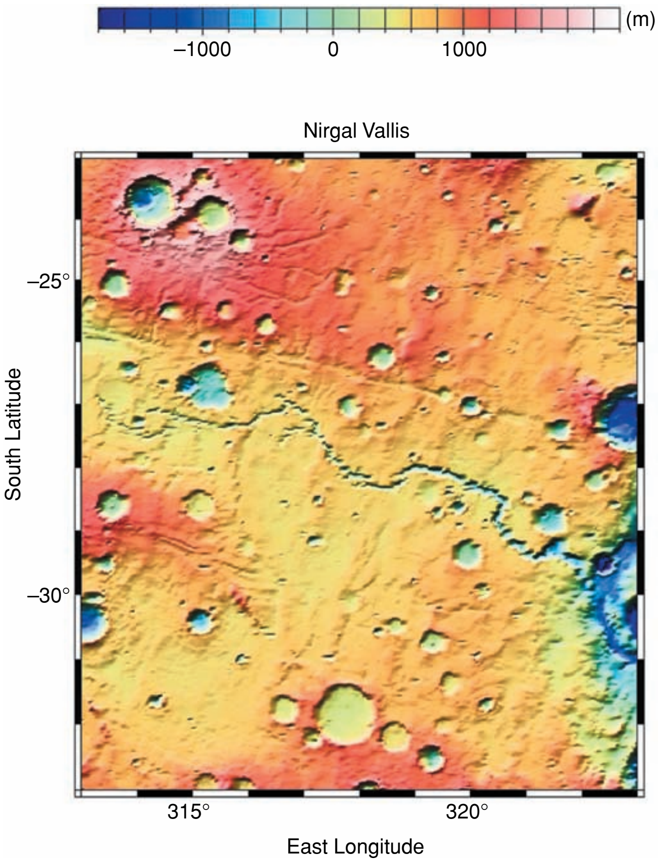


Fig. 2.44 Nirgal Vallis on Mars This valley meanders over 500 kilometers across the heavily cratered southern highlands on Mars, with few tributaries. The valley network is shown in topographic relief with a vertical accuracy of approximately 1 meter provided by a laser altimeter on the *Mars Global Surveyor*. The direction of water flow would be downhill, to the lower right. Nirgal Vallis is located at 28.4 degrees south and 42.0 degrees west longitude (*east longitude shown here*). Nirgal is the Babylonian word for Mars. (Courtesy of NASA/JPL/GSFC.)

Infrared spectroscopy from the ground and space have shown that the rings of Saturn are composed of pure water ice, with no trace of dirt, dust or rock. And the smooth, cold, highly reflective surfaces of Jupiter's moon Europa and Saturn's satellite Enceladus are now known to consist of water ice, encasing great balls of liquid water. After all, ice freezes from the top down, so you might expect Europa and Enceladus to be cold and desolate on the outside, with a warm ocean on the inside.

Images from the *Galileo* spacecraft indicate that Europa almost certainly had a liquid ocean at one time, and they have considerably strengthened the evidence for an ocean of liquid water existing just beneath its icy surface at the present time. The *Galileo* images show evidence for near-surface melting and movements of large blocks of ice, the first icebergs found off planet Earth. In some cases, large ice rafts the size of cities have broken off and drifted apart, sliding away from each other, with edges that fit like the pieces of a jigsaw puzzle. Warm ice or even liquid water must lubricate the moving ice from below.

There is indirect magnetic evidence of a hidden ocean beneath the smooth, icy surface of Europa. Magnetic measurements from *Galileo* indicate that the moon generates internal electrical currents as it sweeps through Jupiter's powerful magnetic field, and these electrical currents generate a temporary magnetic field that briefly alters Jupiter's field near the satellite. For electrical currents to flow in Europa, some part of the satellite must conduct electricity. Ice is not a good conductor, but salty water is. A saltwater ocean about 10 kilometers below the surface can produce the measured changes of Jupiter's magnetic field as it sweeps by in different orientations to the satellite.

What keeps Europa's subsurface ocean from freezing solid? Europa is probably kept warm inside by the gravitational tugging and flexing it experiences when moving toward and away from massive Jupiter in the course of the satellite's elliptical orbital motion. A similar effect melts the inside of rocky Io, producing its ubiquitous volcanoes. The orbit of Europa is pulled slightly out of round by the gravitational action of Io, which is closer to Jupiter, and Ganymede, the next satellite out. Jupiter's varying gravitational pull as Europa moves along its eccentric orbit produces tides of different size, causing the satellite to stretch and distort, heating its interior and keeping the water liquid beneath its icy crust. These internal tides will also cause Europa's overlying ice shell to flex, producing cracks that open and close as Jupiter squeezes the moon in and out.

Jupiter's satellite Ganymede, the largest moon in the solar system, probably contains substantial amounts of water inside, which would help account for its low mean mass density of just 1940 kilograms per cubic meter. Its surface has large dark plates separated by lighter regions, and impact craters that are surrounded by bright material (Fig. 2.45). The dark regions are believed to be part of the original crust of Ganymede, which probably cracked and spread apart. The lighter regions are most likely water ice that has moved in, replacing about half of the ancient, dark surface. The brilliant white material that surrounds some craters is probably clean water ice that splashed out from inside the satellite.

Like Europa, Saturn's moon Enceladus is also an ice world being squeezed inside as it moves along an eccentric orbit. The tidal flexing caused by the varying gravitational pull of nearby massive Saturn apparently heats Enceladus enough to maintain a liquid ocean beneath its

Fig. 2.45 Jupiter's satellite Ganymede Large dark blocks are frozen within the icy surface of Ganymede. They are believed to be part of the original crust of the satellite, resembling frozen-over continents floating on a background of translucent ice. The brilliant white material that surrounds some craters is probably clean water ice or bright snow that was splashed out from inside the satellite. The enhanced color of this *Galileo* image of Ganymede, taken on 29 March 1998, also reveals the two predominant terrain features on the moon: bright grooved terrain and older, dark furrowed areas. The violet hues at the poles may be the result of small particles of frost. (Courtesy of NASA/JPL/DLR.)

ice. As we have previously mentioned, geysers of water ice, water vapor and other substances spray out of fractures in its surface.

Substantial amounts of water ice reside well beyond all the major planets. This is where comets hibernate for billions of years, in the distant Kuiper belt and even more remote Oort cloud (Chapter 14). When one of these comets is deflected into the inner solar system and comes close to the Sun, it can expel about 25 000 kilograms of water every second, propelled by water vapor generated when the Sun's heat melts the water ice. A typical comet contains one million billion (10^{15}) kilograms of water.

3 Atmospheres, magnetospheres and the solar wind

- An atmosphere is a gaseous layer of molecules, with smaller amounts of atoms and ions, which surrounds a planet or natural satellite, held near them by their gravity.
- The present-day atmospheres of the terrestrial planets are thought to be secondary, having originated after planetary formation about 4.6 billion years ago. The atmospheres may have been released from volcanoes or acquired during collisions of comets and asteroids.
- The giant planets retain their primeval atmospheres, created when these planets formed, capturing significant amounts of hydrogen and helium gas.
- An atmosphere is characterized by the pressure and temperature of the molecules in it.
- Pressure increases with the temperature and density of the gas.
- An atmosphere allows the warmth of sunlight in but prevents the escape of infrared heat radiation from the planet's surface; this global warming by heat-trapping gases in an atmosphere is now known as the greenhouse effect.
- The ability of a planet or satellite to retain an atmosphere depends on both the temperature of that atmosphere and the gravitational pull of the planet or satellite. If the gas is hot, the molecules move about with a greater velocity and are more likely to escape the gravitational pull of the planet. A planet with a larger mass is more likely to retain an atmosphere.
- Only the massive giant planets, like Jupiter and Saturn, have a high enough escape velocity that they can retain all atoms and molecules, including the lightest element, hydrogen.
- Not all molecules move at the same average speed. Some of them move faster than the average speed and others move slower, with speeds described by the Maxwellian distribution.
- The lighter, high-velocity molecules can slowly leak out or evaporate from the top of an atmosphere where collisions no longer dominate the velocity distribution; this process is called Jeans escape or thermal evaporation.
- A thin membrane of air protects, ventilates and incubates us.
- Winds move air from hot to cold regions, in an attempt to equalize the temperature differences.

- The surface of Venus now lies under a hot and heavy atmosphere of carbon dioxide; its greenhouse effect has raised the surface temperature on Venus to a torrid 735 kelvin, hot enough to melt lead and zinc.
- High-velocity winds on Venus whip its highest clouds around the planet 60 times faster than the planet rotates.
- Mars now has an exceedingly thin, dry and cold atmosphere of carbon dioxide, with less than one-hundredth the surface pressure of the Earth's atmosphere, but global winds can stir up enough dust to completely cover Mars.
- The red planet breathes about one-third of its atmosphere in and out as its southern polar cap grows and shrinks with the Martian seasons.
- Mercury and the Earth's Moon are surrounded by a tenuous, varying mist of atoms that is continuously escaping into the surrounding space and being replenished from below. It is essentially a layer of exit, or escape, known as an exosphere.
- The atmospheres of Earth, Venus and Mars have evolved to their present states as the result of their varying distances from the Sun, a runaway greenhouse effect on Venus, and the development of life on Earth.
- The Sun generated so little heat more than 2 billion years ago that the Earth's oceans should have been frozen solid. This faint-young-Sun paradox could be resolved if a thick carbon-dioxide atmosphere warmed the planet back then or if the young Sun was more magnetically active or more massive than it is now.
- Like the Sun, the most abundant element in the giant planets is the lightest element, hydrogen, and the next most abundant element is helium; but Uranus and Neptune have lesser amounts of these two gases and relatively greater amounts of the heavier hydrogen compounds like methane, ammonia and water.
- Jupiter is all atmosphere or liquid, with no solid surface to rub against or continents to disturb the flow. Its Great Red Spot has existed for more than 300 years; the location and speed of powerful winds and some violent storms on the giant planet have remained unchanged for at least one century.
- Helium rain has been falling toward the center of Saturn for the past 2 billion years, significantly depleting the amount of helium in the planet's upper atmosphere.
- Saturn's largest moon, Titan, has a substantial Earth-like atmosphere, which is mainly composed of nitrogen and has a surface pressure comparable to that of the Earth's atmosphere. An opaque smog-like haze hides the surface of Titan from view.
- A thin film of oxygen envelops Jupiter's large moon Europa, while a tenuous mist of sulfur dioxide surrounds Jupiter's moon Io.
- An energy-laden, electrically charged solar wind blows out from the Sun in all directions and never stops, carrying with it a magnetic field rooted in the star.
- The radial, supersonic outflow of the solar wind creates a huge bubble of electrons, protons and magnetic fields, with the Sun at the center and the planets inside, called the heliosphere.
- The Earth has a dipolar magnetic field, amplified and sustained by dynamo action in its liquid core.
- Ancient magnetic rocks indicate that Earth's magnetic poles keep switching places every 10 thousand to 10 million years, and that our planet's magnetic field may now be heading for a flip.

- The terrestrial magnetic field deflects the solar wind, hollowing out a cavity called the magnetosphere. The magnetosphere of any planet is the volume of space from which the main thrust of the solar wind is excluded.
- Energetic electrons and protons have penetrated the Earth's magnetic defense. Some of them are confined within two doughnut-shaped radiation belts that encircle the Earth's equator but do not touch it.
- The magnetosphere contains particles from the Sun that arrive via the solar wind and penetrate the Earth's magnetic defense through a temporary opening in it, which is produced by magnetic reconnection when the solar magnetic field and the Earth's magnetic field point in opposite directions at the place where they touch.
- A plasmasphere is found in the inner part of the Earth's magnetosphere. This is located just outside the upper terrestrial atmosphere, which is called the ionosphere. The plasmasphere contains oxygen ions, protons and electrons derived from the ionosphere.
- Cosmic rays produce neutrons in the Earth's atmosphere, and a small fraction of these neutrons move out into the Earth's inner radiation belt before they disintegrate, producing electrons and protons in places they could not otherwise have reached.
- Of the eight major planets, six are known to generate detectable, global magnetic fields; only Mars and Venus do not now have such a dipolar magnetic field.
- Jupiter's magnetosphere is the largest enduring structure in the solar system, more than 10 times larger than the Sun. It was discovered when Earth-based radio telescopes unexpectedly detected the synchrotron radio emission of high-speed electrons trapped in the giant planet's immense magnetic field.
- The magnetic fields of Uranus and Neptune are offset by large amounts from their centers, and tilted by enormous angles from their rotation axes.
- The Earth's aurora is a spectacular multi-colored light-show.
- When viewed from space, the aurora forms an oval centered on the magnetic poles of the Earth; similar aurora ovals have been detected in ultraviolet light at both the north and south magnetic poles of Jupiter and Saturn.

3.1 Fundamentals

What is an atmosphere?

An atmosphere is a gaseous layer of molecules, with smaller amounts of atoms and ions, which surrounds a planet or natural satellite, held near them by their gravity. Since gas has a natural tendency to expand into space, only bodies that have a sufficiently strong gravitational pull can retain atmospheres. So the ability of a planet or satellite to retain an atmosphere depends on its mass; but it also depends on the mass of the gas particles as well as the gas temperature, determined by both the planet's distance from the Sun and the atmosphere's greenhouse effect.

Where do atmospheres come from?

The present-day atmospheres of the terrestrial planets are thought to be secondary, having come from other sources after planetary formation about 4.6 billion years ago. They can acquire atmospheric gas by releasing vapors from within the planet or by capturing volatile materials from comets or asteroids when they strike the planet.

An important source of the atmospheres of the terrestrial planets is the volcanic release of gases trapped inside their hot interiors. Volcanoes can supply water vapor H_2O, carbon dioxide CO_2, nitrogen N_2, and sulfur-bearing gases H_2S and SO_2, where H denotes a hydrogen atom, O an oxygen atom, C a carbon atom, N a nitrogen atom, and S a sulfur atom.

Another source of the terrestrial atmospheres is comets and asteroids. Because these objects formed further from the Sun's heat, they could retain water and carbon dioxide. Early in the history of the solar system, many more comets and asteroids were in orbits that intersected the orbits of Mars, Earth and Venus. The collisions would have released ices and gases, supplying these planets with the volatile substances needed to form their early atmospheres and oceans.

Initially these atmospheres were probably dominated by carbon dioxide and water vapor, and some of these gases could condense to become surface liquids or ices. The Earth's oceans and polar caps probably originated from water vapor that condensed as rain or snow. The polar caps of Mars contain frozen carbon dioxide, and vast amounts of water ice are frozen into its surface. Water vapor or carbon dioxide gas can be returned to these atmospheres by evaporation of surface liquids or sublimation of surface ice into gas.

In contrast to the terrestrial planets, the giant planets retain their primeval atmospheres, created when these planets coalesced in the cold outer precincts of the planetary system. They were large and massive enough to capture significant amounts of hydrogen and helium gas, and far enough from the Sun to be unaffected by its heat and winds.

Pressure and temperature

An atmosphere is characterized by the pressure and temperature of the molecules, which vary as a function of height within the atmosphere. Collisions between molecules in an atmosphere create the pressure, and collisions occur more frequently when a gas is hotter or when the gas has a greater density. The pressure, P, of a gas with molecules of number density, n, and temperature, T, is given by the perfect gas law $P = nkT$, where Boltzmann's constant $k = 1.38 \times 10^{-23}$ joule per kelvin, and the number density $n = \rho/m$ for a gas of mean mass density ρ and a mean molecular mass m. So pressure increases with the temperature and density of the gas.

For comparison purposes, the pressure and temperature are specified at the surfaces of the terrestrial planets and at the cloud tops of the giant planets. The atmospheric pressure is usually measured in a unit called the bar, as in barometer, where one bar is roughly equal to the Earth's atmospheric pressure at sea level. A thick atmosphere has a relatively large surface pressure, while a thin atmosphere produces comparatively little surface pressure.

Solar radiation warms a planet's atmosphere, and as we would expect, the heat is greatest for objects that are closest to the Sun. That is because the intensity of sunlight falls off as the inverse square of distance from the Sun.

We can make an initial estimate for the temperature of the surface of a terrestrial planet, or the cloud tops of a giant planet, by assuming that the surface or cloud tops are not noticeably warmed by heat rising from the planet's interior and that there is no atmosphere above them. The planet is then heated solely by the Sun's radiation, and we can calculate the planet's effective temperature, T_{eff}, from the relation $T_{\text{eff}} = 279/\sqrt{D}$ kelvin, where D is the planet's distance from the Sun in AU (Focus 3.1).

The effective temperatures of the planets are compared to the mean observed surface or cloud-top temperatures in Table 3.1. The surface of Venus is much hotter than expected, and the surface of the Earth is somewhat hotter, both a consequence of the greenhouse effect. The giant planets are also hotter than expected, owing to internal heat left over from their formation or to helium raining down inside them.

Global warming by the greenhouse effect

The surface temperature of a terrestrial planet can increase when its atmosphere traps heat near the surface, warming it to a higher temperature than would be achieved by the Sun's radiation in the absence of an atmosphere. Incoming sunlight is partly reflected by clouds, but the rest passes through the atmosphere to warm the planet's surface. Much of the surface heat is re-radiated in the form of long infrared waves that are absorbed by atmospheric molecules such as carbon dioxide or water vapor. Some of the trapped heat is re-radiated downward to warm the planet's surface and the air immediately above it. The atmosphere thus acts as a one-way filter, allowing the warmth of sunlight in, and holding it close to the planet's surface and elevating the temperature there.

The idea that this atmospheric blanket might warm the Earth was suggested in 1827 by the French mathematician Jean-Baptiste Fourier (1768–1830) and developed by the Irish scientist John Tyndall (1820–1893) in the 1860s. Fourier wondered how the Sun's heat could be retained to keep the Earth hot, concluding that sunlight passes through the atmosphere, which also prevents the escape of heat from the planet's surface.

Global warming by heat-trapping gases in the air is now known as the greenhouse effect, but this is a misnomer. The air inside a garden greenhouse is heated because it is enclosed, preventing the circulation of air currents that would carry away heat and cool the interior. Nevertheless, the term is now so common that we continue to use it to designate the process by which an atmosphere traps heat near a planet's surface.

Focus 3.1 How hot is a planet?

The radiant energy per unit time that a planet of radius R_P receives from the Sun is:

$$\text{Solar energy received} = \pi \ R_P^2 \ f = \pi \ R_P^2 \frac{\sigma \ R_\odot^2 \ T_\odot^4}{D_P^2} \ \text{joule per second}$$

where R_P is the radius of the planet, f is the total amount of radiant solar energy per unit time per unit area reaching the top of the planet's atmosphere, D_P is the planet's distance from the Sun, the Stefan–Boltzmann constant σ is 5.670×10^{-8} J m^{-2} K^{-4} s^{-1}, the solar radius $R_\odot$ is 6.96×10^6 meters, and the effective temperature of the visible solar disk is $T_\odot = 5800$ kelvin. The Stefan–Boltzmann law gives the amount of radiation energy lost per unit time from a planet of effective temperature, T_{eff}:

$$\text{Energy lost} = 4\pi\sigma \ R_P^2 T_{\text{eff}}^4 \ \text{joule per second}$$

Assuming thermal equilibrium between energy lost and received, and collecting terms, we obtain:

$$T_{\text{eff}}^4 = \frac{R_\odot^2 \ T_\odot^4}{4 \ D_P^2} \approx 1.36 \times 10^{32} \left(\frac{1}{D_P^2}\right) \approx 6.089 \times 10^7 \left(\frac{1\,\text{AU}}{D_P}\right)^2$$

or

$$T_{\text{eff}} \approx 279 \left(\frac{1\,\text{AU}}{D_P}\right)^{1/2} \ \text{kelvin}$$

where $1\,\text{AU} = 1.496 \times 10^{11}$ meters is the distance of the Earth from the Sun. Notice that the effective temperature is independent of the planet's radius, and that the effective temperature for planets around other stars depends upon the square root of the star's radius and the star's disk temperature, as well as the planet's distance from the star.

This expression assumes that all of the sunlight falling on the planet is absorbed, but some of it is always reflected. The extent to which a planet or satellite reflects light from the Sun is specified by its albedo, A, the percentage of reflected light. The visual albedo measures the fraction of incoming visible sunlight that is reflected directly into space, on a scale of 0.0 to 1.0. Rocky bodies like the Earth's Moon or Mercury absorb a lot of incident sunlight, while clouds or icy surfaces reflect it. Thus, the Moon and Mercury have a visual albedo of 0.12, while cloud-covered Venus has an albedo of 0.65, helping to make it the brightest planet in the solar system when it is visible from Earth.

Taking the albedo, A, into account, we have:

$$T_{\text{eff}} = 279(1 - A)^{1/4} \left(\frac{1\,\text{AU}}{D_P}\right)^{1/2} \ \text{kelvin}$$

There are two kinds of albedo: the Bond albedo, which measures the total proportion of electromagnetic energy reflected, and the visual geometric albedo that refers only to electromagnetic radiation in the visible spectrum. The geometric albedo of an astronomical body is the ratio of its actual brightness to that of an idealized flat, fully and isotropically reflecting disk with the same cross-sectional area. The Bond albedos for Mercury, Venus, Earth and Mars are 0.119, 0.75, 0.29 and 0.16, respectively, while their visual geometric albedos are 0.106, 0.65, 0.367 and 0.150. When the formula is applied to the Earth we obtain $T_{\text{eff}} \approx 256$ kelvin using the Bond albedo and $T_{\text{eff}} \approx 249$ kelvin using the visual geometric albedo. The Bond albedo for the Earth's Moon is 0.123, so its effective temperature would be higher, at about 270 kelvin.

Tyndall built an instrument to measure the heat-trapping properties of various gases, examining the transmission of infrared heat radiation through them. He found that the main constituents of our atmosphere – oxygen (O_2) 21 percent, and nitrogen (N_2) 77 percent – were transparent to both visible and infrared radiation. These diatomic, or two-atom, molecules are incapable of absorbing any noticeable amounts of infrared heat radiation. He also found that water vapor and carbon dioxide, which are minor ingredients of the Earth's air, absorb significant heat. As Tyndall realized, these gases are transparent to sunlight, which warms the ground, but partially opaque to the infrared heat rays, which are trapped near the surface and warm our globe. Water vapor (H_2O) and carbon dioxide (CO_2) molecules consist of three atoms and are more flexible and free to move in more ways than diatomic molecules, so they absorb the heat radiation.

Once the temperature of a planet has been established, by direct observation or from calculations of solar heating, primeval heat, and the greenhouse effect, we can determine the likely constituents of its atmosphere.

Losing an atmosphere

The escape of gases from a planetary atmosphere plays as big a role in determining its composition as the supply of gases does. One of the most important loss mechanisms is thermal escape, in which the gas gets too hot to hold

Table 3.1 Distances, visual albedos, effective temperatures, and mean temperatures of the planets[a]

Planet	Average distance, D_P (AU)	Visual geometric albedo, A	Effective temperature, T_{eff} (kelvin)	Mean temperature[b] (kelvin)
Mercury	0.387	0.106	436	440
Venus	0.723	0.65	252	730
Earth	1.000	0.367	249	281
Mars	1.524	0.150	217	210
Jupiter	5.203	0.52	102	165
Saturn	9.537	0.47	77.1	134
Uranus	19.19	0.51	53.3	76
Neptune	30.07	0.41	44.6	73

[a] Distances and mean temperatures are from http://www.jpl.nasa.gov/solar_system/planets. Effective temperatures are calculated from the visual geometric albedos, which are from http://ssd.jpl.nasa.gov.
[b] There are the mean surface temperatures for the terrestrial planets and the mean cloud-top temperatures for the giant planets.

on to. Thermal escape provides a straightforward explanation for why the large planets have atmospheres containing hydrogen and hydrogen compounds, the middle-sized planets have atmospheres containing oxygen compounds, and small objects, such as the Earth's Moon and Mercury, have no appreciable atmosphere at all.

George Johnstone Stoney (1826–1911) set forth the simple explanation for these differences in 1898. It partly depends upon the mass and gravitational pull of the object, and it also depends on the atmospheric temperature and the mass of the gas atom or molecule.

A molecule will overcome the gravitational pull of a planet or satellite if the molecule's velocity exceeds the object's escape velocity, which increases with the object's mass (Focus 3.2). Small bodies with low mass, such as our Moon, have a very small escape velocity and insufficient gravitational pull to retain any substantial atmosphere. Middle-sized planets, like the Earth, Venus and Mars, have moderate escape velocities and enough gravity to hold on to heavier, slower-moving molecules, but they are small enough and warm enough for hydrogen and helium to escape. Only the massive giant planets, like Jupiter and Saturn, have a high enough escape velocity that they can retain all molecules, including the lightest one, hydrogen.

Temperature also plays a role, for it helps determine if the molecules can move fast enough to escape an object's gravity. A planet will only retain molecules that are moving at velocities less than the planet's escape velocity, and a molecule's velocity increases with temperature (Focus 3.2). Hotter molecules dart about at faster speeds, and colder molecules move with slower speeds. Since the outer atmospheric temperature falls off with increasing distance from the Sun, molecules tend to have lower velocities out in the realm of the giant planets. At a given temperature, a molecule's velocity increases with decreasing molecular mass (Focus 3.2), so lighter molecules move at faster speeds and are more likely to escape a given planet or satellite than heavier ones.

When the thermal velocity exceeds the escape velocity for a given type of molecule, all of those molecules will promptly flow out into space, and if this happens for every type of molecule, an airless body is left behind – like Mercury, the Earth's Moon, and the four large satellites of Jupiter. For the Earth, Mars and Venus, the thermal velocity of all molecules is smaller than the escape velocity, but the lightest gases can still slowly move out into space from the top of their atmospheres.

To understand why this might happen, note that the molecules in a gas can gain or lose speed by collisions with each other, so not all molecules move at the same average speed, or at the thermal velocity. Some of them move faster and others move slower, with a velocity probability distribution published in 1866 by the Scottish physicist James Clerk Maxwell (1831–1879). His equation, known as the Maxwellian distribution, gives the fraction of gas molecules moving at a specified velocity at any given temperature. In 1871 Ludwig Boltzmann (1844–1906) derived

Focus 3.2 Thermal escape of an atmosphere

The ability of a planet or satellite to retain an atmosphere depends on both the temperature of that atmosphere and the gravitational pull of the planet or satellite. If the gas is hot, the molecules move about with a greater speed and are more likely to escape the gravitational pull of the planet. This is one of the reasons that Mercury, the closest planet to the Sun and therefore the hottest, has no atmosphere. The other reason is that Mercury has a relatively small size and mass, as far as planets go, and thus has a comparatively low gravitational pull. On the other hand, a planet with a larger mass is more likely to retain an atmosphere, which helps explain why massive Jupiter retains the lightest element, hydrogen. Jupiter is also relatively far away from the Sun's heat, so molecules in Jupiter's atmosphere move at a relatively slow speed.

An atom, ion or molecule moves about because it is hot. Its kinetic temperature, T, is used to define its thermal velocity, V_{thermal}, given by equating the thermal energy to the kinetic energy of motion:

$$\text{Thermal energy} = \frac{3}{2}kT = \frac{1}{2}mV_{\text{thermal}}^2 = \text{Kinetic energy}$$

or solving for the thermal velocity:

$$V_{\text{thermal}} = \left[\frac{3kT}{m}\right]^{1/2}$$

where Boltzmann's constant is $k = 1.380\,66 \times 10^{-23}$ joule per kelvin, and the particle's mass is denoted by m. We see right away that, at a given temperature, lighter particles move at faster speeds. Colder particles of a given mass travel at slower speed. Anything will cease to move when it reaches absolute zero on the kelvin scale of temperature.

When the kinetic energy of motion of a particle of mass m moving at velocity V is just equal to the gravitational potential energy exerted on it by a larger mass, M, we have the relation:

$$\text{Kinetic energy} = \frac{mV^2}{2} = \frac{GmM}{D} = \text{Gravitational potential energy}$$

where the Newtonian gravitational potential is $G = 6.6726 \times 10^{-11}\ \text{m}^3\ \text{kg}^{-1}\ \text{s}^{-2}$, and D is the distance between the centers of the two masses. When we solve for the velocity, we obtain:

$$V_{\text{escape}} = \left[\frac{2GM}{D}\right]^{1/2}$$

where the subscript "escape" has been added to show that the small mass must be moving faster than V_{escape} to leave the larger mass, M. This expression is independent of the value of the smaller mass, m. The escape velocities at the surfaces or cloud tops of the planets range between 4 and 60 kilometers per second.

the distribution function independently, as a result of his kinetic theory of gases. Although the function looks symmetric, it cuts off at low velocities and is enhanced in a high-velocity tail (Fig. 3.1). This means that at any instant, a tiny fraction of the molecules are moving fast enough to escape even when the average thermal velocity is less than the escape velocity.

The lighter, high-velocity molecules can slowly leak out or evaporate from the top of the atmosphere where collisions no longer dominate the velocity distribution. At lower altitudes, collisions confine the particles, but above a certain altitude known as the exobase, the atmosphere is so tenuous that gas particles hardly ever collide. Nothing stops an atom or molecule with sufficient velocity from flying away from the exobase into space.

The method of escape from the exobase is known as Jeans escape, after the British scientist James Jeans (1877–1946) who introduced it in 1916. It is also known as thermal evaporation since it is analogous to the slow evaporation of water from the ocean. On average, the molecules in the ocean water do not have enough thermal energy to escape from the liquid, but some of them acquire enough near the surface.

At and below the exobase in the atmosphere, collisions between particles drive the velocity distribution into a Maxwellian distribution, while above the exobase collisions are essentially absent and particles that have velocities greater than the escape velocity may leave the planet. The upward moving atoms in the high velocity tail of the Maxwell distribution can exit the planet, hence the name exobase.

As the lightest element, hydrogen is the one that most easily overcomes the gravity of a terrestrial planet, but first it must reach the exobase. On Earth, the exobase is located about 500 kilometers above the surface, and calculations indicate that about a billion billion billion (10^{27}) hydrogen atoms are still being lost from the Earth's exobase every second. This value is confirmed by satellite ultraviolet observations of hydrogen escaping from the Earth's upper atmosphere.

Notice that lighter particles are lost by thermal evaporation at a much faster rate than heavier ones. Even over

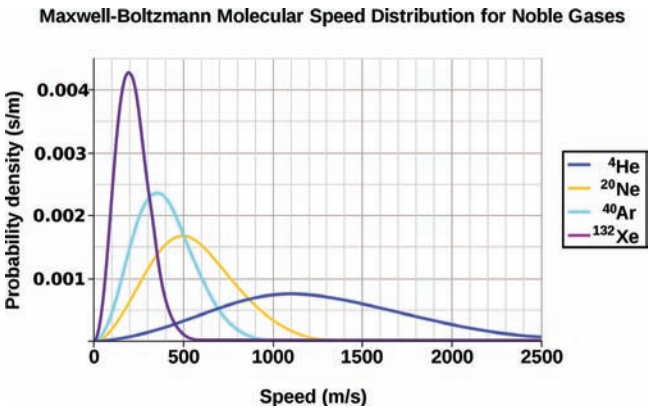


Fig. 3.1 Maxwell-Boltzmann speed distribution The probability distribution of speed density (*vertical axis*) at different speeds, υ (*horizontal axis*) for a temperature of $T = 218.85$ kelvin and four noble gases, helium (*blue*), neon (*yellow*), argon (*turquoise*) and xenon (*purple*). The peak of each curve occurs at the most probable speed, $\upsilon_P = (2kT/m)^{1/2}$, where Boltzmann's constant is $k = 1.38 \times 10^{-23}$ joule per kelvin, T is the temperature, and m is the element's mass. This distribution was first derived by the Scottish scientist James Clerk Maxwell (1831–1879) in 1866. As indicated in this plot, the peak shifts to higher speeds at lower mass provided the temperature is unchanged, because less massive elements tend to move at faster speeds. A similar change occurs at higher temperatures for a given mass. Ninety-nine percent of all molecules have speeds greater than the average molecular speed, given by $\upsilon_{avg} = (3kT/m)^{1/2}$, for every curve has a high-velocity tail which is most evident in the purple line. The equation for the distribution f(υ) is: $f(\upsilon) = 4\pi\ [m/(2\pi kT)]^{3/2}\ \upsilon^2 \exp[-m\upsilon^2/(2kT)]$.

the Earth's lifetime of 4.6 billion years, the total mass of all the hydrogen atoms lost by thermal evaporation is 2×10^{17} kilograms, and the amount lost by heavier molecules would be much less. By way of comparison the total mass of the Earth's atmosphere is about 5×10^{18} kilograms.

3.2 Atmospheres of the terrestrial planets

Earth's unique atmosphere

Our atmosphere forms an indispensable interface with nearby space, but it is often invisible. After all, you look right through the air in your room. Our atmosphere usually goes unseen on a warm, dry, windless day. Yet the slow drift of floating clouds or the sight of birds and airplanes supported by their motion proves that there is something substantial surrounding us. We can sense the touch of the wind on a stormy day, and on cold days we feel the air against our skin.

We find a further clue in the rise of smoke above a candle or a group of hawks circling above a warm meadow. Hot air rises around the flame of the candle, and the flowing air replenishes the supply of oxygen required to keep the candle burning. The hawks are getting free rides in the rising currents of hot air above ground.

When astronauts look down at the Earth at sunrise or sunset, they detect the thin atmosphere that warms and protects us, and permits us to breathe (Fig. 3.2). It is only 10 kilometers from the ground to the top of the sky, or no further than you might run in an hour. Everything beyond that thin layer of air is the black void of space. And everything below it is what it takes to sustain life.

If we were to weigh the air in a one-liter container we would find it tips the scales at slightly more than one gram. This is about one-thousandth the weight of the same amount of water. Determining its constituents is easy – just place the appropriate instrument in the air and see what is there. The major constituents of dry air on Earth are nitrogen molecules (77 percent), oxygen molecules (21 percent) that we breathe, and argon atoms (0.93 percent). Carbon dioxide is a miniscule 0.035 percent. There is almost no hydrogen in our air, and most of the hydrogen on Earth is found in water. The water vapor in wet air is variable in amount, usually no more than 1 percent.

Plants and animals respectively supply almost all of the oxygen and carbon dioxide molecules in our atmosphere, and they are continually being recycled in the photosynthesis and respiration processes. Animals breathe oxygen, and when they exhale they release carbon dioxide and water vapor. Green plants on the land and one-celled plants in the ocean water absorb carbon dioxide and water, use them in the photosynthesis of nourishment and then release oxygen into the atmosphere. This symbiotic relationship is one of the most remarkable features of life on Earth; you might call it the breath of life.

If plants did not continuously replenish the oxygen in our air, animals and humanity would exhaust the available supply in a mere 300 years. All the water on the Earth is split by photosynthesis and reconstituted by respiration every 2 million years or so. For millions of years, our ancestors have breathed the same oxygen and drank the same water, binding them temporarily in their bodies and then releasing them again to the atmosphere.

The Earth is the only place in the solar system where we can stand naked and survive. The air brings oxygen to our lungs and refreshes our bloodstream; sunlight and the "natural" greenhouse effect provide just enough heat to prevent our fluids from freezing or boiling. There is a very different situation on the other terrestrial planets, where there are no plants to supply the oxygen and the temperatures are either boiling hot or freezing cold.

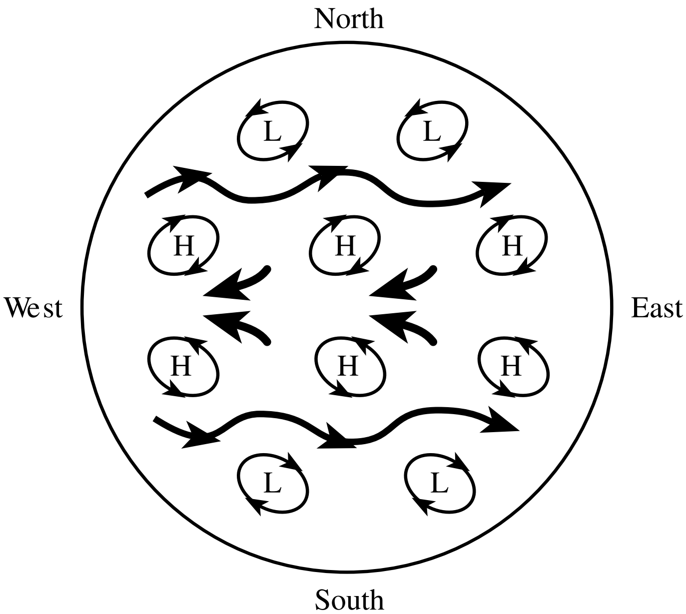
Fig. 3.2 A thin colored line The brilliant red of the setting Sun illuminates the thin atmosphere that warms and protects us. Without this atmospheric membrane we could not breathe and water would freeze. It is dust in the dense lower atmosphere that scatters red sunlight; molecules that are higher up in the air scatter blue sunlight, coloring the sky blue. This image was photographed on 3 June 2007 by an Expedition 15 crewmember on the *International Space Station*. (Courtesy of NASA.)

Earth's climate and weather

Because the Earth is a sphere, there is an unequal distribution of the Sun's heat at the terrestrial surface, producing the climate differences that starkly distinguish one part of the globe from another. The equator, for example, receives more sunlight than the poles, so the equatorial regions, known as the tropics, are the hottest places in the world and the poles are the coldest. The extra warmth at tropical latitudes evaporates seawater and causes the air to expand and rise, carrying freshwater moisture with it. As it rises, the tropical air cools and forms clouds of water ice. These clouds are moved over great distances by winds before condensing again to liquid water and falling to Earth as rain. When arriving on land, the water refreshes lakes and streams, and most of it eventually finds its way back to the sea.

The first to suggest a continuous, global circulation of the atmosphere was the English astronomer Edmond Halley (1646–1742), best known today for the comet that bears his name. Halley reasoned that the high-temperature air in equatorial regions would circulate toward the colder poles, and that the colder air from the north would move away from the poles to replace the warm tropical air. This movement of air in response to unequal temperatures is known as the wind, and the winds are blowing in an attempt to equalize global temperature differences. Since an increase in temperature produces higher pressure, the winds are also attempting to balance pressure differences, made unequal by different amounts of solar heating in various places.

Air currents tend to circulate from the tropics toward the poles and back, in the north–south direction, but they are deflected in the east–west direction by the Earth's rotation. Along the equator the near-surface air currents converge to form the trade winds, that blow mainly from east to west, almost every day of the year (Fig. 3.3). At mid-latitudes, the near-surface winds blow largely from the west and so are called westerlies. The high-altitude jet streams also blow eastward, at speeds of up to 40 meters per second, in a sinuous path that resembles the meandering of a river (Fig. 3.3).



pole
polar easterlies pola
r easterlies polar easterlies polar easterli
es polar easterlies polar easterlies polar easter

roaring forties roaring forties roaring forties roarin
forties westerlies roaring forties roaring forties roar
roaring forties roaring forties roaring forties roaring forties

horse latitudes horse latitudes horse latitudes horse latitudes

the trades the trades the trades the trades the trades the trades the
trades the northeast trades the trades the trades the trades the trades
the trades the trades the trades the trades the trades the trades the tr
ades the trades the trades the northeast trades the trades the trades t

doldrums doldrums doldrums doldrums doldrums doldrums doldrums doldr

the trades the trades the trades the trades the trades the trades t
rades the southeast trades the trades the trades the trades the trades
the trades the trades the trades the trades the trades the trades the
trades the trades the southeast trades the trades the trade

horse latitudes horse latitudes horse latitudes horse latitud

roaring forties roaring forties roaring forties roaring fo
rties westerlies roaring forties roaring forties roaring forties roar
ing forties roaring forties roaring forties roaring f

polar easterlies polar easterlies polar easterli
es polar easterlies polar easterlies pol
ar easterlies polar easterlies pol
pole

Fig. 3.3 Earth's weather patterns Trade winds blow from east to west along the Earth's equatorial regions, in the same direction that the planet rotates. At higher latitudes, there are high-altitude jet streams that move at speeds of up to 40 meters per second in the opposite direction to the trade winds. In the northern hemisphere there are high-pressure cyclones, denoted by H, and low-pressure anti-cyclones, L, rotating in the clockwise and counter-clockwise directions, respectively; in the southern hemisphere the cyclones and anti-cyclones rotate in the opposite direction. The prevailing winds are given in Annie Dillard's poem, "the windy planet".

The atmosphere's attempt to overcome the Sun's unequal heating of the Earth is never ending. The equator is always hotter than the poles, and the Earth is always spinning. The temperatures and pressures are never balanced, and the winds always blow.

Moreover, the winds do not blow in a straight line. The Earth's rotation creates vast swirling eddies in the air, just like those in a river or stream. One type of eddy is known as a *cyclone*, from the Greek word for "wheel". It is a vast whirling mass of wind and precipitation.

A cyclone in the northern hemisphere is a low-pressure cell of air rotating counter-clockwise. An anticyclone is a high-pressure cell rotating clockwise in the northern hemisphere (Fig. 3.3). Between the trade winds and the westerlies, the pattern is primarily a series of high-pressure cyclones rotating clockwise in the north and counter-clockwise in the south. In the polar regions, the weather pattern is mainly a series of low-pressure cyclones rotating counter-clockwise in the north and clockwise in the south (Fig. 3.3).

As the cyclones pinwheel across the globe, they produce most of our stormy weather. Thunderstorms, blizzards and tornadoes are examples. They tend to move from west to east across the North American continent in an almost steady parade. Cyclones also occasionally develop into hurricanes (Atlantic Ocean and Caribbean Sea) and typhoons (Pacific Ocean and China Sea).

Venus' atmosphere

In many ways, Venus is Earth's twin sister, with almost the same weight and waistline. Her mass is 81 percent that of Earth, and her radius 95 percent, so the feel of gravity at the planet's surface is similar to that on Earth. The two planets also have nearly equivalent mean mass densities, of 5244 and 5520 kilograms per cubic meter respectively, indicating that their bulk composition must be nearly the same. Venus is also just a little closer to the Sun than the Earth, orbiting our star at a distance of 0.723 AU compared with the Earth's 1.000 AU. All of these similarities gave rise to the idea that the atmosphere of Venus might resemble the Earth's air, but with a more temperate climate due to its closer distance from the Sun.

About half a century ago, many astronomers believed that the surface of Venus was warm and wet, perhaps with steamy swamps and jungles and even living creatures. Then, in the late 1950s, Earthbound radio astronomers looked beneath the clouds and discovered the extremely hot and inhospitable surface of Venus. Cloudy atmospheres are transparent to the radio waves, so the radio radiation from Venus comes directly from its surface, and can be used to take its temperature. Beneath her gleaming clouds, the planet is an inferno with a temperature hot enough to melt lead or zinc.

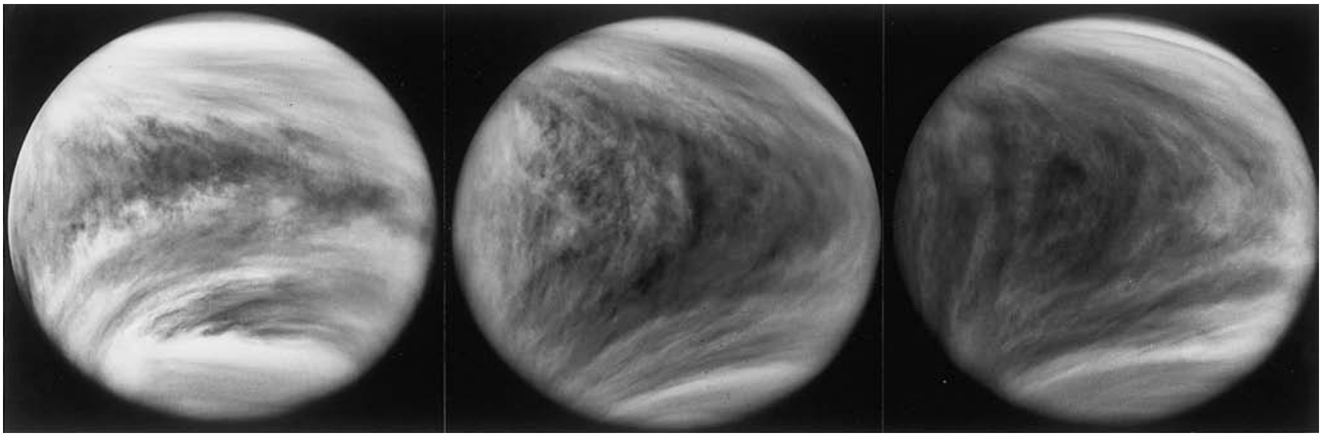


Fig. 3.4 Raging winds on Venus A high-velocity wind whips the upper layer of Venus's cloud deck around the planet's equator once every four Earth days, moving at speeds of up to 100 meters per second. These photographs taken at ultraviolet wavelengths on consecutive days by the *Pioneer Venus* spacecraft in 1979 illustrate this. The Y-shaped clouds move towards the west (*left*). A zonal east-to-west circulation dominates the winds of Venus. The low atmosphere and the planet's surface also rotate westward, but with the much slower period of 243 Earth days. (Courtesy of NASA/Larry Travis.)

Our knowledge of this torrid world was further enhanced when spacecraft directly measured the atmosphere. An entry probe parachuted from the Russian *Venera 7* spacecraft in 1970 transmitted measurements of the temperature and pressure all the way down to the bottom of the atmosphere, where the temperature reaches a sizzling 735 kelvin. Down there the thick, heavy atmosphere produces a pressure of 92 bars – that is, 92 times the air pressure at sea level on Earth. So the atmosphere on Venus is very hot and extremely heavy.

The principal constituent of the thick atmosphere on Venus is carbon dioxide. It was first identified in the upper atmosphere of Venus by infrared absorption lines in the planet's spectrum, obtained in 1932 using modest ground-based telescopes when infrared-sensitive photographic plates had been developed, and confirmed during the *Mariner 2* flyby in 1962. The fact that carbon dioxide is the major atmospheric gas on Venus was firmly established in 1967 and 1969, when three Russian space probes, *Veneras 4*, *5* and *6*, descended by parachute into the atmosphere, and obtained direct measurements of its principal constituents, showing that it consists of 96 percent carbon dioxide.

Strong winds are blowing the highest clouds around Venus at speeds of up to 100 meters per second, racing around the planet's equator in the east–west direction once every four Earth days (Fig. 3.4). Curiously enough, Venus' surface rotates in the same westward direction but with a much longer period of 243 Earth days. So the winds blow the outer atmosphere around the planet much more rapidly than the planet spins. In comparison, terrestrial jet streams move at up to half the speed of the high-flying clouds on Venus, but they are limited to narrow zones high in the Earth's atmosphere.

Near the surface of Venus, at the bottom of its massive atmosphere, the rapid winds have disappeared, and the atmosphere has become as sluggish and turgid as water at the bottom of Earth's oceans. Most of the atmosphere beneath the high-flying clouds probably rotates synchronously with the surface, just as most of the Earth's atmosphere does.

Mars' atmosphere

Mars, fourth planet from the Sun, was also long thought to resemble the Earth. The length of the day on Mars is only 37 minutes longer than our own, the rotational axes of both Earth and Mars are now tilted by about the same amount, and the Martian year is 687 Earth days, or almost two Earth years. Both planets have four seasons – autumn, winter, spring and summer – although the Martian seasons are nearly twice as long. Mars has white polar caps that wax and wane with the seasons. Their alternate growth and recession meant that gases were being extracted from, and released into, an atmosphere. White clouds are also found on Mars, resembling those on Earth, and clouds are not possible without an atmosphere.

Thus, Mars has an atmosphere, with clouds, polar caps and seasons, and these Earth-like qualities led astronomers to speculate that its atmosphere resembles Earth's. The red planet is nevertheless twice as far away from the Sun as the Earth, so it will be warmed less by solar radiation and ought to be colder – if there is no pronounced greenhouse effect on Mars.

Table 3.2 Percentage composition, surface pressures and surface temperatures of the atmospheres of Venus, Mars and Earth

	Venus	Mars	Earth
Carbon dioxide, CO_2 (%)	96.5	95	0.035
Nitrogen, N_2 (%)	3.5	2.7	77
Argon, Ar (%)	0.007	1.6	0.93
Water vapor, H_2O (%)	0.010	0.03 (variable)	1 (variable)
Oxygen, O_2 (%)	0.003	0.13	21
Surface pressure (bars)	92	0.007 to 0.010	1.0 (at sea level)
Surface temperature (kelvin)	735	183 to 268	288 to 293

The composition and extent of the Martian atmosphere wasn't understood until the Space Age. When the *Mariner 4* spacecraft passed behind Mars in 1964, its radio signal was sent through the Martian atmosphere, and a surface pressure of about 0.005 bars, or two-hundredths that of Earth, was inferred from the altered signal. So the atmosphere on Mars is exceedingly thin, with a surface pressure comparable to the pressure high in the Earth's rarefied stratosphere. In addition, by combining the spacecraft measurements of surface pressure on Mars with ground-based spectra of the planet, astronomers quickly concluded that carbon dioxide is the main ingredient of the Martian atmosphere. The exact chemical composition of the atmosphere on Mars was determined by direct measurements in 1976 when the *Viking 1* and *Viking 2* landers arrived at the surface.

The atmospheres of both Mars and Venus have the same ingredients as our air, but the proportions are different (Table 3.2). The principal atmospheric ingredient of both planets is carbon dioxide, at 96 and 95 percent respectively. Carbon dioxide accounts for only 0.035 percent of our air.

There are small, variable amounts of water vapor in the Martian atmosphere, and over the eons it has helped oxidize the planet. The water vapor has been broken down by ultraviolet sunlight into hydrogen and oxygen atoms. The light hydrogen leaked off into space and the heavier, surplus oxygen stayed to oxidize the surface rocks, turning them red. To account for its color, Mars must have lost an ocean of water equivalent to a global layer meters to tens of meters deep.

The atmospheres of Earth, Venus and Mars each have one or more gases that can saturate. That is, the atmosphere fills up with as much of the vapor as it can hold, and then that substance condenses. In the Earth's atmosphere, water vapor condenses to form billowing white clouds of water ice. The same thing happens on Mars, where clouds of water ice are found near volcanoes. Carbon dioxide also condenses out of the thin, cold Martian atmosphere into clouds and onto the surface, to form carbon-dioxide ice, also known as dry ice. On Venus, it is sulfuric acid that condenses to form the thick, unbroken layer of yellow clouds that always enshroud the planet.

The atmospheric pressure on Mars increases and decreases as its largest polar cap, the southern one, grows and shrinks, producing a seasonal change in atmospheric pressure by about 30 percent. It is as if the planet was a giant lung that slowly breathes in and exhales the same gas, carbon dioxide. When the surface temperature drops during the southern winter, the atmospheric carbon dioxide condenses and freezes to enlarge the polar cap, resulting in a drop in atmospheric pressure. In the southern summer, the ice sublimates (goes directly from solid to vapor, without becoming liquid) back into the atmosphere, increasing the atmospheric pressure.

Powerful seasonal winds are driven by temperature differences between the northern and southern hemispheres of Mars. Warm air rises in the summer hemisphere and descends in the winter one, where carbon dioxide is condensing to make the seasonal polar cap grow. The strong Martian winds also strip away light-colored dust in some areas and deposit it in others, accounting for the seasonal growth and decay of large dark areas seen from Earth.

Martian winds stir up small, localized dust storms, and occasionally coalesce to produce a globe-encircling dust storm. When substantial amounts of dust have been tossed aloft, sunlight is absorbed in the atmosphere rather than at the surface, and the storm can sustain itself by converting the Sun's energy into wind energy. The entire planet then becomes wrapped in an opaque yellow veil.

Exospheres of Mercury and the Earth's Moon

A tenuous, variable shroud of hydrogen, helium, sodium, and potassium atoms envelops both Mercury and the Earth's Moon, but always with exceedingly low number densities of less than 0.1 million million (10^{11}) atoms per cubic meter. By way of comparison, a cubic meter of Earth's atmosphere at sea level contains about 25 million billion billion (2.5×10^{25}) molecules.

The gas enveloping Mercury and our Moon is so rarified that the surface pressure is about a trillionth (10^{-12}) of the pressure at the Earth's surface. The atoms are so far apart that they almost never hit and connect with each other. They interact primarily with the surface, rather than constantly ricocheting off each other as the molecules in our air do. Because collisions are rare, each constituent has unique sources and loss mechanisms.

Such a rarefied gaseous envelope is technically known as an exosphere, and not an atmosphere, for the atoms are not permanently bound by gravity but instead escape, or exit, into surrounding space on timescales of hours to days. The neutral, or un-ionized, atoms released from Mercury are, for example, accelerated by solar radiation pressure to form an extended tail of atoms pointing away from the Sun. So the atoms in the exosphere must be continuously replenished, and they are mainly liberated from the surface by the action of sunlight, solar wind particles or micrometeorites.

Small amounts of hydrogen and helium were detected when *Mariner 10* flew past Mercury in 1973–74. In 1985–86, Earth-based telescopic observations detected sodium and potassium in Mercury's exosphere, with abundance variations from hours to years. Sodium and potassium are also found in the Moon's exosphere. Subsequent Earth-based observations of Mercury revealed the presence of calcium, and demonstrated the presence of an extended, anti-sunward tail of sodium atoms.

Observations with an instrument aboard the *MESSENGER* spacecraft, during its flybys of Mercury in 2008 and 2009, revealed the presence of magnesium, calcium and sodium atoms in the anti-sunward tail of Mercury's exosphere, with differing spatial distributions. These atoms must be coming from the planet's surface, but the different sources, as well as the detailed transfer and loss processes, remain unknown.

Evolution of the terrestrial planetary atmospheres

Nothing in the cosmos is fixed and unchanging, and nothing escapes the ravages of time. The atmospheres of the Earth, Venus and Mars are no exception, for they have been slowly altered with the addition and removal of gases as time goes on. In fact, their atmospheres probably weren't even there when the planets formed. They had too little mass to attract and hold on to the abundant hydrogen gas that was around when they accumulated. The building blocks from which the terrestrial planets formed were primarily rocky and metallic objects, and these planets may have been initially too hot to retain substantial amounts of water vapor or carbon dioxide in their early atmospheres.

The atmospheres of Earth and Venus probably began with similar compositions about 4 billion years ago, but their subsequent histories have been quite different. The Earth's atmosphere is now depleted of carbon dioxide and has excessive oxygen, while Venus has no oceans.

A massive carbon-dioxide atmosphere is responsible for the high surface temperature of Venus through the greenhouse effect. On Earth, the surface temperature is raised by about 30 kelvin by this effect, resulting in the mild climate we enjoy today. But the greenhouse effect has raised the temperature of Venus's surface by hundreds of kelvin.

The atmosphere greenhouse effect raised the temperature of young Venus and boiled away any oceans that might have condensed. The increased water vapor blocked more heat, raising the surface temperature in a runaway greenhouse effect. It turned Venus into the torrid world we see today, with a surface temperature of 735 kelvin. Because Earth is slightly further from the Sun than Venus, with a slightly lower initial temperature, the water on Earth remained liquid and it kept its oceans.

Life has been an important influence on the evolution of the Earth's atmosphere. Over the past 4 billion years, living things have caused a decrease in atmospheric carbon dioxide, while also providing an increase in atmospheric oxygen. The carbon dioxide was absorbed in ocean water where tiny marine creatures extracted it to manufacture carbonate shells. When these creatures died, their shells sank, producing carbonate sediments and rocks on the ocean floor. The result was a gradual depletion of carbon dioxide from the atmosphere.

Plant life growing in the Earth's early oceans gradually supplied oxygen to the atmosphere. Substantial amounts of oxygen began to appear in our air about 2 billion years ago, making it breathable by animals and eventually humans.

Thus, the carbon dioxide on Earth probably moved from the atmosphere to the oceans and into the rocks. Some of the carbon dioxide is returned to the air when the spreading sea floor plunges into a deep ocean trench, producing volcanoes. But there is still as much carbon dioxide remaining at the bottom of the ocean and in rocks on Earth as there is in the atmosphere of Venus, enough to exert an atmospheric pressure of 70 bars if released into our air.

The balance is a delicate one. Because Earth is slightly further away from the Sun than Venus, the Earth evolved into a living world capable of sustaining a remarkable diversity of life. If the Earth were placed in Venus's orbit, its atmosphere would get hotter and thus capable of holding more water evaporating from the ocean. The additional water vapor would trap more heat from the Sun, raising the temperature further and evaporating more water. The runaway greenhouse would eventually raise the temperature to values as high as those on Venus today. All of Earth's water would be boiled away, and the Earth would turn into a dried-out, lifeless place like Venus.

Since Mars is further away from the Sun than any other terrestrial planet, it is warmed least by the Sun. The red planet has about half the size and a tenth of the mass of Earth, so its lower gravity would be less likely to hold on to a substantial atmosphere. We might therefore expect a thin, dry and cold atmosphere similar to the one we see today on Mars. Yet, in the distant past, the Martian atmosphere might have been warmer, denser and wetter than it is today, permitting torrents of water to flow across its surface. This would account for ancient water flow that occurred on the surface of Mars 3 to 4 billion years ago.

If Mars once had a thicker atmosphere, it could have slowly evaporated into space under the combined effects of energetic sunlight and the planet's weak gravitational field; ultraviolet light from the Sun breaks up the atmospheric molecules into lighter, more energetic atoms that can escape the relatively weak gravity. More recent climatic change on Mars, during the past few million years, has been induced by the changing tilt of its rotation axis, causing atmospheric water and carbon dioxide to move into polar ice caps and back into the atmosphere again.

The Sun is also changing as time goes on, growing slowly in luminous intensity with age, a steady, inexorable brightening that is a consequence of the nuclear reactions that make it shine. As the Sun burns hydrogen into helium in its energy-generating core, the increasing amounts of helium require a rise in temperature to sustain the nuclear burning, and hence an increase in the rate of the nuclear reactions and a slow brightening of the Sun. You couldn't detect the change over all of human history, but it has profound implications over cosmic periods of time.

Stellar evolution calculations indicate that when the Sun began to shine, about 4.6 billion years ago, it was 30 percent dimmer than it is today. Assuming an unchanging atmosphere on Earth, with the same composition, greenhouse effect, and reflecting properties as today, the decreased solar luminosity would have caused the planet's global surface temperature to drop below the freezing point of water at all times earlier than 2 billion years ago. The Earth's oceans would have been frozen solid, there would be no liquid water, and the entire planet would have been locked into a global ice age something like Mars seems to be in now.

Yet sedimentary terrestrial rocks, which must have been deposited in liquid water, date from 3.8 billion years ago. There is fossil evidence in those rocks for living things at about that time. Thus for billions of years the Earth's surface temperature was not very different from today, and conditions have remained hospitable for life on Earth throughout most of the planet's history.

The discrepancy between the Earth's warm climatic record and an initially dimmer Sun has come to be known as the faint-young-Sun paradox. It can be resolved if the Earth's primitive atmosphere contained about a thousand times more carbon dioxide than it does now. Greater amounts of carbon dioxide would enable the early atmosphere to trap greater amounts of heat near the Earth's surface, warming it by an enhanced greenhouse effect. That would keep the oceans from freezing.

Over time the Sun grew brighter and hotter. The Earth could only maintain a temperate climate by turning down its greenhouse effect as the Sun turned up the heat. Our planet's rocks, oceans, and life itself may have together removed carbon dioxide from the atmosphere. Thus, the increase over time in brightness of the Sun's radiation might have been compensated by a steadily decreasing greenhouse effect, so that the surface temperature of the planet has remained about the same for the past 4 billion years. This convenient thermostat might soon be disrupted as human civilization dumps more and more carbon dioxide into the atmosphere by burning fossil fuels like coal and oil.

A different remedy of the faint-young-Sun paradox is that the Sun was more active in its youth, explosively releasing greater amounts of magnetic energy, while still radiating faint visible sunlight. Because of its faster rotation, the young Sun might have had stronger magnetic fields with enhanced extreme-ultraviolet and X-ray radiation and a greater output of high-energy particles. Or the Sun might have begun shining as a brighter, more massive star, and wasn't faint after all, subsequently losing much of its mass in strong solar winds. As with humans, the Sun's energetic youth might have evolved into a calmer old age, with moderate magnetic activity and weaker winds while shining brighter in visible sunlight.

3.3 Atmospheres of the giant planets

Composition and temperature of giant planet atmospheres

The atmospheres of the giant planets – Jupiter, Saturn, Uranus and Neptune – are very unlike those of the Earth, Mars and Venus. The elemental composition of the giant planets resembles that of the Sun, and they have been

Table 3.3 Percentage composition of the Sun and the outer atmospheres of the giant planets[a]

Molecule or atom	Sun	Jupiter	Saturn	Uranus	Neptune
Hydrogen, H_2 (molecule)	84	86.4	97	83	79
Helium, He (atom)	16	13.6	3	15	18
Water, H_2O (molecule)	0.15	(0.1)	–	–	–
Methane, CH_4 (molecule)	0.07	0.21	0.2	2	3
Ammonia, NH_3 (molecule)	0.02	0.07	0.03	–	–

[a] The percentage abundance by number of molecules for the Sun, cooled to planetary temperatures so that the elements combine to form the compounds listed, and for the outer atmospheres of the giant planets below the clouds. Dashes indicate unobserved compounds. Courtesy of Andrew P. Ingersoll.

warmed by both sunlight and internal heat retained since their formation. Their main ingredient is hydrogen, the lightest element and the most abundant element in the Sun and most stars. This helps explain the low bulk mass densities of the giant planets, between 710 and 1670 kilograms per cubic meter, which are comparable to that of the Sun at 1409 kilograms per cubic meter. Like the Sun, the next most abundant element in the giant planets is helium. The overwhelmingly abundant hydrogen (H) would also combine with the abundant carbon (C), nitrogen (N) and oxygen (O), in the low-temperature environment far from the Sun, to form stable molecules of methane (CH_4), ammonia (NH_3) and water vapor (H_2O).

The composition of the Sun, cooled to planetary temperatures, and that of the outer atmospheres of the giant planets are given in Table 3.3. Jupiter has very nearly the same composition as the Sun, made up mainly of the light gases hydrogen and helium. Saturn has about the same composition, with a bit less helium, but Uranus and Neptune have lesser amounts of hydrogen and relatively greater amounts of the heavier hydrogen compounds like methane.

Definite spectroscopic proof that molecular hydrogen is the most abundant element in Jupiter's and Saturn's upper atmosphere did not occur until the 1960s and late 1970s, when high-dispersion infrared spectroscopy, from both the ground and space, showed several weak absorption features due to molecular hydrogen (Fig. 3.5).

Helium, the second most abundant element in the Sun, has no detectable spectral features to make its presence known in the cold outer atmospheres of the giant planets. Helium is chemically inert and does not combine with other atoms to make molecules. The presence and amounts of helium atoms have nevertheless been inferred from the hydrogen infrared spectral features.

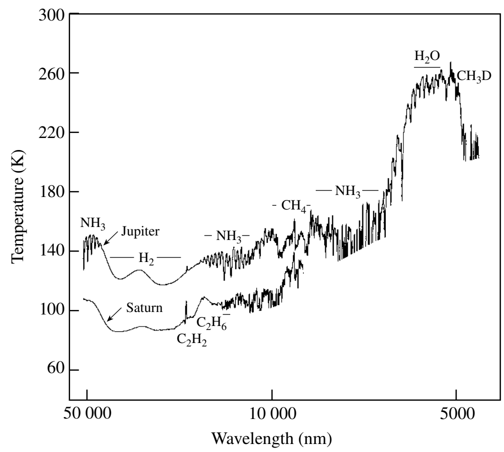


Fig. 3.5 Molecules in the atmospheres of Jupiter and Saturn The infrared radiation from the thin, cold upper atmospheres of Jupiter and Saturn exhibit numerous features that have no counterpart in the spectrum of sunlight. Strong features are seen in Jupiter for molecular hydrogen H_2, ammonia NH_3, methane CH_4, and water vapor H_2O. Saturn's outer atmosphere is also abundant in hydrogen and methane, but the ammonia features are missing and those of acetylene C_2H_2 and ethane C_2H_6 are enhanced. These spectra were taken with instruments aboard *Voyager 1* and *2* during their Jupiter and Saturn flybys in 1979 and 1980, respectively. (Courtesy of Rudolf A. Hanel.)

Collisions between helium atoms and hydrogen molecules alter the latter's ability to absorb infrared light, an effect that the *Pioneer 10* and *11* and *Voyager 1* and *2* instruments could detect. When the infrared measurements were combined with changes in these spacecraft's radio signals, observed when they passed behind the planets, the helium abundance could be determined with an uncertainty of only a few percent.

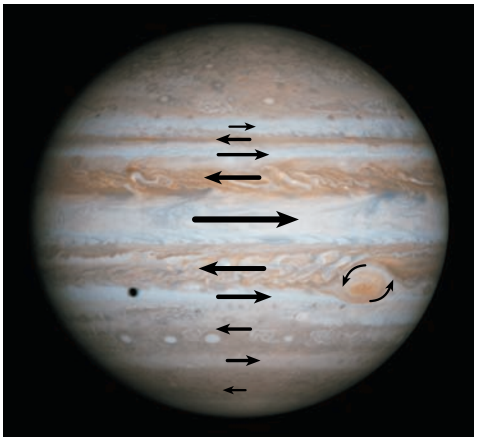


Fig. 3.6 Jupiter's counter-flowing winds The rapid rotation of Jupiter has pulled its winds into bands that flow east to west and west to east, shown in this image taken from the *Cassini* spacecraft on 7 December 2001. The windswept clouds move in alternating light-colored, high-pressure zones and dark-colored, low-pressure belts. The arrows point in the direction of wind flow, and their length corresponds to the wind velocity, which can reach 180 meters per second in the equatorial regions (see Fig. 3.7). The Great Red Spot swirls in the counter-clockwise direction (*curved arrows*), like a high-pressure anticyclone in the Earth's southern hemisphere, but it has lasted for more than 300 years, much longer than terrestrial storms. Jupiter's moon Europa casts a shadow on the planet. (Courtesy of NASA/JPL.)

The helium abundance for Uranus and Neptune is consistent with that expected from a solar composition, but helium has been significantly depleted from Saturn's upper atmosphere and somewhat reduced in Jupiter's. Theoretical calculations indicate that helium rain has been falling toward the center of Saturn for the past 2 billion years, generating heat and producing lower amounts of helium in its outer atmosphere. Helium rain must be similarly settling toward Jupiter's core, but in lesser amounts.

Raging winds on the giant planets

Astronomers have been using telescopes, both small and large, to scrutinize weather patterns on Jupiter for more than a century. They observe clouds of various ices, such as ammonia, that are formed in the cold outer atmosphere. The clouds have been pulled into counter-flowing winds, moving in opposite eastward or westward directions and remaining confined to specific latitudes. These windswept clouds move in alternating light-colored bands called zones and dark ones known as belts (Fig. 3.6). Since Jupiter is all atmosphere and liquid, with no solid surface to rub against or continents to disturb the flow, its winds are free to rage unabated in response to the planet's rotation, with large-scale configurations that have remained unchanged for as long as they have been observed.

The weather patterns on Uranus and Neptune, where clouds of methane ice are observed, more nearly resemble those on Earth, which has low-latitude trade winds that blow westward and a meandering eastward current, the jet stream, in each hemisphere. The Earth has the weakest winds in the solar system; its fastest jet streams move at speeds of about 40 meters per second. In contrast, Jupiter's winds move at constant speeds of up to 180 meters per second, and Uranus's fastest winds are just a little faster (Fig. 3.7). The high clouds on Venus and Mars also move at faster speeds than those on Earth, both with speeds of

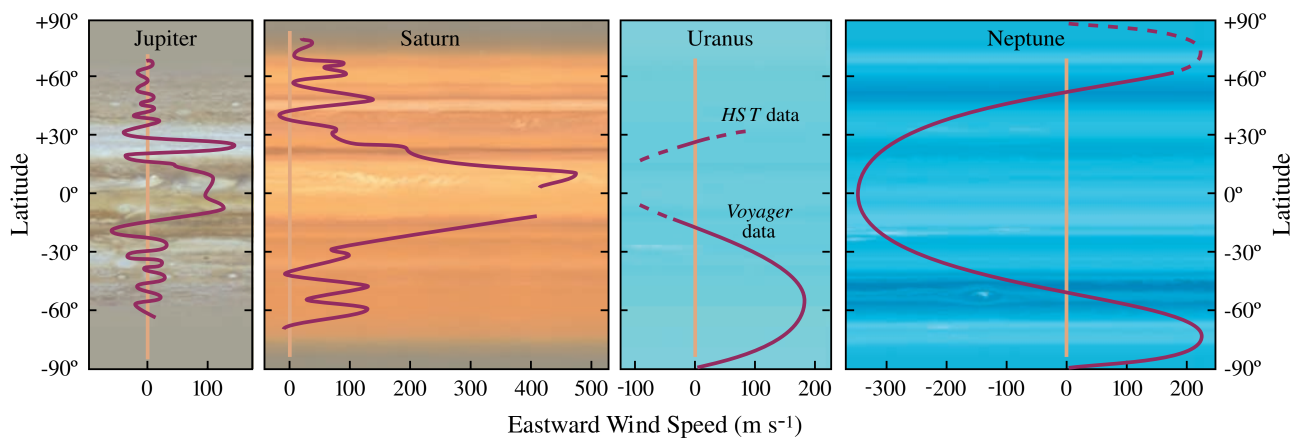


Fig. 3.7 Winds on the giant planets Variation of wind speed and direction as a function of latitude. Since the giant planets have no solid surfaces, the winds are measured relative to the internal rotation speeds; the rotation period is determined from observations of the planet's periodic radio emission. Positive velocities correspond to winds blowing in the same direction but faster than the internal rotation; negative velocities are winds moving more slowly than the rotation. The winds are faster on Saturn than any other planet. (Courtesy of Andrew P. Ingersoll.)

up to 100 meters per second. The winds on Neptune and Saturn move at speeds of up to 400 and 450 meters per second, respectively, ten times the fastest winds on Earth. The speed of the Earth's winds are measured with respect to the rapidly rotating surface beneath them; for the giant planets, which have no solid surface, the wind speed is measured with respect to an internal rotation rate inferred by using radio emission from the magnetic fields that are generated within their cores.

Despite the slow motion of the Earth's winds, the solar energy available to drive them is greater than on any planet. Because it is relatively near the Sun, the Earth intercepts more intense sunlight to power its winds than the more distant planets do. Although Venus is nearer the Sun than the Earth is, the bright clouds on Venus reflect most of the incident sunlight. The nearest planet, Mercury, is too hot to retain a substantial atmosphere.

The circulation of the winds on the giant planets is powered by solar energy, as on Earth, plus internal energy left over from their formation. Even though Jupiter, Saturn and Neptune radiate about 1.67, 1.79 and 2.7 times more energy, respectively, than they absorb from the Sun, the total power per unit area, from both sunlight and internal heat, is much less than that on Earth. The amount of power available to drive winds near the Earth's surface is 25 times that at the cloud tops of Jupiter and 400 times that in Neptune's atmosphere.

Small-scale turbulence in any planet's atmosphere dissipates the energy that is available to drive large-scale winds, and there is more dissipation for the planets that are nearer the Sun. Thus, both the global energy available to drive the winds and the amount of that energy that is dissipated decrease with increasing distance from the Sun, but at different ratios so there is more wind-producing power at the larger distances. The energy-dissipating turbulence is greatest at Earth, where the winds are weak, and least in Neptune's low-dissipation atmosphere, where the winds are stronger. Jupiter lies in between these two extremes.

Stormy weather on the giant planets

Giant anticyclones create continent-sized ovals on Jupiter that roll like ball-bearings between the oppositely directed east–west winds. The large ovals revolve between the jet streams in the anticyclonic direction – clockwise in the northern hemisphere and counter-clockwise in the southern hemisphere. The smaller eddies are soon torn apart by the counter-flowing winds, lasting only about a day or two, but the larger ones can persist for decades or centuries. Jupiter's Great Red Spot, about three times the diameter of Earth in size, has survived for more than 300 years.

How can the violent storms on Jupiter persist for months, decades and even centuries? Some external source must be feeding energy into the spinning vortex, either from the sides or from below. Perhaps each whirling spot draws energy from the shearing motion of the clouds blowing in opposite directions on its sides. Both the biggest long-lived storms and the powerful banded winds on Jupiter may be energized by smaller eddies that merge into them. It is as if the larger features devour the smaller ones, consuming them to help maintain their flow and replenish their energy.

The little, short-lived storms may receive their energy from deep within the hot interior of the planet. In response to heating, gases in lower levels of the atmosphere will expand and thereby become less dense than the gas in the overlying layers. The heated material, due to its low density, rises, just as a hot-air balloon does, cooling at the cloud tops and sinking again. Similar wheeling convective motions occur in a kettle of boiling water. They produce towering thunderclouds here on Earth; lightning is also created in the thunderstorms on both the Earth and Jupiter.

Neptune also has a stormy atmosphere with raging winds, but the big storms are not as long-lasting. The largest observed storm system on Neptune, detected when *Voyager 2* sped past the planet on 24 August 1989, is as broad as the Earth. It is called the Great Dark Spot because it resembles the Great Red Spot of Jupiter. Both storms are found in the planetary tropics, at about one-quarter of the way from the equator to the pole; both rotate counter-clockwise; and both are about the same size relative to their planet. As *Voyager 2* watched the Great Dark Spot it contracted in and stretched out with a regular rhythm, like the mouth of a feeding fish, and drifted slowly toward the planet's equator. Yet, when the *Hubble Space Telescope* was turned toward Neptune in 1995, no trace of the Great Dark Spot could be found. Perhaps it simply ran out of food to supply its energy, or perhaps it moved into strong equatorial winds that could not support it.

3.4 Titan, a satellite with a substantial atmosphere

Saturn's largest moon, Titan, is the only satellite with a substantial atmosphere. Detailed investigations with instruments aboard *Voyager 1* in 1980 showed that the dominant gas surrounding Titan is molecular nitrogen N_2, at 82 to 99 percent, similar to Earth (77 percent). In fact, the satellite is enveloped by about 10 times more nitrogen than we are, yielding a surface pressure 1.5 times greater than

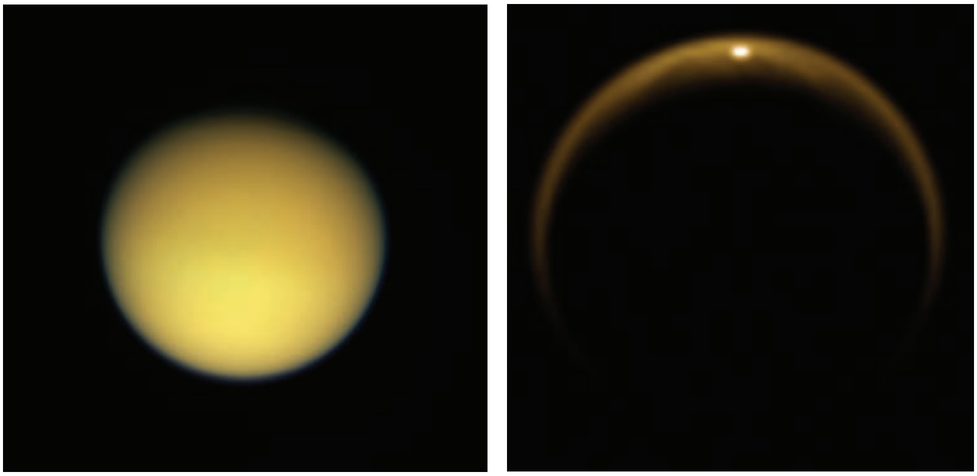


Fig. 3.8 Titan's dense, smoggy atmosphere The surface of Saturn's moon Titan is hidden from view by a hazy layer of smog, giving it a fuzzy, tennis-ball appearance in a *Cassini* natural-color view (*left*) taken on 15 February 2005. When illuminated from behind, the dense atmosphere forms a crescent. By observing radiation at infrared wavelengths, on 8 July 2009, a *Cassini* instrument was able to penetrate through the moon's atmosphere and catch the glint of sunlight reflected off a huge lake in the northern hemisphere of Titan (*right*), which is probably filled with liquid methane and ethane. The lake is named Kraken Mare. [Courtesy of NASA/JPL/SSI (*left*) and NASA/JPL/U. Arizona/DLR (*right*).]

the sea-level pressure of Earth's atmosphere. The surface temperature on Titan is 94 kelvin, as expected for a body so far from the Sun.

The spectrometers on *Voyager 1* showed that the next most abundant gas enveloping Titan is methane CH_4, with an abundance between 1 and 6 percent. Gerard Kuiper (1905–1973) had discovered signs of methane in Titan's spectrum as early as 1944.

Methane molecules rise up to high levels in Titan's atmosphere, where they are broken apart by ultraviolet sunlight and electrons coming from Saturn's magnetic environment. These molecular fragments recombine to form heavier hydrocarbon molecules such as ethane C_2H_6, and familiar gases like acetylene C_2H_2, propane C_2H_8, and hydrogen cyanide HCN.

It doesn't rain water on Titan, but it does rain methane, in large drops that fall like snow. Given the known atmospheric composition and temperatures, scientists speculated that thin clouds of methane ice crystals may form in the lower atmosphere, and that methane, ethane and propane could rain all the way down to the surface, forming seas, lakes and ponds.

Images from the *Voyager* spacecraft in 1981 and the *Cassini* spacecraft in 2005 showed that an opaque haze completely enshrouds the satellite, hiding any lakes or seas from direct view (Fig. 3.8). The smog is unimaginably worse than that over any city on Earth. Compared with any such urban smog, there are relatively few smog particles per unit volume of Titan's atmosphere, but the haze extends to an altitude of about 200 kilometers. This makes the smog thick enough to completely hide Titan's surface from view.

The orange smog must result from ultraviolet sunlight that breaks simple molecules like methane apart and chemical reactions that create more complex substances from these fragments. The exact composition of this photochemical smog remains unknown, but its mere existence suggests that heavy compounds may be falling through the atmosphere and sinking into the hypothetical seas to form an organic sludge on their floors.

As described in greater detail in Section 10.7, radar instruments aboard the *Cassini* spacecraft have been used to map out large lakes of liquid methane and ethane on Titan, and the *Huygens Probe* has been parachuted to the moon's surface, detecting river-like channels suggesting flows of liquid methane.

All of the large satellites in the solar system except Titan have tenuous atmospheres, or no atmosphere at all.

Tenuous exospheres of Europa and Io

The larger moons of the giant planets have sizes that are comparable to the Earth's Moon or Mercury, and these satellites are similarly cloaked in a thin film of gas. Their exospheres are also temporary features, and must be continuously re-supplied from the moon's surfaces.

A tenuous veil of oxygen molecules has been found around Europa, a large ice-covered moon of Jupiter. The

Focus 3.3 Why does Titan have a dense atmosphere?

Why does Saturn's largest satellite, Titan, have such a substantial atmosphere when it is only slightly bigger than the planet Mercury and almost as big as Jupiter's largest satellite, Ganymede, which have exceedingly tenuous atmospheres? The ability of a planet or satellite to retain an atmosphere is determined primarily by the body's mass and temperature, but because Titan, Ganymede and Mercury have nearly the same mass, differences in temperature should account for their atmospheric differences. Mercury is so hot that even the heaviest molecules move fast enough to escape the planet's gravity, while the temperature on Titan is so low that only the lighter molecules, like hydrogen, can escape.

But Ganymede is now sufficiently cold and massive to retain a Titan-like atmosphere. The difference between Titan and Ganymede is more likely a consequence of the temperature at the time of their birth. Titan was born in the remote cooler regions of the solar system, and nearby Saturn never became as hot as Jupiter did during its birth. The low temperatures permitted ammonia, methane and water ice to form on Titan's surface when it was born, and these ices probably sublimated to form a primeval atmosphere of ammonia and methane, while water remained locked into its surface as ice. On Ganymede, water was probably the only ice that formed in the slightly warmer climate. If there was no ammonia or methane ice on Ganymede, and if the temperatures never became high enough for its water ice to sublimate, then Ganymede would be left without any substantial atmosphere.

surface pressure of this atmosphere is barely one-hundred-billionth (10^{-7}) that of Earth's. Scientists believe that the atmospheric oxygen (O_2) on Europa is either created when energetic charged particles bombard water ice (H_2O) on the moon's surface, or ejected from it by volcanoes of ice. Exposure to sunlight and meteor impacts could also create some of the gas. The relatively lightweight hydrogen, H, escapes into space, leaving the heavier oxygen molecules to accumulate to form an atmosphere.

Jupiter's innermost large satellite Io has a varying, thin atmosphere of sulfur dioxide, SO_2, ejected from the moon's active volcanoes. The largest of Jupiter's moons, Ganymede, also has an exceedingly tenuous atmosphere. So we wonder why Saturn's moon Titan is the only large moon to have a dense atmosphere (Focus 3.3).

3.5 The planets are inside the expanding Sun

The space just outside the Earth is not empty. It is filled with pieces of the Sun. Our star is expanding in all directions, filling interplanetary space with electrically charged particles that are forever blowing from the Sun. This solar wind moves past the planets, carrying the Sun's rarefied atmosphere out to the space between the stars. So we are actually living in the outer part of the Sun.

The sharp outer edge of the visible solar disk is illusory. An invisible, rarefied atmosphere expands away from the Sun and extends all the way to the Earth and beyond. This expanding solar atmosphere is so tenuous that we can look right through it to the Sun's bright disk, just as we see through the Earth's clear air.

The outer atmosphere of the Sun, known as the corona, becomes visible to the unaided eye for only a few minutes when the Sun's light is blocked, or eclipsed, by the Moon. During such a total solar eclipse, the corona is seen at the limb, or apparent edge of the Sun, against the blackened sky as a faint halo of white light, or all the visible colors combined (Fig. 3.9).

The amazing thing about the corona is that it is incredibly hot, with a temperature of a few million kelvin. It is so intensely hot that its abundant hydrogen is torn into its component parts. Each hydrogen atom consists of one electron moving about one central, nuclear proton. The solar atmosphere therefore consists mainly of electrons and protons, with smaller amounts of heavier ions created from the less abundant elements in the Sun.

The corona is so hot that it can't stay still. The sizzling heat cannot be entirely constrained by either the Sun's inward gravitational pull or its magnetic forces. An overflow corona is therefore forever expanding in all directions, filling the solar system with a great eternal wind known as the solar wind (Focus 3.4). Interplanetary space probes have been making in-situ (Latin for "in place") measurements of the solar wind for decades, both within the space near the Earth and further out in the Earth's orbital plane.

Unlike any wind on Earth, the solar wind is a tenuous mixture of charged particles and magnetic fields streaming outward in all directions from the Sun at speeds of hundreds of kilometers per second. The seemingly eternal wind carries a magnetic field with it, with one end anchored in the Sun. This interplanetary magnetic field has a spiral shape due to the combined effects of the radial solar wind flow and the Sun's rotation.

Although the Sun is continuously blowing itself away, the outflow can continue for billions of years without significantly reducing the Sun's mass. Every second, the solar wind blows away a million tons, or a billion

Fig. 3.9 Eclipse corona streamers The million-degree solar atmosphere, known as the corona, is seen around the black disk of the Earth's Moon, photographed in white light, or all the colors combined, from atop Mauna Kea, Hawaii, during the solar eclipse on 11 July 1991. The electrically charged gas is concentrated in numerous fine rays as well as larger helmet streamers. (Courtesy of the HAO/NCAR.)

Focus 3.4 Discovery of the solar wind

The existence of the solar wind was suggested from observations of comet tails about half a century ago. When a comet is tossed into the inner solar system, the dirty ice on its surface is vaporized, sometimes forming two kinds of tails that point generally away from the Sun rather than toward it (Fig. 3.10). One is a curved dust tail, pushed away from the Sun by the pressure of sunlight. The other is a straight ion tail that is affected by the solar wind.

The German astronomer Ludwig Biermann (1907–1986) noticed, in the 1950s, that the ions in a comet's tail move with velocities many times higher than could be caused by the weak pressure of sunlight, and proposed that a continuous flow of electrically charged particles pours out of the Sun at all times and in all directions, accelerating the ions to high speeds and pushing them away from the Sun in straight ion tails.

In 1958, Eugene N. Parker (1927–) of the University of Chicago showed how such a relentless flow might work, dubbing it the solar wind. It would naturally result from the expansion of the Sun's million-degree atmosphere, the corona. He also demonstrated how a magnetic field would be pulled into interplanetary space from the rotating Sun, attaining a spiral shape.

The first direct measurements of the solar wind's corpuscular, or particle, content were made by a group of Soviet scientists led by Konstantin I. Gringauz (1918–1993), using four ion traps aboard the *Lunik 2* spacecraft launched to the Moon on 12 September 1959. In the following year, Gringauz reported that the maximum current in all four ion traps corresponded to a solar wind flux of 2 million million (2×10^{12}) ions (presumably protons) per square meter per second. This is in rough accord with all subsequent measurements.

All reasonable doubt concerning the existence of the solar wind was removed by measurements made on board NASA's *Mariner 2*, launched on 27 August 1962. Marcia Neugebauer (1932–) and Conway W. Snyder of the Jet Propulsion Laboratory used more than 100 days of *Mariner 2* data, obtained as the spacecraft traveled to Venus, to show that charged particles are continuously emanating from the Sun, for at least as long as instruments on *Mariner 2* observed them. It also unexpectedly indicated that the solar wind has a slow and a fast component. The slow one moves at a speed of 300 to 400 kilometers per second; the fast one travels at twice that speed.

The solar wind flux determined by Neugebauer and Snyder was in good agreement with the values measured with the ion traps on *Lunik 2*. The average wind ion number density was shown to be 5 million (5×10^6) protons per cubic meter near the distance of the Earth from the Sun. We now know that such a low density close to the Earth's orbit is a natural consequence of the wind's expansion into an ever-greater volume, but that variable wind components can gust with higher densities.

(10^9) kilograms. That sounds like a lot of mass loss, but it is four times less than the amount consumed every second during the thermonuclear reactions that make the Sun shine. To supply the Sun's present luminosity, hydrogen must be converted into helium within the Sun's energy-generating core, with a mass loss of about 4 million tons every second. It is carried away by the Sun's radiation, whose energy vastly exceeds that of the solar wind.

The perpetual solar wind brushes past the planets and engulfs them, carrying the Sun's corona out into interstellar space. As the corona disperses, gases welling up from below to feed the wind must replace it. Exactly where this material comes from is an important subject of contemporary space research, but the main interest for planetary research is how the solar wind affects the planets.

The reason that space looks empty is that these subatomic pieces of the Sun are very small and moving incredibly fast, and there really are not very many of them when compared even to our transparent atmosphere. By the time it reaches the Earth's orbit, the solar wind is diluted to about 5 million electrons and 5 million protons per cubic meter, a very rarefied gas (Table 3.4). By way of comparison, there are 25 million billion billion (2.5×10^{25}) molecules in every cubic meter of our air at sea level. Still, at a mean speed of about 400 kilometers per second, the flux of solar wind particles is far greater than anything else out there in space. Between one million million and ten million million (10^{12} to 10^{13}) particles in the solar wind cross every square meter of space each second (Table 3.4).

The radial, supersonic outflow creates a huge bubble of electrons, protons and magnetic fields, with the Sun at the center and the planets inside, called the heliosphere (Fig. 3.11), from *Helios*, the Greek word for the Sun. Within the heliosphere, conditions are regulated by the Sun. Its domain extends out to about 100 AU, or about 100 times the mean distance between the Earth and Sun. Out there, the solar wind has become so weakened by expansion that it can no longer repel interstellar forces (Section 15.3).

Table 3.4 Mean values of solar-wind parameters at the Earth's orbit[a]

Parameter	Mean Value
Particle density, N	$N \approx 10$ million particles per cubic meter (5 million electrons and 5 million protons)
Velocity, V	$V \approx 400$ kilometers per second and $V \approx 800$ kilometers per second
Flux, F	$F \approx 10^{12}$ to 10^{13} particles per square meter per second
Temperature, T	$T \approx 120\,000$ kelvin (protons) to $140\,000$ kelvin (electrons)
Particle thermal energy, kT	$kT \approx 2 \times 10^{-18}$ joule ≈ 12 eV
Proton kinetic energy	$0.5\, m_p V^2 \approx 10^{-16}$ joule ≈ 1000 eV $= 1$ keV
Particle thermal energy density	$NkT \approx 10^{-11}$ joule m^{-3}
Proton kinetic energy density	$0.25\, N\, m_p\, V^2 \approx 10^{-9}$ joule m^{-3}
Magnetic field strength, H	$H \approx 6 \times 10^{-9}$ tesla $= 6$ nanotesla $= 6 \times 10^{-5}$ gauss

[a] These solar-wind parameters are at the mean distance of the Earth from the Sun, or at one astronomical unit, 1 AU, where 1 AU $= 1.496 \times 10^{11}$ meters. Boltzmann's constant $k = 1.38 \times 10^{-23}$ joule per kelvin relates temperature and thermal energy. The proton mass m_p is 1.67×10^{-27} kilograms.

Fig. 3.10 Comet tails Telescopic photograph of Comet Mrkos taken in August 1957, showing the straight, well-defined ion tail and the more diffuse, slightly curved dust tail. Both comet tails point away from the Sun. The electrified solar wind deflects the charged ions and accelerates them to high velocities, creating the relatively straight ion tails. The radiation pressure of sunlight suffices to blow away the un-ionized comet dust particles, forming a broad arc that can resemble a scimitar. (Courtesy of Lick Observatory.)

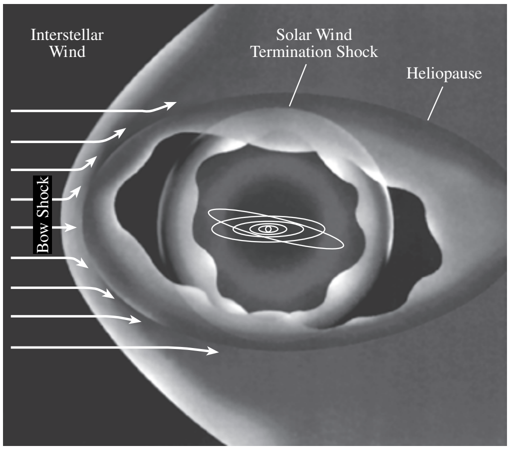


Fig. 3.11 Heliosphere With its solar wind going out in all directions, the Sun blows a huge bubble in space called the heliosphere. The heliopause is the name for the boundary between the heliosphere and the interstellar gas outside the solar system. Interstellar winds mold the heliosphere into a non-spherical shape, creating a bow shock where they first encounter it. The orbits of the planets are shown near the center of the drawing.

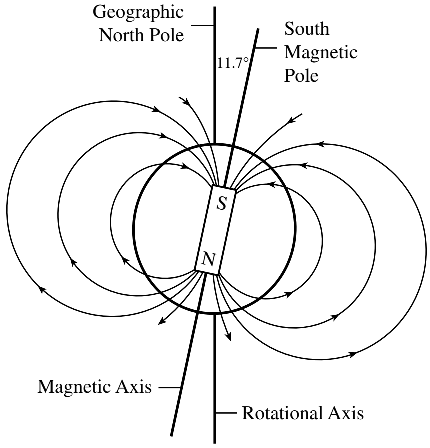


Fig. 3.12 Earth's magnetic dipole The Earth's magnetic field looks like that which would be produced by a bar magnet at the center of the Earth, with the north magnetic pole corresponding to the south geographic pole and vice versa. The Earth's magnetic dipole originates in swirling currents of molten iron deep in the Earth's liquid outer core, and extends more than 10 Earth radii, or 63.7 thousand kilometers, out into space on the side facing the Sun, and all the way to the Moon's orbit at 384 000 kilometers on the opposite side. Magnetic field lines loop out of the north magnetic pole and into the south magnetic pole. The lines are close together near the magnetic poles where the magnetic force is strong, and spread out where it is relatively weak. The magnetic axis is tilted at an angle of 11.7 degrees with respect to the Earth's rotational axis. This dipolar (two poles) configuration applies near the surface of the Earth, but further out the magnetic field is distorted by the solar wind.

3.6 Magnetized planets and magnetospheres

Earth's magnetic dipole

In 1600, William Gilbert (1544–1603), physician to Queen Elizabeth I of England, authored a treatise in Latin with the grand title *De magnete, magneticisque corporibus, et de magno magnee tellure*, translated into English as *Concerning Magnetism, Magnetic Bodies, and the Great Magnet Earth*. In this work, which is still available in its English version, Gilbert showed that the Earth is itself a great magnet, which explains the orientation of compass needles. It is as if there were a colossal bar magnet at the center of the Earth (Fig. 3.12).

At the equator, the two ends of a compass needle point north or south, toward the Earth's magnetic poles. At each magnetic pole, the needle would stand upright, pointing into or out of the ground. And in between, at intermediate latitudes, the compass needles point north or south with a downward dip of one end, but not vertically as at a pole.

Since the geographic poles are located near the magnetic ones, a compass needle is aligned in the north–south direction. We usually put an arrow on the north end of the needle, and an arrowed compass therefore points north. Since the Earth's rotation axis is inclined 11.7 degrees with respect to its magnetic axis, a compass needle does not point exactly toward the geographic north pole, but within 11.7 degrees of it. Currently, the north magnetic pole is located in northern Canada, while the south magnetic pole is off the coast of Antarctica.

We can describe the Earth's magnetism by invisible magnetic field lines, which orient compass needles. These lines of magnetic force emerge out of the north magnetic pole, loop through nearby space and re-enter at the south magnetic pole (Fig. 3.12). According to this convention, the south magnetic pole corresponds to the north geographic pole and vice versa. The lines are close together near the magnetic poles where the magnetic force is strong, and spread out above Earth's equator where the magnetism is weaker. You cannot see the magnetic field lines, but compass needles point along them, and other instruments can be used to measure their strength.

The magnetic field strength at the Earth's magnetic equator is 0.000 030 5 tesla, or 3.05×10^{-5} tesla. Measurements of the surface magnetic fields of Earth show stronger fields near the poles where the magnetic field lines congregate, at roughly twice the strength of the field at the equator. The magnetic strength in both regions is several times weaker than a toy magnet, but the comparison is somewhat misleading since the Earth is a very big magnet.

Magnetism pervades the entire volume of the Earth. This global quality is expressed quantitatively by the magnetic dipole moment, equal to the product of the equatorial magnetic field strength and the cube of the planet's equatorial radius. For the Earth we have an equatorial radius of 6378 kilometers and a magnetic dipole moment of 7.91×10^{15} tesla meters cubed. In comparison, the magnetic dipole moment of a typical laboratory electromagnet is more than 100 billion billion, or 10^{20}, times weaker.

How is magnetism generated within the Earth's hot, molten interior? Heat cooks magnetism out of a permanent magnet, and liquefaction melts it away. But the compass needles are not guided by a permanent magnet. Instead, the inside of the Earth is an electromagnet, generating magnetism by changing electric currents. Heat-driven circulation and the Earth's rotation combine to produce electrically conducting streams of molten iron, which generate the terrestrial magnetic fields by dynamo action.

Electric currents give rise to magnetic fields, and moving magnets generate electric currents. These two effects, in the churning liquid outer core of the Earth, can amplify and sustain the small magnetic field that the planet captured from its surroundings when it formed. As opposing streams of molten iron, carrying tiny magnetic fields, sweep past one another, each induces currents in the other. This creates more magnetism, which induces more currents, and so on. In this way, the dynamo in the Earth's liquid outer core generates the magnetic fields detected at the terrestrial surface, taking energy from both the internal heat and the rotation energy of the planet. And because the currents deep down inside the Earth are always varying, the Earth's magnetism is a dynamic, changing thing.

Ancient magnetic rocks, found on the flanks of volcanoes, indicate that the Earth's magnetic field has not always been the same as it is today. When the molten volcanic lava flows to the surface and hardens into rock, its internal magnetism lines up with the Earth's magnetic field and freezes into position. These magnetic fossils record the direction and intensity of the terrestrial magnetic field when and where the lava solidified.

An inspection of magnetic fossils of differing ages from all parts of the world resulted in an amazing discovery. The great magnet of the Earth has flipped, or reversed its direction, many times in the past. During each flip, the north magnetic pole becomes the south one, and vice versa. The deep electric currents readjust; always remaining nearly aligned with the rotation axis, but with a swap in the magnetic poles.

An examination of volcanoes on land indicates that the Earth's magnetic field has reversed itself at least nine times over the past 3.6 million years. And the ordered succession of magnetized rocks on the spreading ocean floors records 100 and more full reversals of the direction of the magnetic poles during the past 200 million years. Tens of thousands of years separate some of the magnetic field reversals, while tens of million of years separate others.

So the terrestrial magnetic field is inevitably headed for a magnetic flip, but we don't know exactly when. The arrows on compass needles will then point south instead of north, reversing their direction. Animal species and satellites that depend on magnetic fields for guidance will lose their orientation, and will have to adapt to the changing field. The navigation systems of migrating birds and monarch butterflies, for example, depend in part on internal compasses that sense directions from the Earth's magnetic field. Honeybees, some wasps, some fish, sea turtles and even a species of mole rat take bearings magnetically.

The Earth's magnetic fields also influence the space near the Earth. They extend away from the Earth, decreasing in strength as the inverse cube of the distance. Yet they remain strong enough to shield the Earth from the full force of the Sun's charged winds.

Earth's protective magnetosphere

Fortunately for life on Earth, the terrestrial magnetic field deflects the Sun's winds away from the Earth, like a rock in a stream or a windshield that deflects air around a car. It hollows out a protective cavity in the solar wind called the magnetosphere. The magnetosphere of the Earth, or any other planet, is an enveloping bubble of magnetism. It is that region surrounding the planet in which its magnetic field dominates the behavior of electrically charged particles such as electrons, protons and other ions. It diverts most of the solar wind around our planet at a distance far above the atmosphere, thereby protecting humans on the ground from possibly lethal solar particles.

The dipolar (two-pole) magnetic configuration applies near the surface of the Earth, but further out the magnetic field is distorted by the Sun's perpetual wind. Although it is exceedingly tenuous, far less substantial than a terrestrial breeze or even a whisper, the solar wind is powerful enough to mold the outer edges of the Earth's magnetosphere into a changing asymmetric shape, like a tear drop falling toward the Sun (Fig. 3.13).

The hot, high-speed, magnetized solar wind confronts the Earth's magnetic field close to home, usually at a distance from the Earth's center of about 10 times the Earth's radius on the dayside that faces the Sun. Here the solar

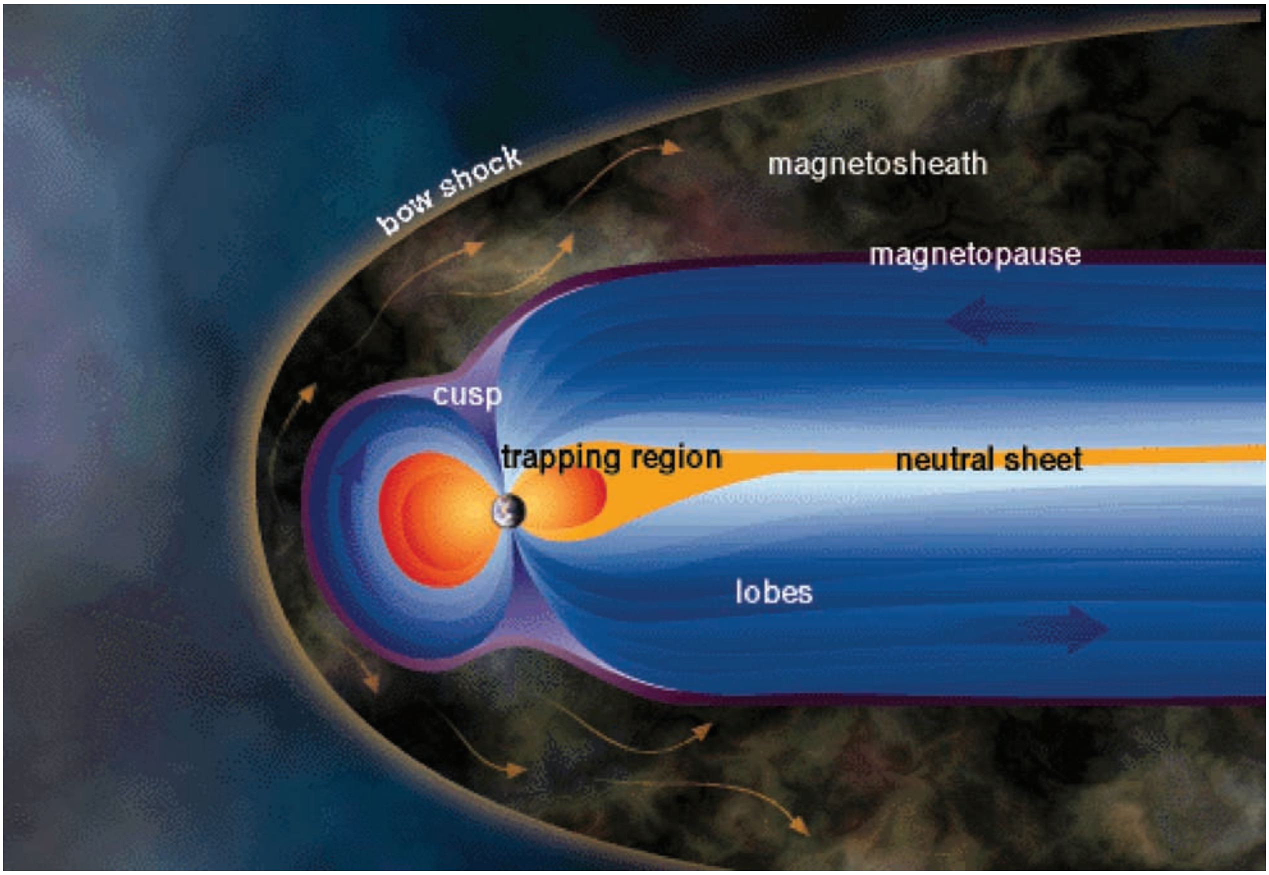


Fig. 3.13 Asymmetric magnetosphere The Earth's magnetic field carves out a hollow in the solar wind, creating a protective cavity called the magnetosphere (*blue*). It is sculpted into an asymmetric shape by the solar wind, with a bow shock that forms at about 10 Earth radii on the sunlit, dayside facing the Sun (*left*). The location of the bow shock is highly variable since it is pushed in and out by the gusty solar wind. The magnetopause marks the outer boundary of the magnetosphere, at the place where the solar wind takes control of the motions of charged particles. The solar wind is deflected around the Earth, pulling the terrestrial magnetic field into a long magnetotail on the nightside (*right*). The red regions in the inner magnetosphere contain the plasmasphere, the ring current and the outer Van Allen belt, where electrons, protons and other ions are trapped in closed paths. (Courtesy of ESA.)

wind pushes the Earth's magnetism in, compressing its outer magnetic boundary and forming a shock wave. It is called a bow shock because it is shaped like waves that pile up ahead of the bow of a moving ship. The bow shock results because the solar wind is supersonic, moving faster than sound waves and other waves that might propagate through the wind. The motion of the solar wind around the magnetosphere has therefore been compared to the flow of air around a supersonic aircraft.

After forming the bow shock, the solar wind encounters and flows around the magnetopause, the boundary between the solar wind and the magnetosphere. The magnetic field carried in the solar wind merges with that of the planet, and stretches it out into a long magnetotail on the nightside of Earth. The magnetic field points roughly toward the Earth in the northern half of the tail and away in the southern. The field strength drops to nearly zero at the center of the tail where the opposite magnetic orientations lie next to each other and currents can flow.

Thus, the Earth's magnetosphere is not precisely spherical. It has a bow shock facing the Sun and a long magnetotail in the opposite direction. The term "magnetosphere" therefore does not refer to form or shape, but instead implies a sphere of influence.

Trapped particles

The Earth's protective magnetic cocoon is not perfect. Energetic charged particles flowing from the Sun can penetrate the magnetic defense and become trapped within the magnetosphere. This was realized in 1958 when James A. Van Allen (1914–2006) and his students used instruments aboard the *Explorer 1* and *3* satellites to unexpectedly discover a large flux of high-energy electrons and protons that girdle the Earth far above the atmosphere, moving within

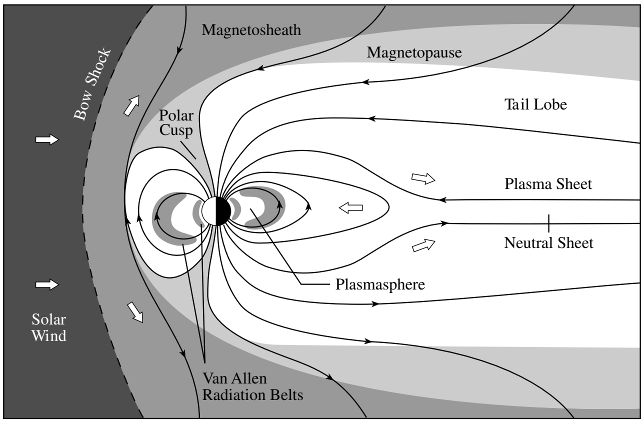


Fig. 3.14 Earth's magnetosphere The Earth, its auroras, atmosphere and ionosphere, and the two Van Allen radiation belts all lie within the Earth's magnetosphere. Similar magnetic cavities are found around other magnetized planets. Electrons and protons in the solar wind are deflected at the bow shock (*left*), and flow along the magnetopause into the magnetic tail (*right*). Electrified particles can be injected back toward the Earth and Sun within the plasma sheet (*center*).

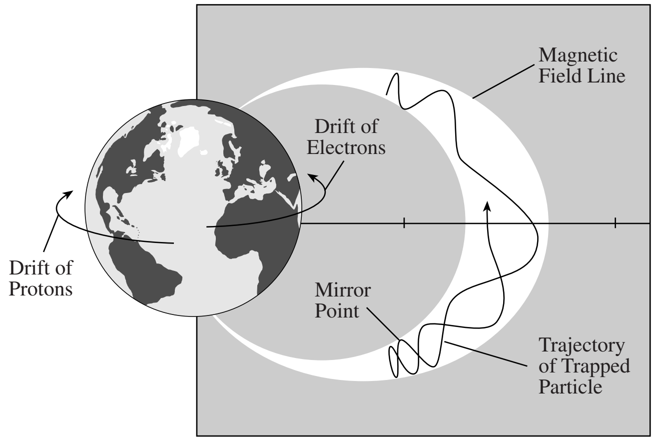


Fig. 3.15 Magnetic trap Charged particles can be trapped by Earth's magnetic field. They bounce back and forth between polar mirror points in either hemisphere at intervals of seconds to minutes, or they also drift around the planet on timescales of about an hour. As shown by the Norwegian scientist Carl Størmer (1874–1957) in 1907, with the trajectories shown here, the motion is turned around by the stronger magnetic fields near the Earth's magnetic poles. Because of their positive and negative charge, the protons and electrons drift in opposite directions.

two belts that encircle the Earth's magnetic equator but do not touch it. They resemble a gigantic, invisible, torus-shaped doughnut. This was the first major discovery of the Space Age.

These regions are sometimes called the inner and outer Van Allen radiation belts. Van Allen used the term "radiation belt" because the charged particles were then known as corpuscular radiation; the nomenclature does not imply either electromagnetic radiation or radioactivity. The radiation belts lie within the inner magnetosphere at distances of 1.5 and 4.5 Earth radii from the center of the Earth, creating a veritable shooting gallery of high-speed electrons and protons in nearby space (Fig. 3.14).

In 1907, about half a century before the discovery of the radiation belts, the Norwegian geophysicist Carl Størmer (1874–1957) showed how electrons and protons can be almost permanently confined and suspended in space by the Earth's dipolar magnetic field. An energetic charged particle moves around the magnetic fields in a spiral path toward one magnetic pole. Its trajectory becomes more tightly coiled in the stronger magnetic fields close to a magnetic pole, where the intense polar fields act like a magnetic mirror, turning the particle around so it moves back toward the other pole.

Thus, the electrons and protons bounce back and forth between the north and south magnetic poles (Fig. 3.15). It takes about one minute for an energetic electron to make one trip between the two polar mirror points. The spiraling electrons also drift eastward, completing one trip around the Earth in about an hour. There is a similar drift for protons, but in the westward direction. The bouncing can continue indefinitely for particles trapped in the Earth's radiation belts, until the particles collide with each other or some external force distorts the magnetic fields.

The problem at the time Størmer developed his theory was that there was no mechanism known to allow electrically charged particles into the dipolar magnetic field. After all, if electrons and protons cannot leave the magnetic cage, how could they get into it in the first place? Nevertheless, instruments aboard spacecraft have shown that energetic charged particles have entered the trap. They include particles from the Sun that arrive via the solar wind and penetrate the Earth's magnetic defense through a temporary opening in it.

The solar wind carries the Sun's magnetic field with it, and the solar magnetism is draped around the magnetosphere when encountering it. The solar magnetic field can open up the Earth's magnetic field when the two fields are pointing in opposite directions where they touch. When this happens, the two fields become linked, just as the opposite poles of two toy magnets stick together. The merging process, known as magnetic reconnection, can create an opening in the Earth's magnetic field, forming a portal through which the solar particles can flow.

The solar wind is then plugged into the Earth's electrical socket, and our planet becomes wired to the Sun along magnetic fields that can stretch all the way back to the solar corona. Tons of high-energy particles may then flow

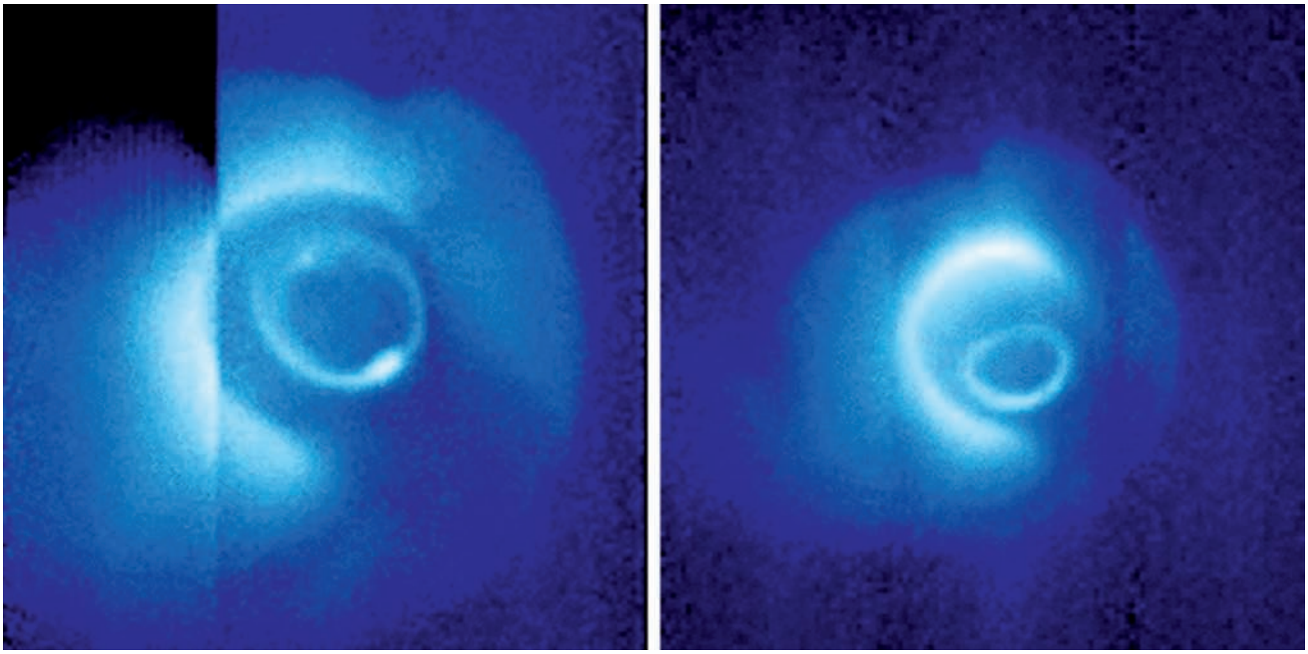


Fig. 3.16 Plasmasphere The Earth is surrounded by a layer of plasma, located in an inner part of the magnetosphere and just outside the ionosphere, which is at the top of the Earth's atmosphere. This plasmasphere is created by energetic sunlight ionizing molecules and atoms in the Earth's upper atmosphere, and it is contained by the Earth's magnetic field. Plasma is an electrically neutral collection of electrons, protons and other ions. These images were obtained on 24 May 2000 in ultraviolet light from the *IMAGE* spacecraft. The Earth is at the center of the images, and the Sun is to the upper left. The view is toward Earth's north pole. The emission of the plasmasphere is brightest on the side pointing toward the Sun. The oval shape of the Earth's northern aurora is detected inside the arc of the plasmasphere. (Courtesy of NASA/EUV *IMAGE* science team.)

into the magnetosphere along this magnetic highway and through the opening before it closes again. The magnetic reconnection can occur during either a frontal assault near the bow shock and magnetic poles, or from the rear in the immense magnetotail.

In June 2007, for example, coordinated measurements from five identical *THEMIS* satellites, placed in carefully coordinated orbits, discovered a giant breach in the Earth's magnetic field wider than the Earth itself. The opening was explained by brief magnetic reconnection on the dayside of the Earth, facing the Sun. The magnetic portal formed over the Earth's equator, and then rolled over the planet's magnetically open polar regions.

The immense magnetic tail provides another location for breaking into the Earth's magnetic domain. When the solar and terrestrial magnetic fields touch each other, the magnetotail can be punctured, providing a back-door entry that funnels energy and particles into the magnetosphere. The magnetotail snaps like a rubber band that has been stretched too far. The snap catapults the outer part of the tail away from the Earth and propels the inner part back toward it. Once inside the magnetic trap, the charged particles can be additionally accelerated to higher energies.

A plasmasphere is found in the inner part of the Earth's magnetosphere (Fig. 3.16). It is located just outside the upper terrestrial atmosphere, called the ionosphere. The upper reaches of our planet's atmosphere are exposed to ultraviolet light from the Sun, which ionizes the atmosphere's atoms and molecules. The charged particles that result from the ionization are electrons, protons and other ions. The electrons have the lowest mass and highest velocities, and some of them move fast enough to overcome the Earth's gravitational pull. The growing number of escaped, negatively charged electrons electrically attracts the positively charged protons and other ions, pulling them out. The Earth's magnetic field then traps oxygen ions, protons and electrons derived from the ionosphere, creating the plasmasphere that rotates with the Earth's magnetic field and extends outward to include the radiation belts. Farther out, the magnetosphere is dominated by its interaction with the solar wind.

Energetic charged particles coming from interstellar space, known as cosmic rays, may also play a role in feeding the radiation belts, supplying the inner one with its high-energy protons. When cosmic rays bombard the Earth's atmosphere, which lies below the radiation belts, they collide with atoms in the air and eject neutrons from

Table 3.5 Planetary magnetic fields[a]

Planet	Magnetic dipole moment (Earth = 1)	Magnetic field at the equator, B_o (Earth = 1)	Tilt of magnetic axis center (degrees)	Offset from planet R_{MP}	Bow shock stance, (R_P)	Planet equatorial radius, R_P (km)
Mercury	0.0007	0.0033	+14	0.05 R_M	1.5 R_M	$R_M = 2439$
Earth	1	0.305	+11.7	0.07 R_E	10 R_E	$R_E = 6378$
Jupiter	20,000	4.28	−9.6	0.14 R_J	42 R_J	$R_J = 71\,492$
Saturn	600	0.22	<1.0	0.04 R_S	19 R_S	$R_S = 60\,268$
Uranus	50	0.23	−58.6	0.3 R_U	25 R_U	$R_U = 25\,559$
Neptune	25	0.14	−47.0	0.55 R_N	24 R_N	$R_N = 24\,764$

[a] The magnetic field strengths are given at the surface of Mercury and the Earth and at the cloud tops for the giant planets. Venus and Mars have no detected global, dipolar magnetic field, with respective upper limits of 2×10^{-9} and 10^{-8} tesla. Here the magnetic dipole moment, $D_P = B_o R_P{}^3$, is given in units of the Earth's magnetic dipole moment of 7.91×10^{15} T m^3. The tilt is the angle between the magnetic axis and the rotation axis. Here we use the SI unit for magnetic field strength, the tesla (T). The c.g.s. unit of magnetic field strength, the gauss (G), can be computed from 1 T 10^4 G. The nanotesla (nT) is also used, with 1 nT = 10^{-9} T, and the nanotesla has historically also been called the gamma. A dipole moment of 1 T m^3 equals 10^{10} G cm^3, where m and cm respectively denote meter and centimeter. The equivalent unit of 1 G cm^3 = 10^{-3} A m^2 is also used, where the current is in units of amperes (A). The equatorial radius of the planets is given in kilometers (km).

the atomic nuclei. These neutrons travel in all directions, unimpeded by magnetic fields since they have no electrical charge.

Once it is liberated from an atomic nucleus, a neutron cannot stand being left alone. A free neutron lasts only about 10 minutes on average before it decays into an electron and proton. A small fraction of the neutrons produced in our atmosphere by cosmic rays move out into the inner radiation belt before they disintegrate, producing electrons and protons in places they could not otherwise have reached. These electrically charged particles are immediately snared by the magnetic fields and remain stored within them, accumulating in substantial numbers over time.

Planets with magnetospheres

Magnetic fields are ubiquitous in the solar system. Earth, Mercury and all the giant planets have strong magnetic fields generated within the planet, and Jupiter and Saturn have extensive magnetospheres. A magnetic field has been found on Mars, but the field's patchy nature suggests that it is not the result of an active internal dynamo. Instead the magnetic field on Mars is probably a remnant of former times, frozen into expanses of solidifying lava. Venus is the only major planet to have no detectable magnetic field. A dipolar magnetic field has even been found on at least one satellite, Jupiter's Ganymede.

Of the eight major planets, six are known to generate detectable, global magnetic fields. As with the Earth, the magnetic fields near these planets can be described by a magnetic dipole, and the best characterization of each planet's intrinsic magnetism is the magnetic dipole moment. The dipole moment divided by the cube of the planet's radius yields the average strength of the magnetic field along the magnetic equator. Jupiter's magnetic moment is 20 000 times that of Earth. Saturn's magnetic moment is 600 times larger than the Earth's, but still about 30 times weaker than Jupiter's. The magnetic dipole moments, equatorial magnetic field strengths and planetary radii are given in Table 3.5 for the six planets with known dipolar magnetic fields.

Because the characteristic timescale for the decay of a magnetic field in a planetary interior is much less than the age of the planet, its global magnetic field must now be amplified and rejuvenated by an internal dynamo. Such a magnetic dynamo exists within a large, fluid, electrically conducting region.

A planetary magnetosphere is the volume in space from which the solar wind is deflected. The extent of a planet's magnetosphere depends upon both the strength of its magnetic field and the intensity of the solar wind at the planet's distance. The strong magnetic fields generated in Jupiter and Saturn, for example, as well as the weak solar-wind pressures present in the outer solar system, permit them to hollow out a larger cavity in the solar wind, with magnetospheres that are much larger than the Earth's.

The magnetospheres of Jupiter and Saturn are dominated by planetary rotation; satellites are a major source of

Focus 3.5 Planetary magnetospheres

Six planets are known to have magnetospheres. The size of the magnetosphere, on the day side facing the Sun, is determined by the distance, R_{MP}, along the planet–Sun line at which the pressure of the planetary magnetic field balances the dynamic ram pressure of the solar wind. The magnetic pressure at the surface of the planet is given by $B_o^{\,2} / (2\mu_0)$, where B_o is the equatorial magnetic field strength, and $\mu_0 = 4\pi \times 10^{-7}$ is the permeability of free space. Since the dipole's magnetic field strength falls off as the cube of the distance from the planet, the magnetic pressure decreases as the sixth power of that distance. This means that the standoff point where the two pressures are equal occurs when:

$$\text{Magnetic pressure} = \frac{R_P^6 B_o^2}{2\mu_0 R_{MP}^6} = m_p N V^2$$
$$= \text{Wind ram pressure}$$

where the planet's radius is R_P, the proton mass is $m_p = 1.67 \times 10^{-27}$ kilograms, N is the number density of the protons in the solar wind at the planet's distance from the Sun, and V is the solar wind velocity at that distance. Solving for R_{MP} we have:

$$R_{MP} = \left(\frac{B_o^2}{2\,\mu_0 m_p N V^2} \right)^{1/6} R_P$$

At the Earth's distance from the Sun, the number density of the solar wind is about $N = 5$ million protons per cubic meter and the wind velocity is about $V = 400$ kilometers per second. The equatorial surface magnetic field strength of the Earth is $B_o = 3.05 \times 10^{-5}$ tesla. With these numbers our equation gives $R_{ME} = 10\ R_E$, so the bow shock of the Earth is out at about 10 times the Earth's radius. The bow shock distance can be reduced to half this value when a powerful coronal mass ejection from the Sun hits the Earth, producing extra ram pressure. Moreover, unusual drops in the wind's pressure have very occasionally inflated the leading edge of the Earth's magnetosphere five or six times farther out in space, until it engulfed the Moon.

The values of R_{MP} for the other planets can be inferred by noting that the solar wind number density N falls off with the inverse cube of the distance of the planet from the Sun, while the solar wind velocity remains relatively constant. Values of R_{MP} were given in Table 3.5 for the six planets with detected dipolar magnetic fields.

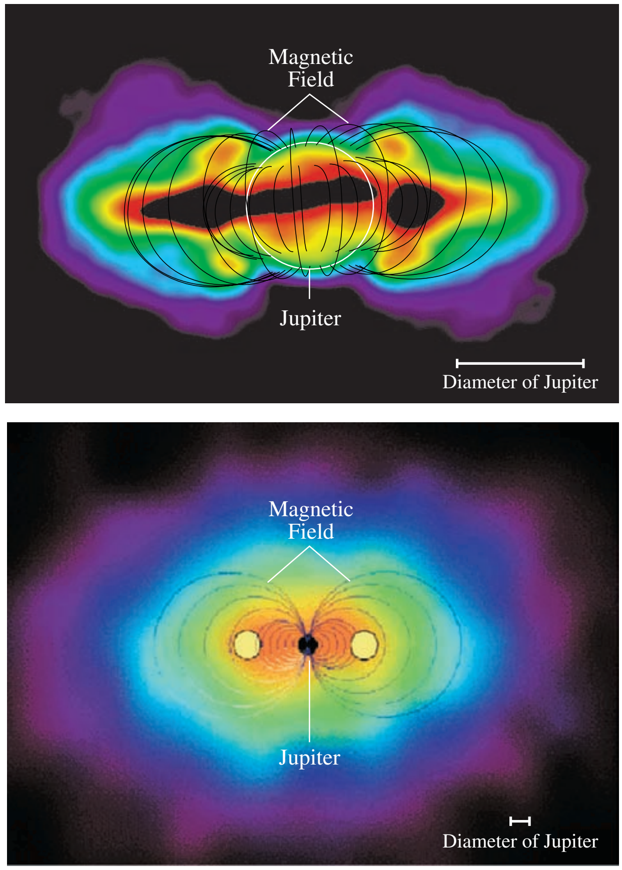


Fig. 3.17 Jupiter's magnetosphere High-speed electrons that are trapped in Jupiter's magnetosphere emit steady radio radiation (*top*) by the synchrotron process; it is detected by ground-based radio telescopes such as the Very Large Array. An instrument aboard the *Cassini* spacecraft measured energetic atoms (*bottom*) created when fast-moving ions within the magnetosphere picked up electrons to become neutral atoms. The two open circles denote Io's orbital position on each side of the planet; the central black disk denotes Jupiter. [Courtesy of Imke de Pater, U. C. Berkeley (*top*) and NASA/JPL/JHUAPL (*bottom*).]

their charged particles, and internal forces shape those particles into an equatorial disk. Unlike Jupiter and Saturn, the magnetospheres of Uranus and Neptune are largely empty, and their magnetic fields are unexpectedly tilted and offset from the centers of these planets.

On the side facing the Sun, each of the planetary magnetic fields is compressed by the solar wind, forming a bow shock. It is located at the place where the pressure of the planet's magnetic field just equals the pressure of the solar wind (Focus 3.5). Such standoff distances for the six planets with detected magnetic fields were given in Table 3.5, but the varying solar wind pressure can alter this distance by a factor of 2.

The general form of Jupiter's magnetosphere resembles that of the Earth (Fig. 3.17), but its dimensions are at

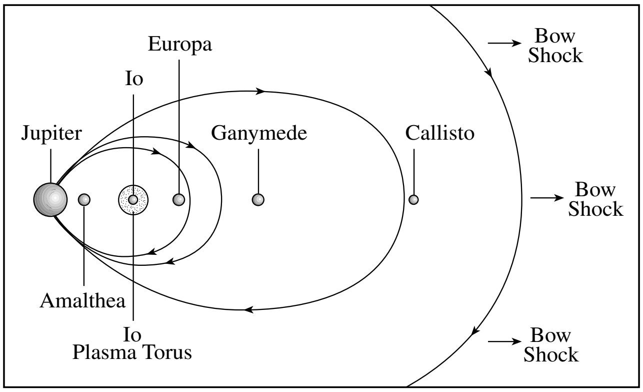


Fig. 3.18 Satellites within Jupiter's magnetic field This cross-section shows that the four Galilean satellites, Io, Europa, Ganymede and Callisto, are all embedded within Jupiter's magnetosphere. Small satellites orbit Jupiter within the orbit of Io, the innermost Galilean satellite. The outermost Galilean satellite, Callisto, orbits Jupiter near its bow shock. All of these satellites are being continuously bombarded with energetic charged particles that are trapped within Jupiter's magnetosphere. The distance from Jupiter to Callisto is 1.88 million (1.88×10^6) kilometers, while the radius of the Sun is 0.696 million kilometers, so Jupiter's bow shock is bigger than the Sun.

least 1200 times greater. It is larger than the Sun in size, with a bow shock at more than 3 million kilometers or at least 42 times the planet's radius. *Pioneer 10* and *Voyager 1* first encountered Jupiter's bow shock at 95 and 86 planetary radii, but the shock moved in and out due to the variable solar wind. Jupiter's largest satellites are all embedded within its magnetosphere and interact with it (Fig. 3.18).

Jupiter's enormous magnetic tail, driven outward by the solar wind, is almost a billion kilometers long. It spans the distance between the orbits of Jupiter and Saturn, which is as great as the distance from the Sun to Jupiter itself. In contrast, the Earth's magnetic tail barely flicks across our Moon's path, less than a half-million kilometers from the Earth. Thus, Jupiter's magnetosphere is the largest enduring structure in the solar system, although it is occasionally and temporarily exceeded in size by comet tails.

Jupiter's magnetic field was first recognized in 1954–55 by Earth-based observations of the planet's intense radio emission, and then directly measured by visiting spacecraft. High-speed electrons trapped within the planet's magnetic field generate the radio signals (Focus 3.6). The synchrotron radio emission is beamed in a direction nearly parallel to the magnetic equator, and it is therefore swept past an Earth-based observer as the planet rotates and brings the magnetic equator in and out of alignment with the line of sight. Periodic variations in the strength of the

Focus 3.6 Radio broadcasts from Jupiter

The discovery of intense radio emission from Jupiter is one of the many examples of the accidental discovery of an unexpected phenomenon while looking for something else. In 1954–55 Bernard F. Burke (1928–) and Kenneth L. Franklin (1923–2007) were using ground-based radio telescopes to observe the meter-wavelength emission of the Crab Nebula, a famous remnant of a stellar explosion in 1054 AD. They planned to study the changes in the Crab's radio signal, at a wavelength of 13.6 meters, as it passed behind the Sun, thereby determining properties of the outer solar atmosphere.

Their observations were hampered by radio bursts that resembled terrestrial interference, but the alert scientists noticed that they only appeared when the radio telescope was pointing in a certain direction in the sky. This meant that the radio bursts had an extraterrestrial origin. They were at first assumed to be coming from the Sun, an intense source of variable radio radiation, but the calculated position in the sky coincided with Jupiter.

Soon thereafter, a steady Jovian radio signal was found at shorter wavelengths of several centimeters, and scientists interpreted this emission as synchrotron radiation emitted by electrons trapped in the Jovian magnetic field and moving at relativistic speeds approaching the velocity of light. The electrons spiral about the magnetic field, emitting the radio radiation, named after the synchrotron particle accelerator on Earth where similar radiation was first observed visually.

The new theory for Jupiter's radio signals was confirmed by the observation of two characteristic signatures of synchrotron radiation. The radio broadcasts were weaker at shorter wavelengths, and stronger at the longer ones, unlike the thermal emission of a hot gas that is most intense at shorter wavelengths. The non-thermal radio radiation was also polarized, with a preferred orientation or direction, which ought to coincide with that of the magnetic fields. In addition, radio interferometer measurements indicated that the radio emission was much larger than the planet and roughly aligned with its equator.

As Jupiter rotates, it carries the magnetic fields and their trapped electrons with it. Since the radio signal of the electrons is beamed in a particular direction, it sweeps past the observer once every rotation, providing a precise measurement of the planet's rotation period: 9 hours 55 minutes 29.7 seconds, or 9.9249 hours.

radio emission indicate that the magnetic field rotates with a period of precisely 9 hours 55 minutes 29.7 seconds, or 9.9249 hours. Since the magnetic fields are generated deep within the planet, this is assumed to be Jupiter's rotation period; it differs from the rotation speed inferred from visible clouds that are blown in different directions and at various speeds by powerful winds.

In December 1973 and December 1974 *Pioneer 10* and *11* confirmed Jupiter's strong dipolar magnetic field and energetic trapped electrons. In 1979 *Voyager 1* and *2* obtained information about Jupiter's outer magnetosphere, as well as the magnetic interaction between Jupiter and Io.

The *Pioneer* data showed that the magnetic field at the cloud tops is 4.28×10^{-4} tesla, or about 14 times stronger than the Earth's equatorial magnetic field strength. As near the Earth's surface, the magnetic fields at Jupiter's cloud tops are strongest at the poles and weakest along the equator. The magnetic field can be described as a dipole with a magnetic axis that is titled at 9.6 degrees with respect to the rotation axis, similar to the Earth's tilt of 11.7 degrees; but the magnetic poles are reversed in comparison to those of the Earth, so a north-seeking terrestrial compass would point south in the vicinity of Jupiter.

Like its terrestrial counterpart, Jupiter's magnetosphere contains electrons and protons that are supplied from outside by the variable solar wind. Unlike the Earth, the giant planet's magnetosphere is also fed from within, by ions erupted from the active volcanoes on its innermost large satellite, Io. The dominant ions are sulfur and oxygen, both products of Io's unique volcanic activity. Jupiter's belts of charged particles resemble the terrestrial Van Allen belts in shape, but the Jovian belts are up to a million times more densely filled with particles than those near the Earth are.

Ions and electrons within Jupiter's inner magnetosphere are accelerated by the spinning magnetic field of the planet, eventually reaching very high energies. The inner magnetosphere is a stiff, permanent structure that is tied to the planet and rotates with it. Once an electrically charged particle enters this region, the magnetic field picks the particle up and takes it for long rides around the planet. As the powerful magnetic field spins, it extracts rotational energy from the planet, lashing and accelerating the charged particles to nearly the speed of light.

Thus, the energy that populates and maintains the magnetosphere of Jupiter comes principally from the planet's rotation, as well as the tidal flexing of Io that results in its volcanoes. In contrast, the Earth's magnetosphere is principally energized by the solar wind. The numerous high-energy particles in Jupiter's magnetosphere are capable of destroying sensitive electronic circuits in spacecraft that pass near the planet.

The high-speed charged particles exert an outward pressure on Jupiter's magnetic field, inflating it like an air-filled balloon. Because the field is weakest in its equatorial plane, the forces and pressures associated with the rapid rotation stretch the equatorial regions outwards in the form of a thin, elongated disk.

The varying solar wind pressure buffets Jupiter's outer magnetosphere, so it expands and contracts. As varying solar activity pushed the magnetosphere in and out, approaching *Pioneer 10* and *11* and *Voyager 1* and *2* spacecraft crossed the bow shock several times. The changing shape and location of the magnetotail similarly caused the outward-bound spacecraft to cross it many times.

When gusts in the solar wind compress Jupiter's outer magnetosphere, some of its high-speed, electrically charged particles squirt out into interplanetary space with energies that exceed the typical energy of electrons and protons in the solar wind. The Jovian particles are continually replenished by acceleration within the planet's magnetosphere, spraying energetic electrons and protons throughout the solar system. Some of them reach the orbit of the Earth and even that of Mercury.

The discovery of Saturn's magnetosphere did not occur until September 1979 when *Pioneer 11* first crossed its bow shock, at 24 Saturn radii, closely followed by the *Voyager 1* and *2* encounters in November 1980 and August 1981, respectively. The magnetic field strength at Saturn's cloud tops is 70 percent of that at the Earth's equator, but spread over a much bigger volume. Saturn's dipolar magnetic field is almost precisely aligned with its poles of rotation. Like Jupiter, the magnetic poles of Saturn are reversed compared to those of the Earth. High-speed electrons in the spinning magnetic field give rise to periodic radio emission, and the inferred rotation period is 10 hours 39 minutes 22.3 seconds, or 10.6562 hours, and just 44 minutes longer than Jupiter's rotation period.

Saturn's sizeable satellites and its rings absorb energetic electrons and protons. Perhaps as a result, its magnetosphere does not contain a high density of high-energy electrons. The planet also does not have a volcanically active satellite to generate sulfur and oxygen ions. Its magnetic trap is instead permeated with low-energy ionized material, protons and oxygen ions, chipped or sputtered off the water ice in the planet's rings and on its satellite surfaces. A vast dense cloud of neutral, or un-ionized, hydroxyl (OH) molecules envelops the rings; it is also derived from the water ice H_2O.

The magnetic axes of Uranus and Neptune are tilted at a large angle with respect to their rotational axis, by 59 degrees for Uranus and 47 degrees for Neptune (Fig. 3.19),

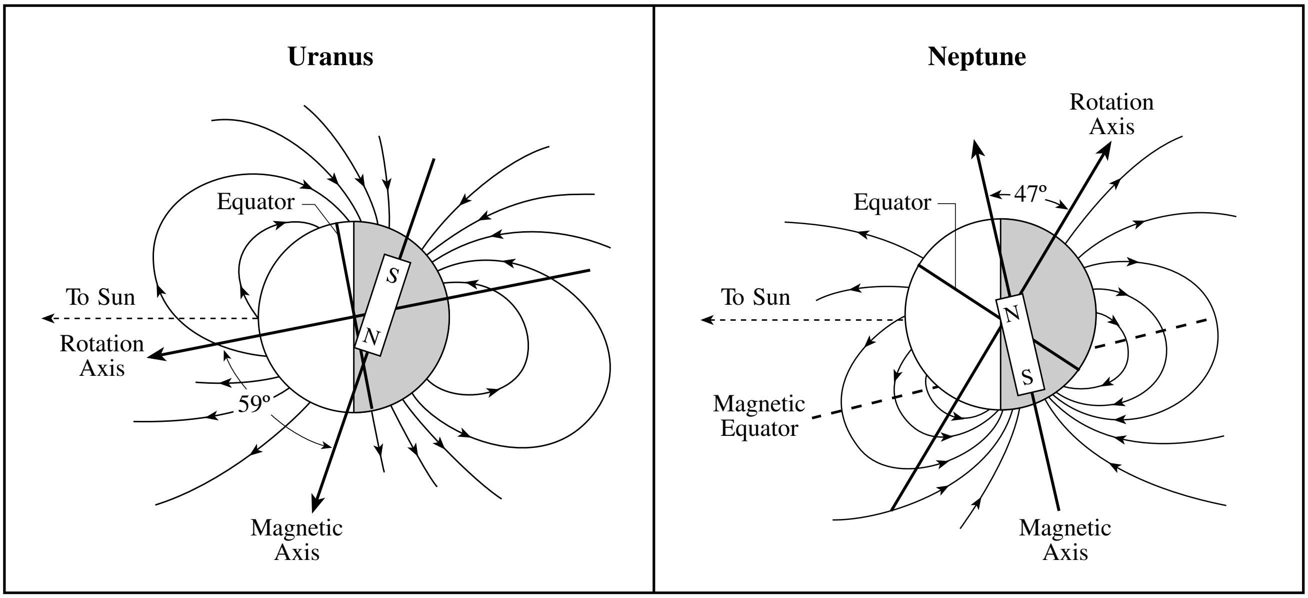


Fig. 3.19 Tilted magnetic fields The magnetic fields of Uranus and Neptune can be represented by a simple bar magnet, or dipole, embedded in the planet, but with a magnetic axis that is tilted with respect to the rotation axis. For Uranus the tilt is about 59 degrees; Neptune has a tilt of 47 degrees. In contrast, the magnetic axes of Jupiter, Saturn and Earth are much more nearly aligned with their rotation axes. The arrow of the rotation axis points from the geographic south towards geographic north, and the magnetic axis similarly points from magnetic south to magnetic north. On Uranus and Neptune a terrestrial compass would point toward the southern hemisphere of the planet, while on Earth it points toward the geographic north pole. In addition to the dipole part of their magnetic field, Uranus and Neptune have a large additional component known as the quadrupole. A method of visualizing this is to imagine that the dipole has a magnetic center that is offset from the center of the planet. As shown here, the equivalent offset for Uranus is almost a third of the planet's radius, and there is a larger offset for Neptune of nearly half its radius. But such off-center dipoles are only useful as a picture of what the external field looks like and do not help in understanding how it is produced deep down.

and these planets have fully developed magnetospheres. The rotation periods of the magnetic fields of Uranus and Neptune, inferred from their periodic radio emission, are 17.24 and 16.11 hours, respectively.

All planetary magnetic fields are generated by the dynamo action of moving electrically conducting material in their interior. Internal rotation and convection produce the motions, somewhat like a spinning and boiling pot of water. Vigorous internal convection is powered by the decay of radioactive elements in the Earth and Mercury; the giant planets have retained much of their primordial heat to drive the internal convection and power their dynamo. The combination of convection and rotation concentrates the magnetism, amplifying its strength and regenerating the magnetic fields.

Mercury and the Earth have cores of molten iron alloys. At the high pressures inside Jupiter and Saturn, their most abundant ingredient, hydrogen, behaves like a liquid metal. Their strong magnetic fields are attributed to electrical currents driven by the fast rotation of their liquid metallic interiors. For Uranus and Neptune, water-rich material within their vast internal oceans most likely provides the electrical conductivity.

3.7 Aurora

Terrestrial aurora

Curtains of green or red light dance and shimmer across the night sky in the Earth's polar regions, far above the highest clouds (Figs. 3.20, 3.21). This light is called the *aurora* after the Roman goddess of the rosy-fingered dawn, a designation that has been traced back to Galileo Galilei (1564–1642). The aurora seen near the north and south poles have been given the Latin names *aurora borealis*, for Northern Lights, and *aurora australis*, for Southern Lights.

Most people never see the awesome lights, for aurora are normally confined to high latitudes in the north or south polar regions. But this does not mean that the aurora occur infrequently. Residents in far northern locations can see a faint aurora every clear and dark night.

The northern aurora borealis has been documented for centuries. Ancient Vikings (500–1500 AD) thought they were the spirits of fallen warriors being carried to Valhalla, the home of the gods. The southern aurora australis have never achieved a renown comparable to the Northern Lights, probably because the southern ones are not

Fig. 3.20 Northern Lights Spectacular green curtains of light dance and shimmer across the northern sky. High-energy electrons are funneled down the Earth's polar magnetic field lines into the atmosphere, where they excite oxygen atoms that fluoresce green light, like a cosmic neon sign. The color is usually green, but the aurora can also have red bottoms, arising from excited nitrogen molecules. This photograph of the fluorescent Northern Lights, or *aurora borealis*, was taken over Fairbanks, Alaska. (Courtesy of Jan Curtis.)

Fig. 3.21 Twilight aurora The form and brightness of the aurora will vary with time of night, appearing near twilight as bands or diffuse arcs, then rising and brightening as the night progresses and the display forms a curtain or drapery formation. The dark gaps located between the green bands, known as black aurora, have been attributed to negatively charged particles that are sucked out of the Earth's ionosphere along adjoining magnetic field lines, climbing to over 20 000 kilometers and lasting for several minutes. This photograph was taken over Fairbanks, Alaska. (Courtesy of Jan Curtis.)

usually located over inhabited land and are instead seen from oceans that are infrequently traveled.

Rare, brilliant aurora can extend down to the Earth's equator, becoming visible as far south as Athens, Rome or Mexico City. They were noted by the ancient Greeks. Plutarch (*c.* AD 46–120) reported one that occurred in 427 BC, but aurora do not extend down to Greece very often, perhaps every 50 or 100 years.

Since aurora become more frequent as one travels north from tropical latitudes, it was thought that they would become brighter and occur most frequently at the highest northern latitudes. Arctic explorers were therefore surprised to see that the intensity and frequency of aurora did not increase all the way to the poles and instead peaked in an oval-shaped band that encircles the North Pole. This northern aurora oval has a radius of about 2250 kilometers and is centered on the Earth's north magnetic pole, with an inner and outer radius separated by about 500 kilometers.

Nowadays we can use spacecraft to view both the northern and southern lights from space (Figs. 3.22, 3.23). The *Space Shuttle* has even flown right through the Northern Lights. While inside the display, astronauts could close their eyes and see flashes of light caused by the charged aurora particles speeding through the spacecraft walls and into their eyeballs.

When viewed from above, the aurora form a luminous oval centered at each magnetic pole, resembling a fiery halo (Fig. 3.23). The aurora oval is constantly in motion, expanding a little toward the equator or contacting a bit

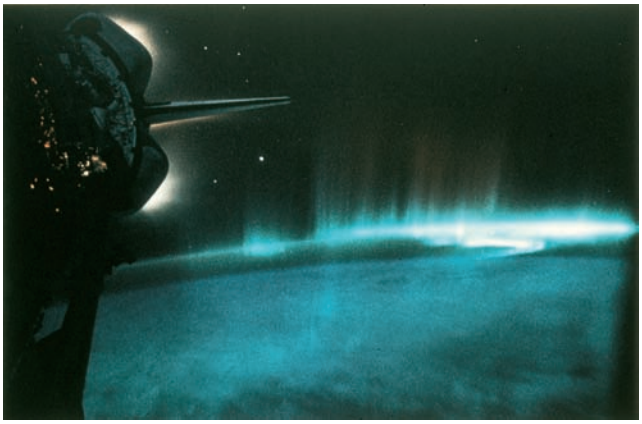


Fig. 3.22 Aurora australis The eerie, beautiful glow of auroras can be detected from space, as shown in this image of the *aurora australis*, or Southern Lights, taken from the *Space Shuttle Discovery*. The colored emission of atomic oxygen extends upward to between 200 and 300 kilometers above the Earth's surface. (Courtesy of NASA.)

toward the pole, and constantly changing in brightness. Such ever-changing aurora ovals are created simultaneously in both hemispheres and can be viewed at the same time from the Moon.

Visual auroras normally occur at 100 to 250 kilometers above the ground. This height is much smaller than either the average radius of the oval, at 2250 kilometers, or the radius of the Earth, 6380 kilometers. An observer on the ground therefore sees only a small, changing piece of the aurora oval, which can resemble a bright, thin, wind-blown curtain hanging vertically down from the Arctic sky.

Energetic electrons bombarding the upper atmosphere cause most aurora on the Earth. The reason that aurora are usually located near the polar regions is that the Earth's magnetic fields guide the energetic electrons there. Electrical currents as great as a million amperes can be produced along the aurora oval, and the electric power generated during the discharge is truly awesome – about ten times the annual consumption of electricity in the United States.

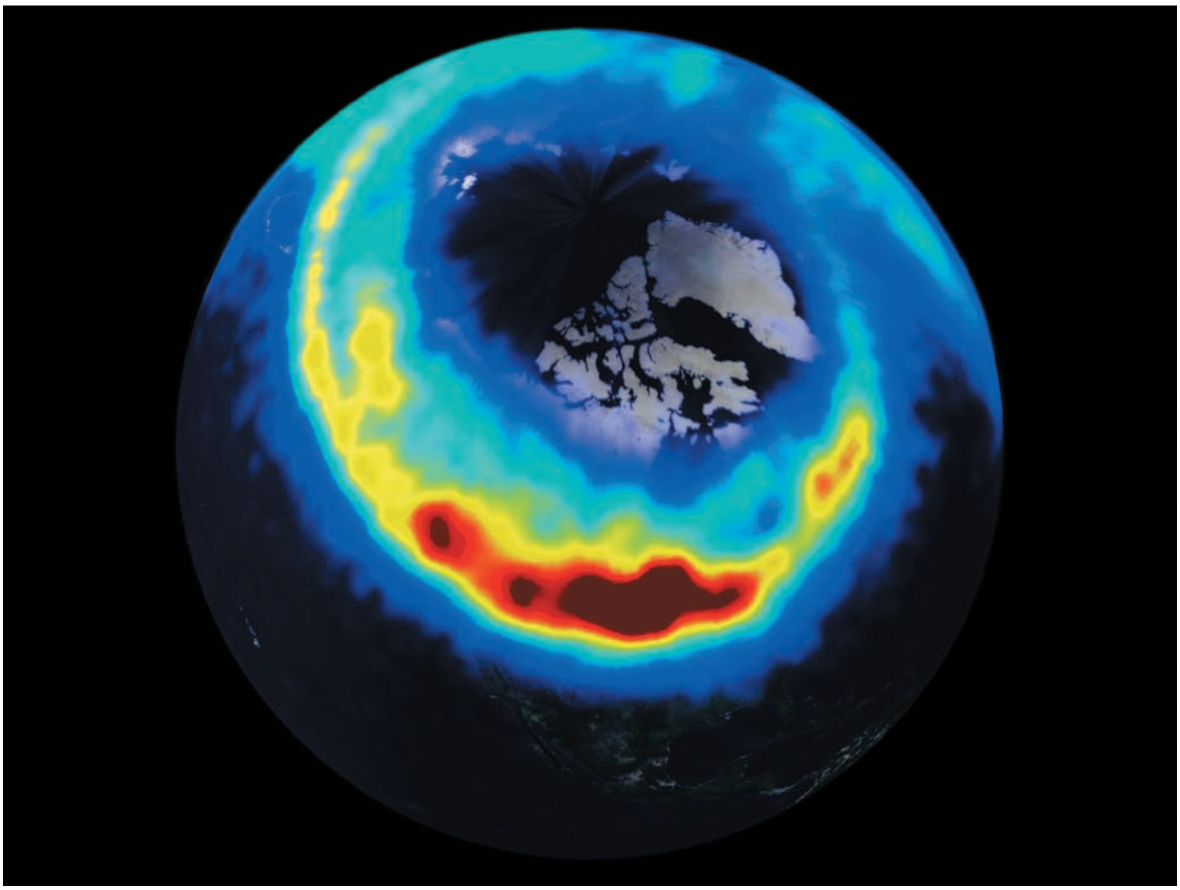


Fig. 3.23 The aurora oval The *POLAR* spacecraft looks down on an aurora from high above the Earth's north polar region in February 2000, showing the Northern Lights in their entirety. The glowing oval is 4500 kilometers across. The most intense aurora activity appears in bright red or yellow. (Courtesy of NASA/U. Iowa.)

Table 3.6 Frequent spectral features in the aurora emission

Wavelength (nanometers)	Emitting atom, ion or molecule	Altitude (kilometers)	Visual color
391.4	N^+ (nitrogen ion)	1000	violet-purple
427.8	N^+ (nitrogen ion)	1000	violet-purple
557.7	O (oxygen atom)	90–150	green
630.0	O (oxygen atom)	>150	red
636.4	O (oxygen atom)	>150	red
661.1	N_2 (nitrogen molecule)	65–90	red
669.6	N_2 (nitrogen molecule)	65–90	red
676.8	N_2 (nitrogen molecule)	65–90	red
686.1	N_2 (nitrogen molecule)	65–90	red

When the fast-moving electrons slam into the upper atmosphere, at speeds of about 50 kilometers per second, they collide with the oxygen and nitrogen atoms or molecules there and excite them. The pumped-up particles quickly give up the energy they acquired from the electrons, emitting a burst of color in a process called fluorescence. It is something like electricity making the gas in a neon light shine or a fluorescent lamp glow.

The color of the aurora depends on which atoms or molecules are struck by the precipitating electrons, and the atmospheric height at which they are struck (Table 3.6). Low-altitude oxygen atoms produce green, a common aurora color, while the high-altitude oxygen atoms cause the rare all-red aurora. Nitrogen molecules create low-altitude red light, below the oxygen's green, while nitrogen ions can produce violet-purple light at high altitudes. The green oxygen emission appears at about 100 kilometers and the red oxygen light at 200 to 400 kilometers. At these heights, the aurora shines from the ionosphere, an electrically conducting layer in the Earth's upper atmosphere.

Even though changing conditions on the Sun may trigger exceptionally intense Northern and Southern Lights, we now know that the electrons that cause some of the everyday aurora arrive indirectly at the polar regions from the Earth's magnetic tail, and that these electrons can also be energized locally within the magnetosphere. As the solar wind flows past the Earth, terrestrial magnetic fields can capture and store energy from the winds, and the solar-wind magnetic fields can merge, or reconnect, with the terrestrial magnetic fields. The stretched-out magnetism eventually gets overloaded with too much energy, temporarily pinching off the Earth's magnetotail.

During this magnetic reconnection process, the magnetic fields heading in opposite directions – having opposite north and south polarities – break and reconnect at 140 000 to 160 000 kilometers downwind of Earth on its nightside. Electrons are pushed up and down the tail, and can be accelerated within the magnetosphere as they travel back toward the Earth and into its polar regions. The electrons that are thrown Earthward follow the path of magnetic field lines, which link the magnetotail to the polar regions and map into the aurora oval.

Magnetic reconnection between the solar-wind magnetic fields and the Earth's magnetic fields can also occur on the dayside of the Earth, facing the Sun, and this can also open a valve that lets the solar-wind energy cross into the magnetosphere. In this event, high-energy electrons also follow the path of magnetic field lines into the polar regions where they produce the aurora and create the aurora oval.

Although most aurora are caused by collisions between high-energy electrons and the atmosphere, protons can also be funneled down along the polar magnetic fields and sometimes cause aurora. The ultraviolet emission of such a proton aurora was, for example, recorded from the *IMAGE* spacecraft in 2000. Solar-wind protons, which enter through rare temporary openings in the Earth's magnetic field, cause them.

The rare bright aurora seen at low latitudes in more clement climates tend to occur when the Sun is near the peak of its 11-year magnetic activity cycle, when sunspots are most numerous; such exceptionally bright aurora occur less often at the minimum of the cycle when there are few sunspots. The sunspots do not themselves cause the intense aurora, but are instead a sign of strong magnetic fields on the Sun, which can explosively eject huge magnetic bubbles known as coronal mass ejections. At maximum activity, the Sun emits coronal mass ejections more frequently, and when they chance to hit the

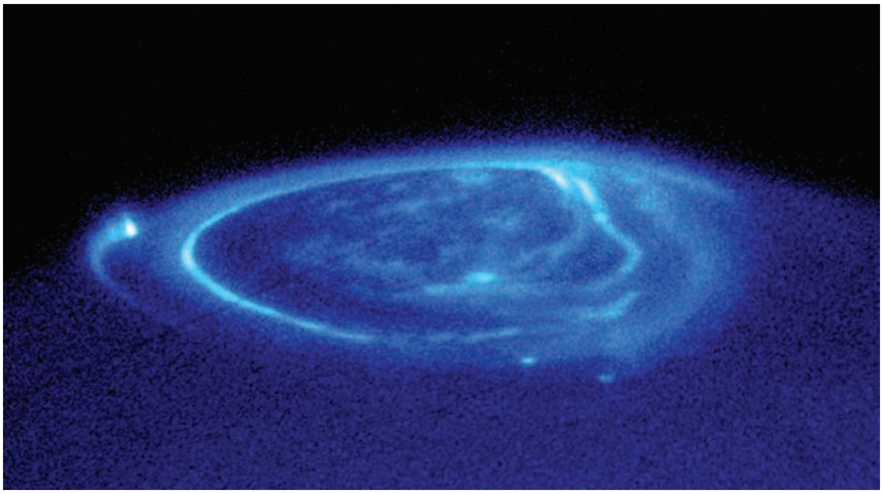


Fig. 3.24 Jupiter's aurora The aurora oval or ring over the north polar region of Jupiter, as imaged from the *Hubble Space Telescope* in ultraviolet light in September 1997. High-energy electrons and ions cascade into Jupiter's upper atmosphere and create the bright ultraviolet aurora. Several of the bright spots are believed to mark the magnetic "footpoints" of three of Jupiter's largest moons. The footpoints are locations where powerful streams of electrons follow the magnetic fields of Jupiter's magnetosphere from the moons down into Jupiter's atmosphere. (Courtesy of NASA/STScI/U. Michigan.)

terrestrial magnetic field with the right magnetic orientation, they reconnect with it and compress the Earth's magnetic field. The aurora ovals then intensify and spread down as far as the tropics in both hemispheres. So it is really the Sun that controls the intensity of the brightest, most extensive aurora, like the dimming switch of a cosmic light.

Aurora on Jupiter and Saturn

Charged particles are funneled into the magnetic polar regions of Jupiter and Saturn, producing aurora ovals that shine in ultraviolet light, rather than Earth's green, red or violet. The atmospheres of these giant planets are primarily composed of hydrogen, unlike Earth's oxygen and nitrogen, and their ultraviolet auroras are emitted when charged particles stream into and excite the atomic and molecular hydrogen. Moreover, many, but not all, of the aurora at Jupiter and Saturn are caused by particles trapped in their immense magnetospheres, rather than being directly connected to solar-wind particles.

Like their terrestrial counterparts, the curtains of light on Jupiter are found in two oval-shaped regions circling the magnetic poles of the planet, just above the clouds. The aurora glows are produced in these high-latitude regions because that is where the magnetic fields direct electrically charged particles: electrons, protons and other ions. When these particles hit the planet's upper atmosphere, they collide with atoms and molecules there, leaving them in an excited state. As on Earth, the atoms and molecules release the extra energy in the form of light, and return to their normal state. But unlike Earth's colored light show, Jupiter's aurora ovals were first observed from space at ultraviolet wavelengths (Fig. 3.24).

Jupiter's aurora is the most powerful in the solar system. At about 10^{14} watts, it is typically one thousand times more powerful than Earth's aurora. The Jovian lights are powered largely by energy extracted from planetary rotation, although there seems to be a contribution from the solar wind. Thus, internal processes seem to be the dominant source of power for Jupiter's aurora. This contrasts with Earth's aurora, which is mainly generated externally through the interaction of the solar wind and the terrestrial magnetosphere.

Jupiter's satellite Io affects the aurora on the planet. Electrons and ions spewed out by volcanoes on Io are captured by the intense, rapidly rotating magnetic field and spiral inward at high energies toward the planet's polar regions. As the rotating magnetic field sweeps past Io, an invisible current of charged particles flows along Jupiter's magnetic field lines into the polar regions, producing bright trails in the ultraviolet images (Fig. 3.24).

On Jupiter one can normally see a main ultraviolet oval, and in addition bright swirling streaks are sometimes detected both within and outside the oval (Fig. 3.24). They have been attributed to electric currents from three of the planet's large moons, Io, Ganymede and Europa. Very intense bursts of aurora activity have also

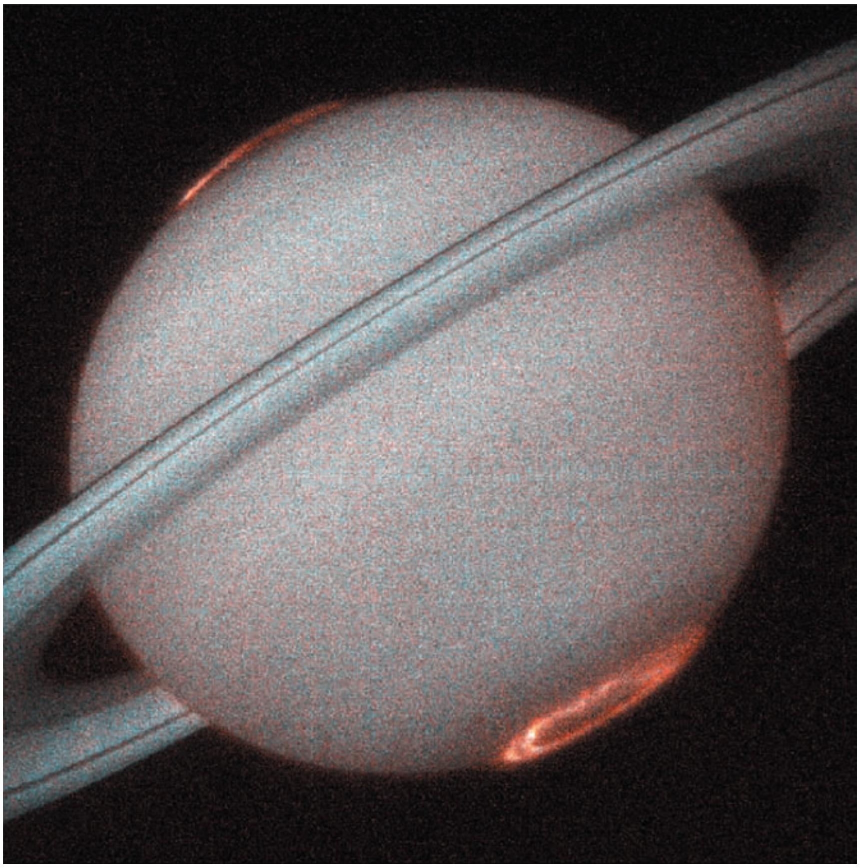


Fig. 3.25 Aurora on Saturn High-energy electrons and ions are captured from the solar wind and funneled down into Saturn's upper atmosphere, creating aurora ovals at its northern (*upper left*) and southern (*lower right*) magnetic poles. This ultraviolet image was recorded from the *Hubble Space Telescope* in October 1997. The bright red aurora features are dominated by emission from atomic hydrogen, while the white regions within them are emitted by molecular hydrogen. (Courtesy of NASA/STScI/JPL.)

been detected; they are apparently regulated by the variable solar wind, perhaps because it affects the size of Jupiter's magnetosphere.

Saturn's ultraviolet aurora (Fig. 3.25) is most likely caused when the gusty solar wind sweeps over the planet, perhaps like the Earth's aurora. But unlike the Earth, Saturn's aurora oval has only been seen from spacecraft in ultraviolet light, at least so far. It could not be detected from beneath the Earth's atmosphere that absorbs the ultraviolet.

In Chapter 4 we turn our attention to our home planet, Earth, third rock from the Sun.

Part 2 The inner solar system: rocky worlds

4 Restless Earth: third rock from the Sun

- Seismic waves generated by earthquakes have been used to look inside the Earth, determining its internal structure.
- There is a crystalline globe of solid iron at the center of the Earth that spins faster than the rest of the planet. This inner solid core is suspended in a much larger, fluid, outer core of molten iron, which is itself encased in a thick mantle of solid rock.
- The continents disperse and then reassemble, over and over again, roaming about the planet in an endless journey.
- Sound waves and gravitational data have been used to effectively empty the Earth's oceans and see their floors, revealing an underwater range of active volcanoes that snakes its way around the middle of the ocean floor.
- The bottom of the oceans remains in eternal youth as new floor spills out of mid-ocean volcanoes and old floor is pushed back inside the Earth, but the water above the floors has remained for billions of years, shifting about the globe as new oceans open up and old ones close.
- The outer part of the Earth is broken into a mosaic of large plates, like the cracked pieces of an eggshell; these plates move across the Earth at the rate of a few centimeters per year, or about as fast as your fingernails grow.
- Wheeling, churning motions deep inside the Earth's hot interior move continents sideways all across the planet.
- The Earth's moving plates squeeze oceans out of existence, grind against each other to create earthquakes, and dive into the Earth to produce volcanoes that make continents grow at their edges.
- Boston and Italy were once part of Africa, a glacier of ice once covered the Sahara Desert, and the Pacific Ocean once washed against the shores of Colorado.
- A colossal alp can erode away into a small, round knob of a hill in just a few hundred million years, while continents can also weld together to form new mountain ranges.
- The Earth's upper atmosphere is heated and ionized by the Sun's variable X-ray and extreme ultraviolet radiation.

- Ultraviolet radiation from the Sun creates the protective ozone layer in the stratosphere of the Earth's atmosphere.
- Synthetic chemicals called chloroflurocarbons (CFCs) have been destroying the thin layer of ozone that protects human beings from dangerous solar ultraviolet radiation. The production of these ozone-destroying chemicals was outlawed in 1987 by an international agreement named the *Montreal Protocol*.
- Invisible gases help to warm the Earth by trapping the Sun's heat and preventing some of it from being reflected back into space. This process is commonly known as the greenhouse effect.
- Warming of the Earth's surface and lower atmosphere by the greenhouse effect keeps the Earth from becoming a frozen ball of ice.
- Carbon dioxide and other heat-trapping gases, such as methane and nitrous oxide, have been increasing in the Earth's atmosphere for more than a century as the result of human activity.
- By burning coal and oil, humans have increased the amount of carbon dioxide in the Earth's atmosphere by 30 percent since the industrial revolution.
- Rising seas, retreating glaciers, melting ice caps, and increasing sea and air temperatures are all recent signs of global warming from increased emissions of heat-trapping gases.
- If current emissions of carbon dioxide and other greenhouse gases go unchecked over the next 100 years, global warming could produce agricultural disaster in the world's poorest countries, rising seas with coastal flooding throughout the world, and the spread of diseases carried by mosquitoes.
- An international agreement to limit the emission of heat-trapping gases was made in December 1997. Known as the *Kyoto Protocol*, it has had a limited effect on curbing global warming because it has not been ratified by China or the United States, two of the main climate-altering polluters.
- The world's most influential science academies have warned national leaders that global warming from emissions of carbon dioxide and other heat-trapping gases poses a clear and increasing threat.
- The 2007 Nobel Peace Prize was awarded to an Intergovernmental Panel on Climate Change and to Albert Gore Jr. for their contributions to knowledge about man-made climate change and for laying foundations to measures needed to counteract the change.
- The Copenhagen Summit in December 2009 sought international consensus on ways to combat global warming, but it did not result in any legally binding treaty on limiting carbon-dioxide emissions. Both China and the United States refused to accept such mandatory limits, but agreed with a hypothetical climate-change accord that has voluntary curbs and varying emission reductions for different countries.
- The major ice ages, which repeat every 100 000 years, are caused by astronomical rhythms that alter the angles and distances from which sunlight strikes the Earth.
- The Sun is slowly getting brighter as time goes on. It will become hot enough in 3 billion years to boil the Earth's oceans away, and 4 billion years thereafter our star will balloon into a giant star, engulfing the planet Mercury and becoming hot enough to melt the Earth's surface.

- Space weather refers to conditions on the Sun and in the Sun's winds, the Earth's magnetosphere, and the Earth's outer atmosphere that can influence the performance and reliability of space-borne and ground-based technological systems, and can affect human life and health.
- Explosive outbursts of solar flares and coronal mass ejections from the Sun can cripple spacecraft and seriously endanger unprotected astronauts that venture into outer space. Sun storms can also disrupt global radio communications and disable satellites used for navigation, military reconnaissance or surveillance, and communication, from cell phones to pagers, with considerable economic, safety and security consequences.
- Solar protons are the most energetic and therefore the most dangerous solar energetic particles. They can severely affect the health of unprotected astronauts traveling outside the Earth's magnetosphere, and they are capable of penetrating spacecraft to damage or disrupt sensitive technical systems. The strongest events produce radiation doses that might be lethal to astronauts fixing a spacecraft in outer space or taking a walk on the Moon or Mars.
- Interplanetary magnetic clouds travel behind interplanetary shocks, which are driven by coronal mass ejections. Such a magnetic cloud contains a well-organized, twisted magnetic flux tube, which can provide a "highway" for the transport of solar energetic particles.
- When encountering Earth with the right magnetic alignment, coronal mass ejections can trigger intense geomagnetic storms, accompanied by exceptionally bright aurora, and compress the magnetosphere, exposing geosynchronous satellites to the full force of the solar wind.
- Solar X-rays and extreme ultraviolet radiation both produce and significantly alter the Earth's ionosphere. The solar X-rays fluctuate in intensity by two orders of magnitude, or a factor of 100, during the Sun's 11-year magnetic activity cycle. Near activity maximum, greater amounts of X-rays produce increased ionization, greater heat, and expansion of the Earth's upper atmosphere, altering satellite orbits and disrupting communications.

4.1 Fundamentals

Our home planet Earth is larger and denser than the other terrestrial planets, and the only one with large oceans of liquid water. It revolves about the Sun once a year, at a mean distance of one astronomical unit, and rotates on its axis to view the same star every 23 hours 56 minutes and 4 seconds, the Earth's sidereal rotation period. Table 4.1 summarizes the physical properties of the Earth.

4.2 Journey to the center of the Earth

Looking inside the Earth's hidden interior

The internal structure of the Earth can be mapped with the help of earthquake waves. The Greek word for earthquake is *seismos*, meaning "to quake or tremor". Today, earthquake waves are often called *seismic waves*, and the study of earthquakes is known as *seismology*.

Earthquakes that originate in the planet's outer shell set the seismic waves in motion, and their velocities are determined by the density, temperature and chemical composition of the rocks they travel through. The waves become sluggish in hot, low-density rock, and they speed up in colder, denser regions. When moving between materials of differing physical properties, the seismic waves change their speed and direction of movement, enabling seismologists to determine boundaries between the Earth's internal layers.

The seismic investigations indicate that the Earth is layered inside like a peach. Its deeper layers are denser, and they are separated from one another in sharp transitions. There are three major parts: (1) the rocky *crust*, (2) a *mantle* of hot, plastic rock, and (3) the dense *core* (Fig. 4.1). They are the skin, pulp, and pit of the Earth, so to speak.

Table 4.1 Physical properties of the Earth

Mass	5.972×10^{24} kilograms
Mean radius	6371 kilometers
Bulk density	5513.4 kilograms per cubic meter
Sidereal rotation period	23 hours 56 minutes 4 seconds $= 8.6164 \times 10^4$ seconds $= 0.997\,27$ days
Sidereal orbital period	1 year = 365.24 days $= 3.1557 \times 10^7$ seconds
Mean distance from Sun	$1.495\,98 \times 10^8$ kilometers $= 1.000$ AU
Orbital eccentricity	0.0167
Tilt of rotational axis, or obliquity	23.27 degrees
Age	4.6×10^9 years
Atmosphere	77 percent nitrogen, 21 percent oxygen
Surface pressure	1.013 bars at sea level
Surface temperature	288 to 293 kelvin
Magnetic field strength	0.305×10^{-4} tesla at the equator
Magnetic dipole moment	7.91×10^{15} tesla meters cubed

Table 4.2 The five most abundant elements in the Earth

Element	Symbol	Average abundance (percent by mass)
Iron	Fe	34.6
Oxygen	O	29.5
Silicon	Si	15.2
Magnesium	Mg	12.7
Nickel	Ni	2.4

The core has a liquid outer component and a solid inner one.

These different internal layers can be distinguished by the different chemical composition of their rocks. Most of the rocks of the mantle consist of minerals in which silicon (Si) and oxygen (O) are linked to other atoms. Such minerals are known as *silicates*. The core is composed mainly of iron (Fe) with some nickel (Ni). It is made up of an outer core of liquid molten iron and an inner core of solid iron. The abundance of these ingredients in planet Earth is given in Table 4.2.

Most earthquakes occur just beneath the Earth's surface at depths of no more than 700 kilometers, when massive blocks of rock grind, lurch and slide against one another. But they shake the Earth to its very center, at 6371 kilometers below the surface, causing the planet to vibrate and ring like a bell. The reverberations resemble ripples spreading out from a disturbance on the surface of a pond. These seismic waves move in all directions and their arrivals at various places on the Earth can be detected by seismometers. By combining the arrival times of different seismic waves that have traveled through the Earth's interior to various points on the surface, seismologists can determine the hidden interior structure of the Earth (Fig. 4.2).

Rock layers of different density and stiffness will propagate the waves at different speeds, much the way that a tightened violin string will sound at a higher pitch. As a

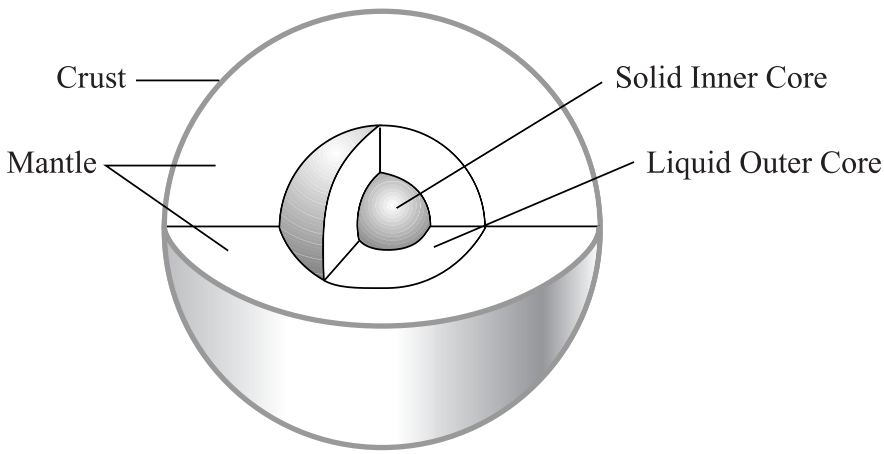


Fig. 4.1 Crust, mantle and core A relatively thin, rocky crust covers a thick silicate mantle. They overlie a liquid outer core, composed mainly of molten iron, and an inner core of solid iron. These nested layers have been inferred from seismic waves that travel through the Earth, changing velocity and direction at the layer boundaries.

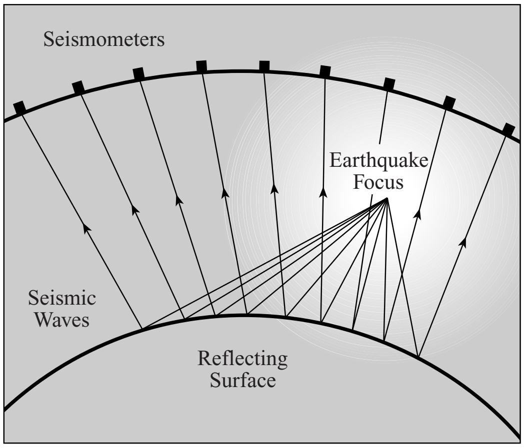


Fig. 4.2 Measuring earthquakes When an earthquake occurs beneath the surface of the Earth, it becomes the focus of seismic waves that travel through the Earth. Seismometers on the surface of the Earth record the arrival of the waves, and locate the position of the boundaries between internal layers of different composition, density, pressure and temperature. Seismic waves known as S-waves (or "shear and shake" waves) cannot pass though a fluid, and are reflected by it. The reflected waves shown here mark the boundary of the Earth's liquid outer core.

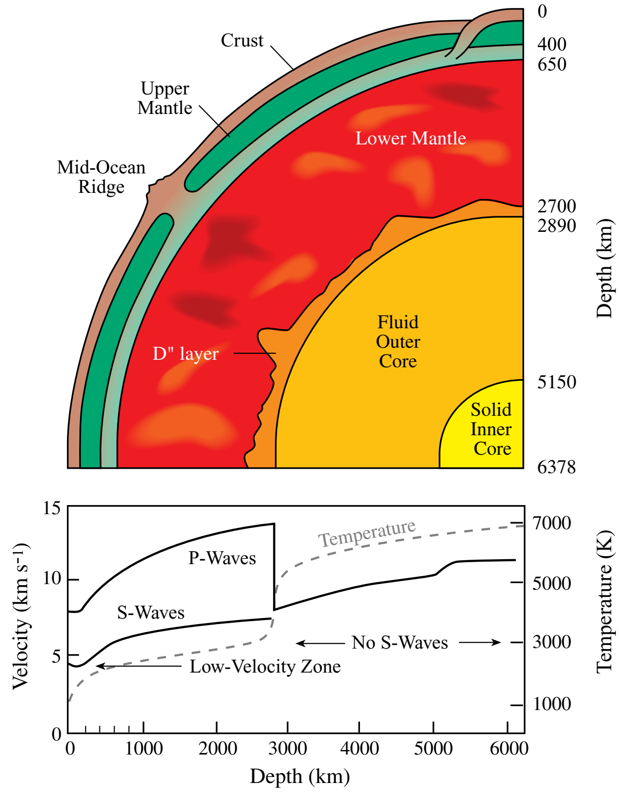


Fig. 4.3 Layered structure of the Earth The Earth's internal structure is determined by the varying velocity of earthquake waves. There are two types of waves that travel through the Earth. They are known as the compression P-waves (or "push and pull" waves) and the shear S-waves (or "shake" waves). The P-waves move almost twice as fast as the S-waves, and the P-waves pass through the fluid outer core, which the S-waves cannot do. The boundary between the mantle and core is marked by a precipitous drop in the velocity of the P-waves at a depth of about 2890 kilometers. The S-waves do not propagate beyond this boundary. The liquid outer core is separated from the solid inner core at a radius of 1220 kilometers where the P-waves increase in velocity.

result, the paths of seismic waves are bent and focused by their passage through the Earth's interior.

By careful mapping of the patterns of many earthquakes that travel to different depths, seismologists have peeled away the Earth's outer layers and looked at various levels within it (Fig. 4.3; Focus 4.1). It is similar to using an ultrasonic scanner to map out the shape of an unborn infant in a mother's womb. Seismology is also somewhat like using computed axial tomography (CAT) scans to derive clear views of the insides of living bodies from the numerous readings of X-rays that cross through the body from different directions.

The crust and mantle of the Earth

The outer skin of the Earth, its crust, is a thin veneer of rocky material that covers the planet like the lumpy and split crust of an apple pie. At the base of the Earth's crust lies the Mohorovicic discontinuity, a tongue-twisting name shortened by most geologists to "Moho".

The Moho separates the dense mantle from the light crust. The boundary lies at a depth of about 5 kilometers from the ocean bottom and 35 kilometers below most places on continents, but as much as 60 kilometers below mountains. So the crust is thinnest under the oceans and thickest under the continents. Because the continental and oceanic crusts are both less dense than the underlying material, they both tend to float on the mantle. High mountains have deep crustal roots that provide buoyancy and keep them afloat, much the way icebergs float on the ocean.

The Earth's buoyant crust is made up of two different materials. There is the oceanic crust, which is made of the black, shiny, volcanic rock known as basalt, and the continental crust that contains granite. The ocean floor covers more than half the Earth's surface and has been produced by an outpouring of lava from volcanoes at the bottom of the sea. Volcanic islands like Hawaii and Iceland are also largely composed of basalt. The tough continental granites were formed in the fiery melts of magma, and include hard, colorless quartz.

The bulk of the Earth is in its mantle, the region that reaches down some 2890 kilometers, on average, from the

Focus 4.1 Taking the pulse of the Earth

Earthquakes produce three types of seismic waves. There are the *P-waves* and *S-waves* that travel into the Earth, and the *surface waves* that propagate around it. The P-waves consist of compression pulses through the Earth, expanding and compressing the rocks. They are analogous to sound waves in air, although the vibrations of the P seismic waves are much slower than audible sound. The P-waves arrive at monitoring stations before the S-waves, which set the Earth vibrating at right angles to the path of the waves, advancing like snakes. The P stands for "push and pull", while the S denotes "shear or shake".

The P-waves can propagate through every part of the Earth, even its center. A large portion of them penetrates the deep interior and then re-emerges toward the surface on the other side. The S-waves do not travel in a fluid; they propagate only in resilient, solid substances that have elastic resistance to twisting.

In 1906, the British geologist Richard D. Oldham (1858–1936) found that at a certain depth the P-waves slowed sharply and the S-waves couldn't propagate. These changes mark the bottom of the mantle and the top of the Earth's liquid outer core. The core–mantle boundary is sometimes called the *Gutenberg discontinuity*, after Beno Gutenberg (1889–1960), from the California Institute of Technology, who made the first accurate determination of its depth at an average of 2890 kilometers. In 1909, the Croatian geologist Andrija Mohorovicic (1857–1936) discovered that the speed of seismic waves increases at 35 to 60 kilometers below some continents. This *Mohorovicic discontinuity* marks the place where the crust ends and the mantle begins. The inner core, with a radius of about 1220 kilometers, was discovered in 1936 by the Danish seismologist Inge Lehmann (1888–1993).

thin crust to the top of the core. The difference between the crust and the mantle is one of chemical composition. Material brought up by volcanic eruptions, as well as eroded mountains, indicate that the upper mantle is composed of dense minerals known as olivine and pyroxene, which are silicates with a little magnesium or iron mixed in as minor constituents.

Lithosphere and asthenosphere of the Earth

The outermost parts of the Earth can be divided by their physical properties into the lithosphere and asthenosphere. The lithosphere is the solid region beneath the familiar oceans and mountains. It extends to depths of about 100 kilometers, which includes both the crust and upper mantle. The lithosphere consists of rocky crust and mantle down to a zone in the mantle that is lubricious enough to move. Beneath the lithosphere lies the warm and plastic asthenosphere that reaches to a depth of about 300 kilometers. Its material is revealed by the slowness with which it propagates seismic waves.

The distinction between the lithosphere and asthenosphere is one of stiffness. The lithosphere takes the root of its name from the Greek *lithos*, for "stone". The lithosphere is the solid "plate" of the plate tectonic theory mentioned later in this section. The word "asthenosphere" comes from the Greek *asthenos*, meaning "without strength" or "devoid of force".

The radioactive elements responsible for the warmth of the asthenosphere are too weakly concentrated to melt the rock, but they cause it to soften and behave like putty. Rock in the asthenosphere flows slowly when strained for a long time, like applying slow pressure to an open tube of toothpaste, but the asthenosphere responds like a solid when it is struck by an earthquake.

Two cores of the Earth

If you pick up a typical rock in the Earth's crust and determine its mass density, it will be roughly 3000 kilograms per cubic meter, or about three times that of water. By way of comparison, the mean mass density of the Earth is 5513.4 kilograms per cubic meter, which means that there must be high-density material located deep inside the Earth. The fact that the Earth is not homogeneous, with its densest parts concentrated inside, has been known since the time of Isaac Newton (1642–1727), from the varying gravitational pull measured by pendulums located at different places on the Earth's surface. The material with the greatest density is concentrated in the planet's core, with a mass density of about twelve times that of water.

The Earth's core reaches about halfway to the surface, implying a volume that is one-eighth that of the entire Earth. If the mass density of the Earth were uniform, the core would have an equal share, one-eighth, of the mass of the Earth, but its actual mass is nearly three times greater. This points to iron as the most likely core material, since it is also the most abundant heavy element in the Sun and in some meteorites.

Laboratory measurements also show that the densities and seismic-wave velocities of the core are more closely matched by iron than any other element. The seismic evidence indicates that the core is less dense than pure iron would be at the pressures there. Although the core is mostly iron, it must consist of an alloy of iron that includes light elements, one of which may be hydrogen.

By weight, the Earth is mostly iron, but relatively little of the metal is found in the Earth's crust. It is principally made of lighter elements like silicon and oxygen. Billions of years ago, during the planet's very early history, the Earth must have been molten, permitting the iron to sink to the interior because of its enormous weight. The planet would have then cooled from the outside, forming solid rocks in the crust and mantle that consist of lighter elements that are locked together and do not sink into the molten core.

Examination of earthquake waves has shown that there are two cores: an inner crystalline solid core and an outer fluid one. Compressed by the immense weight of the overlying material, the inner core solidified, while the outer core remained liquid. The two cores are very different in size. The solid inner core has a radius of about 1220 kilometers, which is slightly smaller than the Moon whose radius is 1740 kilometers. The outer fluid core is about 3490 kilometers in radius, or 55 percent of the Earth's radius.

The seismic evidence indicates the presence of a rugged, interactive zone, known as the D″ layer, where the liquid-metal outer core meets the lowermost part of the rocky mantle. This turbulent, irregular boundary contains troughs and swells, deeper than the Grand Canyon and higher than Mount Everest, spreading across continent-sized areas. It may represent material that was once dissolved in the underlying fluid core, dense material that sank through the mantle but could sink no further down, or material formed as a result of chemical reactions between the core and mantle. Seismologists speculate that irregularities in the D″ layer at the core–mantle boundary may channel heat flows to produce giant rising plumes of molten rock capable of penetrating the thick mantle and occasionally making their way to the Earth's surface.

The temperature of the deep core is difficult to determine, but we certainly know that the planet is hot inside. The central inner core appears to be about 6900 kelvin, which is a bit hotter than the visible disk of the Sun at 5780 kelvin. At first glance, this would seem to imply that the center of the Earth must be liquid, but this is contradicted by seismic evidence, which indicates a solid region in the deep interior. The clue to the apparent paradox is the high pressure at the center of the Earth, about 3.6 million times the pressure of the atmosphere at sea level. These pressures have been imitated in laboratory experiments, and they lead to a remarkable conclusion. At high pressures, iron can persist as a fairly rigid solid even at a temperature of thousands of kelvin.

Most liquids will solidify if the pressures are high enough and the temperatures are relatively low. Probably the entire core was once molten, but the drop in temperature associated with a loss of heat permitted the inner portion to solidify under the high pressures. The pressures are low enough, and the temperatures still high enough, to sustain a liquid outer core. It also remains liquid because iron alloys melt at lower temperatures than most rocks.

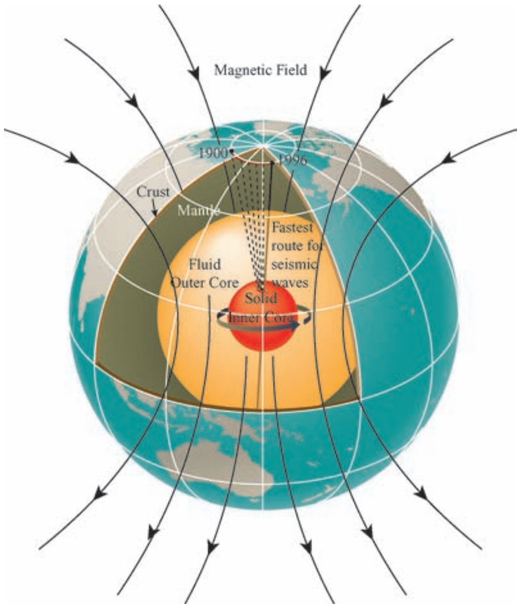


Fig. 4.4 The Earth's double core The mantle and part of the crust have been cut away here to show the relative sizes of the Earth's fluid and solid cores. The outer fluid core is about 55 percent of the radius of the Earth, and the inner solid core is slightly smaller than the Earth's Moon. The Earth's magnetic field is thought to be generated and sustained by moving currents in the planet's electrically conducting, fluid outer core, which is composed of molten iron. Geophysicists have discovered that the route of the rapid polar (north–south) waves through the Earth's interior is gradually shifting eastward because the inner core is rotating slightly faster than the rest of the planet. The fast rotation of the inner solid core may help explain how Earth's magnetic field reverses polarity. (Courtesy of Paul Richards, Lamont-Doherty Earth Observatory.)

As our planet grows older and colder, the solid inner core is growing continuously at the expense of the liquid outer core. The iron liquid at the base of the fluid outer core is freezing, solidifying and snowing on the surface of the solid inner core, making it grow at the rate of about 0.01 meters every 100 years. At the same time, the rocky mantle may be slowly dissolving into the liquid metal of the outer core.

The Earth's inner core is a solid lump of iron suspended at the center of the much larger, fluid outer core, something like a golf ball levitated in the middle of a fish bowl (Fig. 4.4). The faint seismic vibrations that pierce the inner core move through it at different speeds that depend on their direction, faster on polar north–south paths than equatorial east–west ones. This directional dependence of seismic-wave velocities is explained by the crystalline structure of the inner core. The crystals give the solid inner core a texture with a preferred orientation, like the grain in wood. By lining up along the Earth's rotation axis, iron crystals make the inner core stiffer along this axis, thus making sound waves travel faster in this direction. Some scientists have even speculated that the inner core may be just one, single, gigantic crystal of iron atoms rather than a mass of tiny crystals, like a huge diamond suitable for an interplanetary engagement. In either case, each crystal probably takes its direction from either the stress generated by Earth's rotation or from the terrestrial magnetic field.

Recordings of weak earthquake rumbles, which have traveled through the central core of the Earth, indicate that it spins faster than the outer Earth, but that they both rotate in the same direction. The fast lane for seismic waves is tipped slightly with respect to the Earth's north–south axis, and it moves around it (Fig. 4.4). This shift in orientation means that the crystalline globe at the center of the Earth is turning slowly within its liquid metal enclosure. It is spinning with respect to the Earth's surface at between 0.2 and 0.3 degrees per year, completing one lap in between 1200 and 1800 years.

The Earth's magnetic field threads the solid inner core and could make it turn faster, much as a magnetic field turns the shaft of an electric motor. The magnetic coupling between the two cores could also account for reverses in the Earth's magnetic polarity; the planet switches its north and south magnetic poles a few times every million years. Currents in the electrically conducting, fluid outer core generate and maintain the magnetic fields, and turbulence in the fluid is always trying to toss the magnetism into a polarity reversal. The inner solid core exerts a stabilizing influence on this tendency, forcing the fields to stay in place, but the magnetic connection between the two cores is probably pulled apart as they rotate with respect to each other. The coupling eventually gives way and the magnetism flips.

Origin of the Earth's layered interior

The origin of the layered structure of the Earth's interior is still a geological mystery, but there are two extreme alternatives. According to one theory, the Earth accumulated rapidly (in 100 000 to 10 million years), and the kinetic energy of the impacting material that coalesced to form the Earth kept the planet hot and molten as it formed. If the rocks were molten as the Earth grew, its constituents would separate, with the dense, heavy material sinking toward the interior, creating the dense core, and the light substances rising to the surface to form the low-density mantle and crust.

In the alternative scenario, the Earth gathered itself together relatively slowly, in 100 million to 1 billion years, and the planet started out cold, homogeneous and solid. The globe then became heated by emission from radioactive material that was uniformly distributed through the interior, and its temperature gradually rose to the melting point. When the planet melted, the heavy elements fell toward the center, forming a dense core, while the lighter elements rose toward the surface, producing chemically distinct layers.

In both the hot and cold theories, the layered internal structure of the Earth results from a process known as differentiation, in which gravity separates elements in a molten state, pulling the heavier ones down. A similar thing takes place in a blast furnace or smelter. Slag-forming rock is loaded into the furnace, and molten metal is tapped periodically from the bottom. Thus, in both theories the Earth was once molten and after a process of gravitational separation or differentiation, the Earth began cooling from the outside. The solid crust formed and then the mantle, and the basic layered structures remained a feature of the Earth since its early history.

4.3 Remodeling the Earth's surface

Earth's continents, oceans and ocean floors

There are two major types of terrain on Earth – the high, dry continents and the low, wet floor of the ocean (Fig. 4.5). Between them, and partially surrounding many continents, is a narrow strip of shallow ocean called the continental shelf. Today, the oceans cover 71 percent of the Earth's surface, and the world's continents amount only to scattered and isolated masses surrounded by water.

To those of us who are confined near the surface of the globe, the Earth seems rugged, with towering mountains rising several kilometers above the ocean (Fig. 4.6). But a scale model of the Earth would have to be quite smooth. The highest and lowest places reach only one-tenth of one percent, or 0.001, of the Earth's radius above and below the ocean surface. A basketball this smooth would have bumps no more than 0.0001, or 10^{-4}, meters high, roughly the size of the dot at the end of this sentence.

The smoothness of the Earth is due to the immense force of its gravity and the weight of its outer layers, which largely overcome the electrical force inside the solids making up the Earth and cause them to lie in concentric shells. In smaller bodies, such as asteroids less than a few hundred kilometers in diameter, the interior is strong enough to remain rigid and they retain their original irregular shapes and internal composition.

Moreover, ongoing erosion will wear down the world's highest mountains in just a few hundred million years, which is just a fraction of the Earth's age of 4.6 billion years. The Rocky Mountains were, for example, once

Fig. 4.5 Planet Earth from space As illustrated in this image of Africa, the Arabian Peninsula, and the Indian Ocean, the Earth's surface consists of continents and oceans. Continents cover a little more than one-quarter of the Earth's surface, while ocean water covers almost three-quarters of the surface. Our home planet has a thin atmosphere with white clouds of water ice, and enough transparency that you can usually look right through it. The Antarctica ice cap gleams white at the bottom. *Apollo 17* astronauts took this image in December 1972 as they left the Earth en route to the Moon. (Courtesy of NASA.)

Fig. 4.6 Topography of the Earth's landmass Global topographic image of land on the Earth is displayed by both colored height and shaded relief in this Mercator projection. Color-coding is directly related to topographic height, with green at the lower elevations, rising through yellow and tan, to white at the highest elevations. The shaded image displays topographic slopes in the northwest–southeast direction, with bright northwest slopes and dark southeast slopes. The elevation data were acquired from the Shuttle Radar Topography Mission (SRTM) aboard the *Space Shuttle Endeavor*, launched on 11 February 2000. (Courtesy of NASA/JPL/NIMA.)

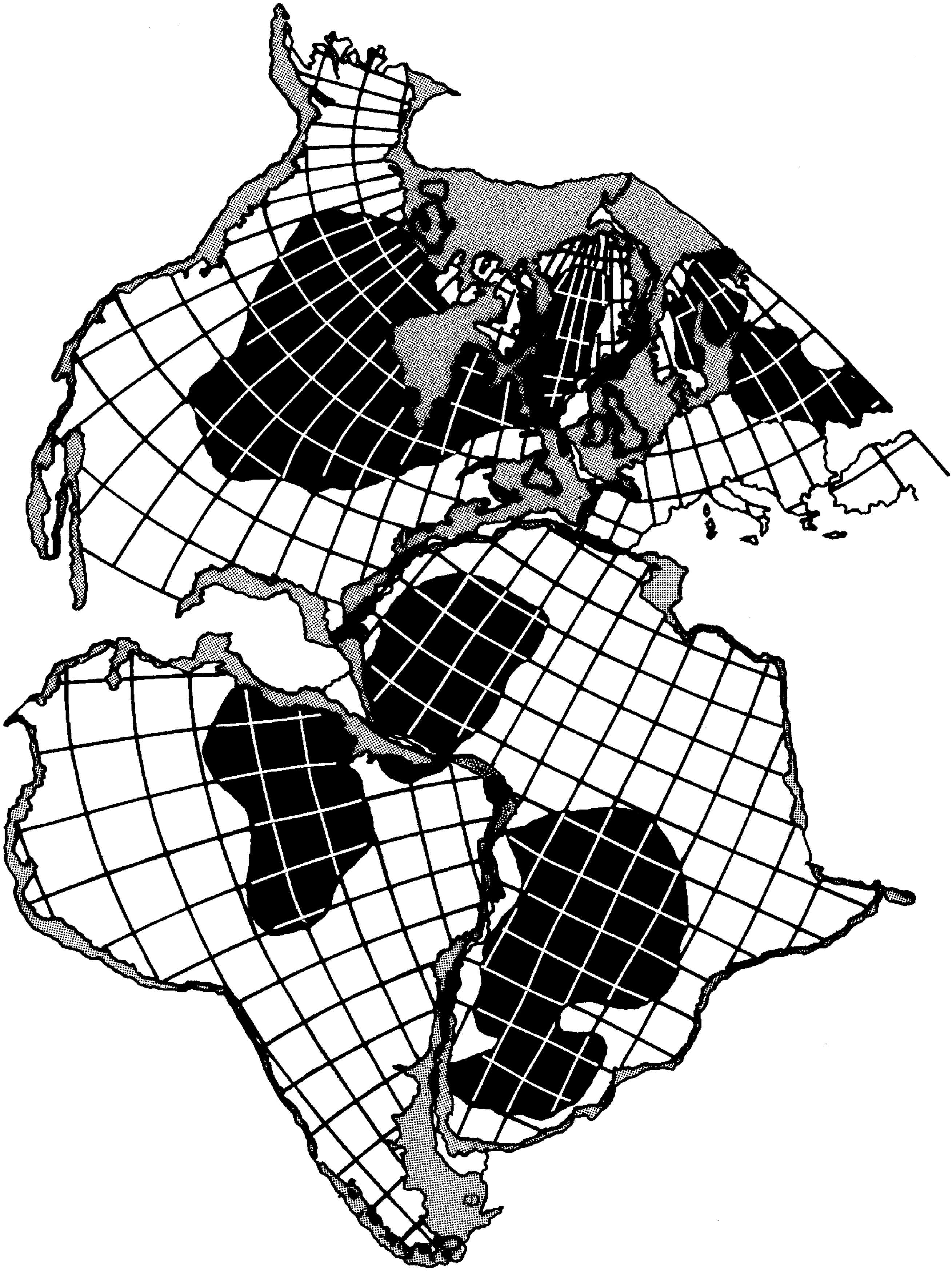


Fig. 4.7 Continental fit The continents fit together like the pieces of a puzzle. Here the fit has been made along the continental slope at the depth of 910 meters, or 500 fathoms (*gray areas*). Within the present continents are ancient terrains between 1.7 and 3.8 billion years old (*black areas*). The close fit of the shorelines of the continents suggests that they once formed a single land mass known as Pangaea shown in Fig. 4.8.

twice as tall as they are today; they have been worn down through tens of millions of years of erosion by wind, rain and ice. If the planet were a perfectly smooth sphere, the oceans would cover the entire globe to a depth of 2.8 kilometers. So we can tell right away that high, dry land must be continuously recreated and pushed up out of the water.

There's water just about everywhere, but most of it is in the salty seas. Less than three percent of the Earth's water is fresh, and most of that is locked up in polar ice caps and glaciers. Lakes, rivers, and other sources of drinkable water make up less than one percent of the planet's total water. As time goes on, the demand for drinking water will increase, and the supply will diminish or even disappear in many dry places within a few decades or less.

Earth's drifting continents

The continents and oceans are not eternal, unchanging aspects of the Earth. Their appearance of permanence is an illusion caused by the brevity of the human lifespan. Just as an entire human lifetime is just a fleeting moment in the history of the Earth, today's map of the world is just a brief snapshot of its evolving, mobile, ever-changing surface. Over hundreds of millions of years, blocks of the Earth move about, producing drifting continents that completely alter our picture of the world.

The idea that continents have not always been fixed in their present positions was suggested more than three centuries ago, in 1596 by the Dutch mapmaker Abraham Ortelius (1527–1598) in his work *Thesaurus Geographicus*. However, the theory of moving continents was not developed into a thorough scientific hypothesis until the early 20th century, by the German meteorologist Alfred Wegener (1880–1930) in his influential and controversial book *Die Entstehung der Kontinente und Ozeane*, or *The Origin of Continents and Oceans*. Wegener noticed that the outlines of the continents exhibit a number of remarkable symmetries. For example, the eastern edge of South America would fit snugly into the western edge of Africa, a fit originally noticed by Ortelius. In fact, large parts of the east and west shores of the Atlantic are as well matched as the shores of a river (Fig. 4.7).

Wegener based his concept of continental drift not only on the similar shapes of the present continental edges, but also on the striking match of certain rocks and geologic formations, fossil creatures, and ancient climates along the borders of continents on opposite sides of the ocean. He concluded that all of the continents were once a part of a single landmass that fragmented and drifted apart (Fig. 4.8). If spacecraft had existed back then, their camera eyes would have seen one large continent and a single ocean surrounding it.

This hypothetical super-continent is called *Pangaea*, a Greek word meaning "all lands" and pronounced *pan-gee-ah*. After all, if today's continents spread apart from their obvious puzzle fit, they had to have been together in the first place. This would also account for the Earth's curiously asymmetric face, in which the ocean waters dominate the southern hemisphere while the continents dominate the northern hemisphere.

Pangaea broke into pieces about 200 million years ago when large amphibians and reptiles ruled the land, leaving many fossils and forming the various smaller continents we see today. As the once-joined continents moved apart, the water rushed in to fill the gap caused by their separation. This led to the various smaller drifting continents that we see today.

Modern geologists have now pieced together the past, reconstructing the pieces of the Earth's moving jigsaw puzzle. They have shown that Boston and southern Florida are both former pieces of Africa, which were left behind when the Atlantic Ocean opened up. China used to be separated from Siberia by at least one ocean. Japan was once attached to Asia, and it may become part of Alaska in 800 million years.

Several super-continents formed and split before Pangaea. There was Rodinia, which covered much of the southern hemisphere 800 million years ago, and Gondwana that was located near the South Pole about 500 million years ago.

Today's continents will continue to move apart, and since the globe is round, all lands will eventually converge again. Thus, in about 250 million years, many of the continents will drift together and reposition themselves into another single, dominant landmass, forming a new Pangaea. And then, inevitably, another break-up will ensue as our restless planet continues to reform and reshape itself. In the process the continents will continue on their endless journey, forever roaming and wandering about the planet with no final destination.

Wegener's theory of continental drift was disparaged, ridiculed and even scorned by most geologists for at least half a century. In retrospect, their objections are hard to understand. The discovery of glacial deposits in Africa and of fossils of tropical plants, in the form of coal deposits, in Antarctica certainly meant that these two continents had once been located at different parts of the globe with climates much different from their present ones.

The main difficulty was understanding how the continents could move across the Earth and plow their way

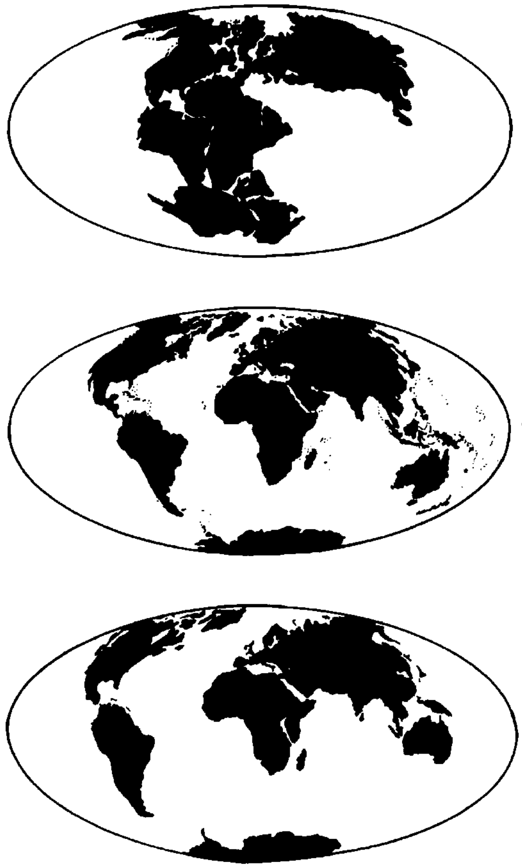


Fig. 4.8 Earth's continental drift Two hundred million years ago all of the continents were grouped into a single super-continent called Pangaea and the world contained only one ocean (*top*). The continents then drifted away from Pangaea, riding on the back of plates to the positions they now occupy (*middle*). The bottom diagram depicts the world geography 50 million years from now.

through solid rock at the bottom of the ocean. A possible mechanism had been proposed by the Scottish geologist Arthur Holmes (1890–1965), who noticed that both the Earth's surface and interior could be in motion. Internal heat could drive churning motions that might propel the continents from below. But these prescient ideas, developed in the 1930s and revitalized in Holmes' 1944 classic *Principles of Physical Geology*, were not widely accepted. Exploration of the ocean floor by sound waves provided the first evidence that Wegener and Holmes were on the right track after all.

Sea-floor spreading

The bottom of the ocean is not flat. It contains underwater mountains and valleys that are as grand as those on any continent. Although we cannot see these features in the inky darkness of the deep sea, we can use sound waves to reach down and touch them. Their distance is determined by recording the time it takes for electrically generated sound signals, called pings, to travel from a ship to the floor and back.

The German Navy used such echo-sounding measurements to reveal the rugged sea-floor soon after World War I (1914–1918). They showed that a chain of submarine mountains runs right across the middle of the floor of the Atlantic Ocean. Known as the Mid-Atlantic Ridge, it extends about 2 kilometers above the adjacent sea-floor, which is at a depth of about 6 kilometers.

Nowadays, the United States Navy detects enemy submarines or ships with sonar, an acronym for sound navigation and ranging, transmitting a continuous train of pulsed sound waves and using the same echo technique to measure distance. Many modern ships, including warships and some commercial fishing boats, are also equipped with sonar to aid in navigation. Navy ships and research vessels can now use sonar to map a two-kilometer swath at the bottom of the ocean in a single ping of the sonar. Gravitational data, obtained from satellites that bounce radio beams off the sea surface (Focus 4.2), complement the sonar data, and they together result in highly detailed maps of the entire ocean floor. They have shown that the Mid-Atlantic Ridge is just a part of a global mid-ocean ridge that snakes its way across the bottom of the world's oceans (Fig. 4.9).

The global mid-ocean ridge is a gigantic network of underwater mountain ranges. The submerged mountains stand higher than the greatest peaks on land, and meander

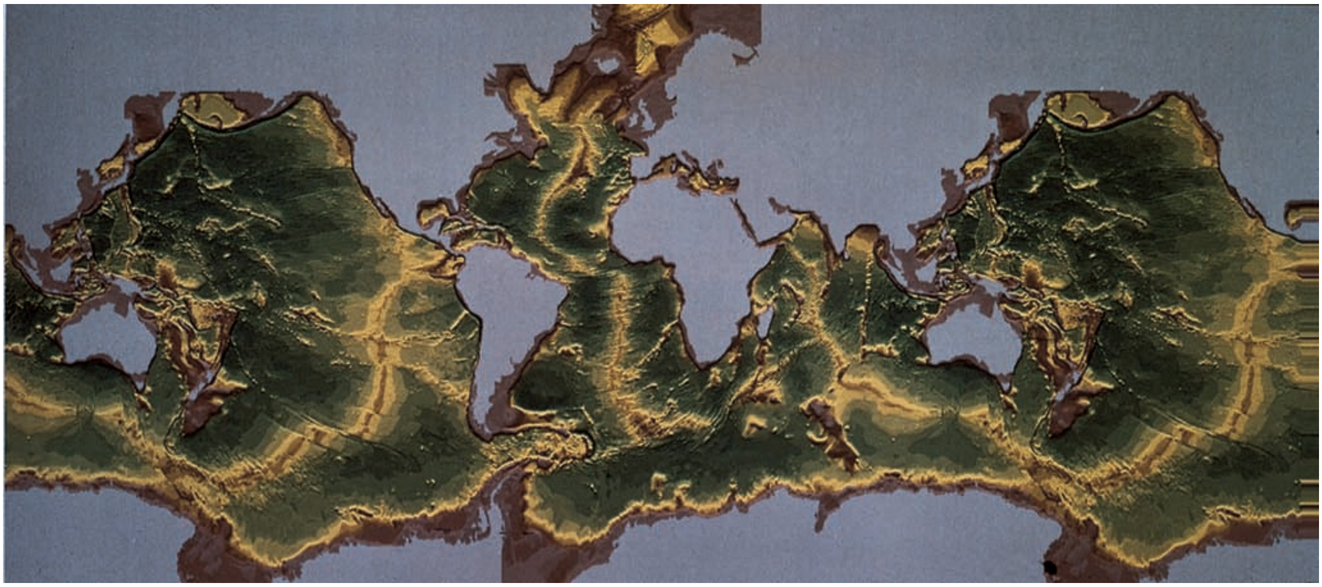


Fig. 4.9 Bottom of the Earth's oceans This map of the world's ocean floors was acquired by the *Seasat* satellite. The Mid-Atlantic Ridge runs down the middle of the ocean floor separating Africa from North and South America. As shown here, a succession of great ridges runs through all of the world's ocean floors, although not always in the middle. (Courtesy of William F. Haxby, Lamont-Doherty Geophysical Observatory, Columbia University.)

Focus 4.2 Mapping the Earth's ocean floor from the top of the sea

The ocean depths have recently been charted with great accuracy by measuring the height of the sea surface. Through the action of gravity on water, seabed mountains produce swells at the surface and canyons or valleys produce dips. So the top of the ocean mimics the topography at its bottom.

A satellite is used to beam pulses of microwaves, or short radio waves, down at the ocean, and to determine the time for the reflected pulse to bounce back. Since the microwave pulse travels at the speed of light, the distance between the satellite and the top of the sea is half the product of that speed and the round-trip travel time. Because the separation between the satellite and the center of the Earth is known from the satellite's orbit, one can establish the distance between the center and the top of the ocean by subtraction.

The topographical contours of the ocean top move up and down with features on the bottom, faithfully tracing out their highs and lows with an extraordinary precision of about one-tenth of a meter. This information helps submariners glide stealthily through the sea. Knowing the precise direction of gravity also improved the accuracy of guided missiles fired from the submarines.

With the cold war ending in the mid-1990s, the United States Navy declassified the data, obtained from the *Geosat* satellite, and it has been used with other satellite altimeter data, of the European Space Agency and NASA, to provide a full, detailed map of the sea-floor.

for more than 75 000 kilometers, creating the longest mountain chain on Earth. It is long enough to accommodate the total length of the Alps, Andes, Himalayas and Rockies. The mid-ocean ridge winds around the Earth, girdling the globe like the stitched seams of a baseball, not in simple lines but in offset segments. When the undersea mountains reach the surface they can form islands, like Iceland and its relatively new neighboring island Surtsey, named after the Icelandic god of fire, Surtur (Fig. 4.10).

Even more remarkable are the deep canyons, collectively known as the Great Global Rift, that run along the mid-ocean ridges, splitting them as though they had been sliced with a giant's knife. Discovered in 1953 by the American scientists Maurice Ewing (1906–1974) and Bruce Heezen (1924–1977), the rift marks a line where

Fig. 4.10 Volcanic islands on Earth Lava erupting from the volcanic island Surtsey on 19 August 1966, almost three years after it rose out of the sea near the coast of Iceland. The volcanic island of Jolnir is in the background. It disappeared back into the sea about one month after this picture was taken, but Surtsey is still visited for research purposes today. All of these volcanic islands, including Iceland, mark points where a mid-ocean ridge has risen out of the ocean. (Courtesy of Hjalmar R. Bardarson, Reykjavik, from his book *Ice and Fire*.)

much of the Earth's internal heat is released. It is filled with hot molten rock, or magma, coming up from inside the planet.

Amazing creatures live down there in the eternal darkness at the bottom of the sea, where life doesn't need the Sun. Giant clams, tubeworms and crabs are warmed and fed by the superheated water. They thrive without light by digesting sulfur minerals emitted from the hot vents, nutrients that other animals would find poisonous. Some of the heat-loving microbes breathe iron, and survive well above the boiling temperatures usually associated with sterilization.

The mid-ocean ridge is more accurately described as a tear in a sheet of paper rather than a cut, for it represents the place from which the ocean floor moves outward on both sides. It is as if the Earth was pulling itself apart and becoming unstitched.

Hot magma emerges from beneath the sea-floor, and oozes into the canyons of the Great Global Rift, filling them with lava. As the lava cools in the ocean water, it expands and pushes the ocean crust away from the ridge. More lava then fills the widening crack, creating new sea-floor that moves laterally away from the ridge on both sides, with bilateral symmetry.

If sea-floor is continuously created in the middle of the ocean, where does it go? If all of the new material kept on piling up, the Earth would grow bigger as time goes on, and that is not observed. The size of the Earth has not changed significantly during the past 600 million years, and very likely not since shortly after its formation 4.6 billion years ago. The Earth's unchanging size implies that the ocean floor must be destroyed at about the same rate as it is being created. The floor disappears back inside the Earth, where it is transformed by the heat and eventually recycles to rise again.

As it migrates away from the hot rift of its beginning, the new ocean floor grows colder and denser, subsiding to greater depths as it ages. After traveling across the Earth, in conveyer-belt fashion for many millions of years, the older, heavier floor bends and descends back into the Earth, often at the edges of continents, creating a deep-ocean trench in the underlying rock. Such trenches are found all around the edges of the Pacific Ocean, and they can sink as far below sea level as the tallest mountains rise above it.

The overall concept is known as sea-floor spreading, an idea introduced by the American geologist Harry H. Hess (1906–1969) in his 1962 paper entitled "History of Ocean

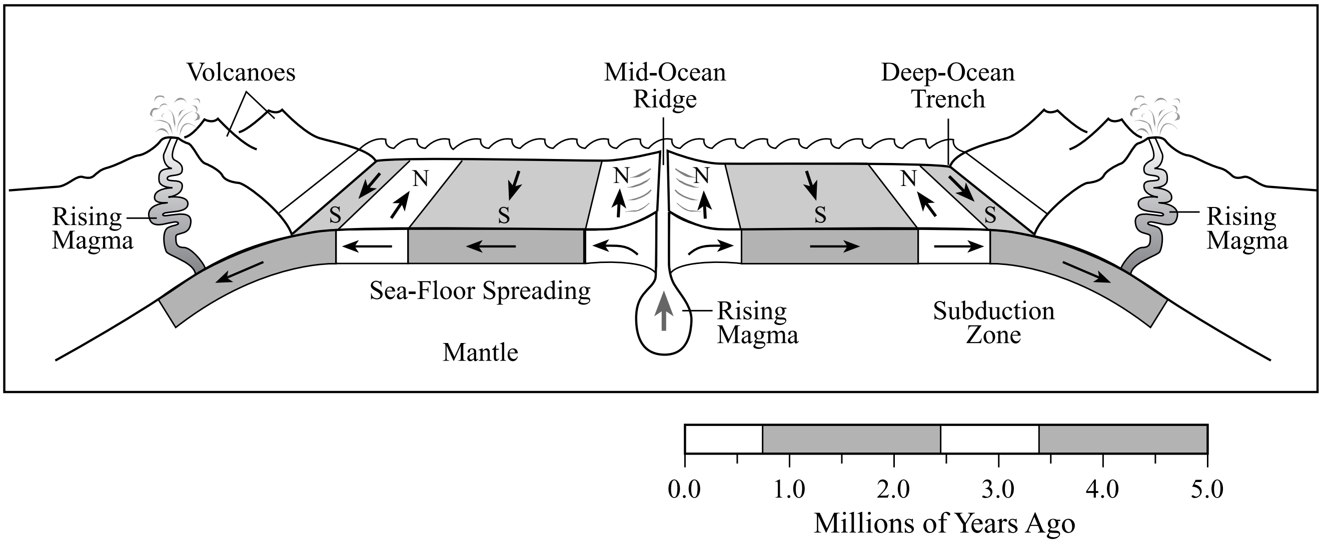


Fig. 4.11 Magnetic reversals and sea-floor spreading Radioactive dating of volcanic rocks on land have been used to determine the timescale of magnetic reversals on the Earth (*bottom*). They indicate that the Earth's magnetic field has flipped, or changed direction, several times during the past 5 million years. The data describe normal epochs (*white*) when compasses would have pointed toward the geographic north, as they do now, and reversed epochs (*gray*) when compasses would have pointed south. The pattern of magnetic reversals on both sides of the volcanic mid-ocean ridge (*top*) is the same, indicating that sea-floor spreading has carried the solidified lava away from the central ridge. The sea-floor is consumed at the other end, when it slides into a deep-ocean trench at a subduction zone.

Basins". In brief, new sea-floor is formed at a rift in the mid-ocean ridge, turning cold and heavy as it spreads away from its source in two directions; the sea-floor eventually sinks and disappears in a deep-ocean trench, where it is consumed. Its material is then recycled and born again as new floor emerges from the central ridge.

Sea-floor spreading accounts for the fact that the ocean floor and the sediments on it are both relatively young. Core samples recovered from the seabed during petroleum exploration show that the floor of the Atlantic Ocean is youngest in the middle, and grows progressively older with increasing distance from the mid-ocean ridge. Both the average age and the thickness of the sediment increase away from the ridge. Moreover, the thickness of the sediments indicates that none of them has been accumulating for more than 200 million years. The oldest fossils found down there are also no more than 200 million years old. From creation to extinction, from rift to trench, the ocean floor completely cleans house, erasing any record of its previous history in less than 200 million years.

In contrast, marine fossils in rock strata on land can be considerably older, including those found near the tops of the highest mountains, and the oldest known continental rocks have ages dating back to about 4 billion years. The sea most probably dates back to the formative stages of the young Earth, more than 4 billion years ago. Thus, young ocean floors have been replacing older ones, while the water above them has remained for billions of years.

Perhaps the most decisive evidence for sea-floor spreading was the discovery of regular magnetic-field patterns in the ocean floor. Magnetic detectors towed behind ships and carried in aircraft could measure very small differences in the Earth's magnetic field from place to place, known as magnetic anomalies. Positive magnetic anomalies are places where the magnetic field is stronger than expected, and negative ones are weaker than anticipated.

One of the motivations of these studies was to detect perturbations in the magnetism caused by enemy submarines, but the results had far greater consequences. The pattern of magnetic anomalies was symmetrically placed, or mirrored, on each side of the mid-ocean ridge (Fig. 4.11). Frederick Vines (1939–1988), then at Princeton University, and Drummond Matthews (1931–1997), working at Cambridge University, compared the magnetic irregularities with the known history of magnetic field reversals found on the flanks of volcanoes on land. These continental lavas show that every so often the Earth's magnetic field has changed its direction, or polarity, and the symmetric magnetic anomalies on the sea-floor exactly match these polarity reversals recorded on land (Fig. 4.11).

Each time the anomaly changes from positive to negative, the Earth's magnetic field turns upside down. Its magnetic poles flip, so the south magnetic pole switches from the north geographic pole to the south geographic pole or vice versa, and the north magnetic pole moves to the opposite geographic pole. Lava emerging at the present

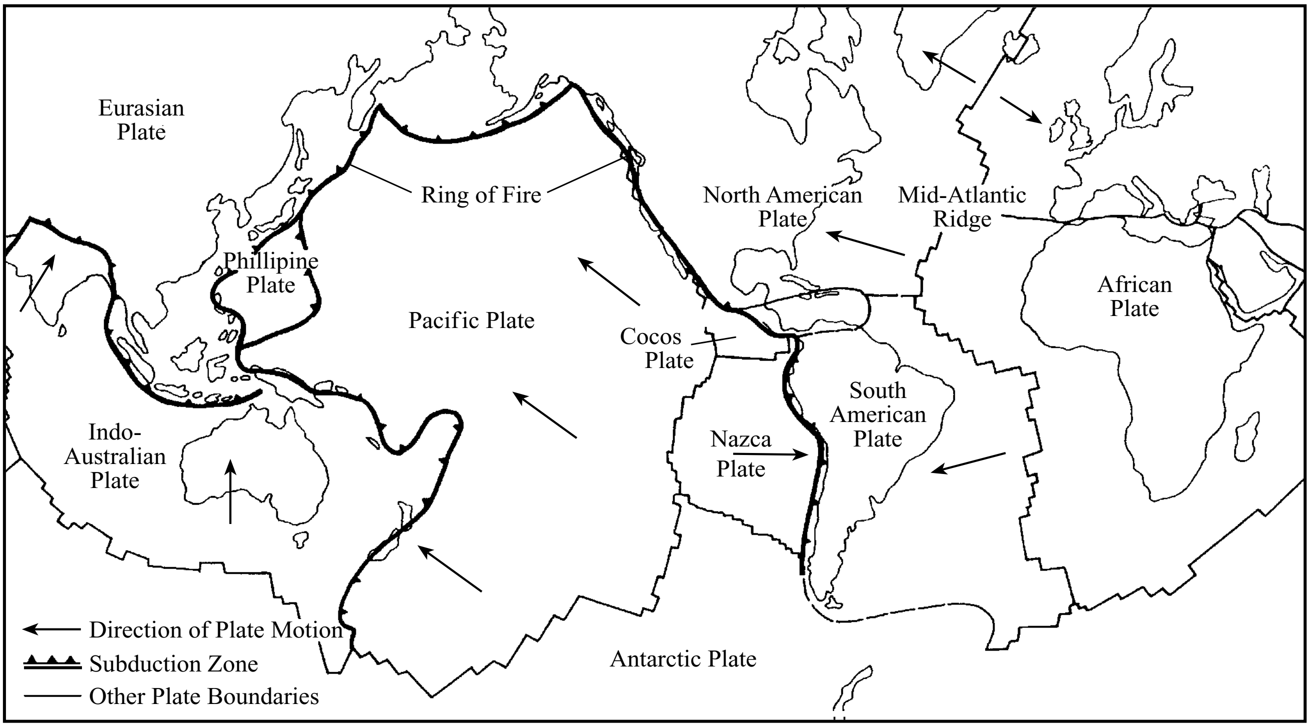


Fig. 4.12 Earth's moving plates The Earth's lithosphere is broken into numerous plates. They move in the directions shown by the arrows at rates of about a tenth of a meter per year. The lithosphere dives into the underlying asthenosphere at zones of subduction. They are denoted by the thick line with triangles, forming the famous Ring of Fire around the edge of the Pacific and Nazca Plates. Most of the Earth's earthquake and continental volcanic activity is concentrated along the subduction zones.

time would have a positive magnetic anomaly, with the Earth's south magnetic pole located at the north geographic pole.

The orientation of the Earth's dipolar magnetic field imprints itself on the volcanic rocks at the time they form, whether on land or under the sea. When fresh molten lava pours out of them, magnetic minerals within the cooling lava become aligned with the Earth's magnetic field, and this orientation or polarity remains as a fossil magnetic record, locked into the rock when it solidifies. Vine and Matthews proposed that the lava on both sides of the mid-ocean rift solidified and moved away, freezing in the magnetic direction at the time, and when the Earth's poles flipped and reversed the lava flows preserved a set of parallel bands with opposite magnetic direction. Thus, the symmetric magnetic-anomaly stripes were recording the Earth's past magnetic field, and providing dramatic support for sea-floor spreading.

By radioactive dating of volcanic rocks on land, it is possible to tell when they solidified and to build up a chronology of the magnetic changes. This chronology can then put dates on the reversals found in the sea-floor, and from the distances traveled it is possible to compute the rate of sea-floor spreading, assuming that the floor has moved at a constant rate. These rates have been independently calibrated astronomically by comparison of the seabed sediments with the orbital parameters that govern climate changes recorded in fossil organisms in the sediments. They indicate that the sea-floor has indeed been spreading for the past 5 million years, moving away from the ridge at rates of 0.02 to 0.20 meters per year depending on the location, or just a little faster than your fingernails grow.

When sustained for 200 million years, the spreading sea-floor can push continents apart by between 400 and 4000 kilometers – entirely adequate to explain the widths of the great oceans. At the measured rate, it took just 150 million years for a slight fracture in an ancient former continent to widen into today's Atlantic Ocean.

Plate tectonics on Earth

The surface of the Earth is continually shaped and molded by plate tectonics, a process that does not occur on any other planet in the solar system. The rind of the Earth, its outer shell known as the lithosphere, is subdivided into a mosaic of large moving plates, each thousands of kilometers across (Fig. 4.12). They vaguely resemble the cracked pieces of an eggshell. The plates move horizontally atop a viscous layer of much hotter, softer, more malleable rock called the asthenosphere. Because of the high temperatures and immense pressures found there, the uppermost part of the asthenosphere flows along at the base of the lithosphere.

Plates are composed of the Earth's crust and the rigid upper mantle just beneath it. Most plates contain both continental and oceanic crust, and they all include oceanic crust. Six of the nine major plates are named for continents embedded in them: the North American, South American, Eurasian, African, Indo-Australian and Antarctic Plates. The other three are almost entirely oceanic: the Pacific, Nazca, and Cocos Plates. Accompanying them is a host of smaller plates.

Driven by heat from below, the plates move with respect to one another, accounting for most of our world's familiar surface features and phenomena, such as mountains, earthquakes and ocean basins. The continents are implanted within the moving plates, and continental drift is a consequence of the motion of plates carried along by the sea-floor spreading.

The rigid plates are in continual, relentless movement, and they deform at their boundaries. Like drops of olive oil gliding across a warm frying pan, the continents sometimes collide and coalesce, sometimes slide and rub against each other, and at other times break up and scatter. The transformations produced by these interactive motions are known as plate tectonics, from the Greek word *tectonic* for "carpenter or building". They are forever reconstructing the face of the Earth.

A mid-ocean ridge is a crack in the sea-floor that is filled in by magma from the underlying mantle as two diverging plates separate. So the ridge marks the boundary between two plates. As the plates move apart in opposite directions, a crack or rift opens up at the crest of the ridge, allowing more molten rock to move up and feed the spreading plates, like blood in an open wound that will not heal.

The spreading ocean floors are eventually pushed back down into the planet's hot interior, at subduction zones where two converging plates meet. When a moving oceanic plate encounters a light continental plate, or a younger, lighter oceanic one, the heavier oceanic plate plunges steeply into the Earth along a subduction zone, like a down-going escalator, producing a deep-ocean trench. Because the continental rock has the lowest density, it remains on the top, while the ocean floor slides underneath. The buried material is consumed within the Earth, only to re-emerge, recycled and transformed at some other location.

Magma is generated at the sinking subduction zones where dense oceanic plates are pushed under lighter continental ones, producing volcanoes that help us locate the edges of the plates. A dramatic example is the circular line of volcanoes that borders the Pacific Ocean. This active belt is known as the Ring of Fire because it is often the site of fiery volcanic eruptions (Fig. 4.12).

These volcanoes recycle atmospheric carbon dioxide, which was dissolved in rainwater and ingested by organisms in the sea. Their shells are deposited on the sea-floor, carried along by the diving oceanic plate, and released in the volcanic eruptions.

Scientists can track the plate motions using Very-Long-Baseline Interferometry (VLBI), and by Satellite Laser Ranging (SLR). In VLBI, radio receivers at widely separated telescopes record the strength and arrival times of cosmic radio signals from quasars, and the interference between the recorded data is used to determine the distance between the telescopes, with an accuracy of less than 0.01 meters. The SLR targets are satellites covered with tiny mirrors called corner cubes. Pulsed laser light, generated at stations on the ground, is bounced off the mirrors and the round-trip time for the light to return is used to establish the distances to the satellites. When combined with precise orbital information for the satellites, these distances can be used to determine the changing separation of the ground stations and the rate of tectonic plate motion.

The VLBI and SLR measurements indicate that the plates move laterally, or horizontally, across the Earth at rates of 0.02 to 0.20 meters per year depending on the plate, which is consistent with the rate of sea-floor spreading. When scientists extrapolate the current plate motions into the past, like running a movie backwards, they find that all of the continents converge, joining together into a single super-continent, Pangaea, about 200 million years ago. As suggested by the Canadian geophysicist J. Tuzo Wilson (1908–1993) in the 1970s, the heat-driven plate motions cause the continents to disperse and then reassemble over and over again, as oceans open and close.

The radio interferometer measurements indicate that the Pacific Plate is carrying Los Angeles northwest with respect to the plate that holds most of North America, at a velocity of 0.048 meters per year, while also producing earthquakes along the edge of the plate (Fig. 4.13). At this rate, Los Angeles will be a suburb of San Francisco in 10 million years. The interferometer observations also indicate that the Atlantic Ocean is widening by 0.017 meters per year, so it was 8.7 meters narrower when Columbus crossed it in 1492.

Earthquakes

An earthquake is a trembling or shaking of the ground caused by a sudden release of energy stored in the rocks below the Earth's surface. The devastating tremors and after-shocks can ravage large sections of the land, flattening

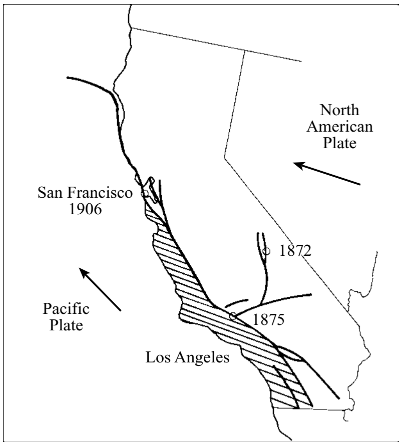


Fig. 4.13 Los Angeles is moving The Pacific Plate is moving northwestward at the rate of 0.048 meters per year, or about 5 meters every century, slowly carrying Los Angeles towards San Francisco. The North American and Pacific Plates strike and rub against each other like immense grindstones, producing earthquakes along their boundary known as the San Andreas Fault. The dated circles denote places where very major earthquakes have occurred with magnitudes of 8 and over on the Richter scale.

entire cities, awakening dormant volcanoes and creating new ones, draining lakes and causing floods, avalanches and fires. Enormous tidal waves, called tsunami, can be generated by the quakes, racing across the ocean at high speed and wiping out everything on the shores they reach. Tens of thousands of people can be killed during a major earthquake, and hundreds of thousands more can be left homeless. Earthquakes are erratic, unpredictable and inevitable, with the power to merely shake you or completely destroy you.

Like volcanoes, the world's earthquakes do not occur just anywhere, but usually along the edges of plates. They occur most often where the ocean floor is being either created or destroyed along the mid-ocean ridges and the deep-ocean trenches. Since an earthquake in the middle of the ocean floor is not likely to disrupt human life, we are naturally most interested in earthquakes that occur near the edges of continents where cities are located.

In addition to diverging plates, which are moving apart at a mid-ocean ridge, and convergent plates, which are heading toward a collision, there is a third type of plate boundary proposed in 1965 by J. Tuzo Wilson. It is known as a transform fault, a place where plates move past one another, neither toward nor away, and this is where earthquakes can occur. When the two plates meet along a transform fault, they "transform" their encounter into a slipping, sliding horizontal motion, and a sudden lurch in this motion can produce an earthquake.

The two plates on each side of a transform fault bump, crush, grind, rub and slide against each other, without creating or destroying crust, like two high-speed cars sideswiping each other, but in slow motion. A famous and visible example is the San Andreas Fault in California that marks the meeting of the Pacific Plate with the North American Plate. In 1906 a great earthquake devastated San Francisco, which is located at the edge of this fault.

The plates on each side of a transform fault build up stress along the line where they meet, and the stress is greatest where they are most tightly locked. As the friction and strain accumulate and rise over the years, a moment comes when the rock can't take it anymore, as when a festering problem of a family can surface into a screaming fight. In effect, the strain surpasses the strength of the rock. The stress is pushed to the limit, the two plates cannot slide further, and the accumulated energy is released as an earthquake. That part of the fault line then lurches back to its original equilibrium position, waiting for the next big one.

Since the time between major earthquakes in a given location can be a little longer than a human lifetime, many imagine that the danger is over, but the ground beneath their feet can remain unstable. Thus people living near the San Andreas Fault should no longer be concerned about whether another earthquake will occur, but when it will happen.

An instrument called a seismograph can measure the relative amount of energy released by an earthquake, its magnitude. The earthquake magnitude is given on a numerical scale, named after the American seismologist Charles F. Richter (1913–1984) who established it. Each increase of one unit on the Richter scale represents a 32-fold increase in the intensity of an earthquake. Major earthquakes usually measure between 6.0 and 9.1 (the highest ever recorded) on the Richter scale. The 1906 San Francisco earthquake would have measured 8.3 on the Richter scale, and the one that occurred there in 1989 measured 6.9.

An even more powerful earthquake of magnitude 9.1 struck off the west coast of northern Sumatra on 26 December 2004, producing devastating tsunamis in the Indian Ocean, sending shock waves that rippled and ricocheted around the globe, and lifted the Earth's surface by a couple of centimeters halfway around the world. A strong earthquake of magnitude 7.0 hit Port-au-Prince, Haiti, on 12 January 2010 with devastating effects, killing more than 200 000 people and significantly damaging or destroying major buildings.

An earthquake of magnitude 8.8 occurred near the shores of Chile on 27 February 2010. Although loss of life and building destruction were not as great as the Haiti earthquake, the one in Chile shook the entire planet, shifting the Earth's axis by about 8 centimeters and shortening the length of an Earth day by about 1.26 microseconds, or 1.26 millionths of a second. The Sumatran earthquake should have shortened the length of day by 6.8 microseconds and shifted the Earth's axis by about 7 centimeters.

An international network established in the late 1990s monitors the globe for clandestine nuclear bomb blasts, and incidentally records the seismic din of earthquakes. Scientists can sometimes distinguish between the two by the form of their shock waves. A nuclear explosion begins with a sharp spike, while an earthquake starts with a gentle shaking that subsequently becomes more violent.

The Earth's internal heat engine

What pushes the tectonic plates across the globe? Like humans, most of the driving forces that transform the Earth's face are hidden below its surface. Or, as the saying goes, it's what inside that counts. Heat, bottled up deep inside the Earth, produces internal currents that move the plates and propel the drifting continents.

The Earth's internal heat is left over from the time of the planet's formation, within the liquid outer core, and augmented by the continued radioactive decay of elements such as uranium and thorium. As the internal heat tries to escape, it maintains a ceaseless, wheeling, churning and roiling motion, called convection, which turns and rolls over very slowly. Convection occurs when molten rock becomes swollen by heat and rises through the cooler overlying material of lower pressure, like the currents in a pot of thick soup or oatmeal about to boil.

The relatively low density of the hottest rock makes the material buoyant, so it ascends slowly; in contrast, the colder, denser rock sinks until heat escaping from the molten core warms it enough to make it rise again. Thus, the hot mantle rock flows in a circular pattern with hot rock rising in some places and cooler rock descending in others. The heated rock moves upward, spreads sideways, dragging portions of the lithosphere with it, and then cools and sinks, to be reheated and pushed upward again. And the crust rides passively atop these giant convection cells, like dirt on a conveyor belt.

Powerful motions deep inside the planet do not only push the plates horizontally; they also produce vertical changes at the surface. After all, the wheeling convection moves up and down, as well as sideways. Huge rising plumes of hot, buoyant, molten rock, originating and channeled at the rugged core–mantle boundary, can expand upward, piercing the mantle to lift and lower entire continents. The rising heat is now pushing South Africa up from below, and has been doing so for the past 100 million years.

The continents are poor conductors of heat, and therefore act like an insulating blanket that tries to block the heat's escape. The pent-up force of the trapped heat can become powerful enough to split a continent apart. A gigantic plume of hot magma may even have played a major role in the break-up of Pangaea.

Thus, the energy that drives the continents, spreads the sea-floor, sets off earthquakes and ignites volcanoes is ultimately derived from the hot interior of the Earth. From the inner core all the way to the surface of the Earth, the dynamical activity of the Earth is driven by heat, and like any heat engine the Earth must be gradually running down. When the Earth's internal heat becomes totally depleted, it will become a geologically dead planet, and erosion will gradually flatten the mountains. In the meantime, continents continue to be renewed as the result of the Earth's internal heat engine.

The Earth's impermanent face

Moving plates provide the tools for sculpting the Earth's surface and altering its landscape. They have profoundly changed the way we view the world. Its entire surface is continuously shifting about and changing in shape and form. It is something like a rich old lady who keeps going in for face-lifts, in a futile attempt to resurrect her youth. Indeed, nothing on the Earth's face resembles itself as it was even several millions of years ago.

As an ocean plate disappears into the trenches, great chains of volcanoes are created along the margins of continents. The descending slab of lithosphere causes underground rock to melt, and the magma generated rises buoyantly to widen the continents at their edges. The Andes are still growing higher in this way, as the floor of the Pacific Ocean plunges beneath the west coast of South America.

The Pacific Ocean once reached to Colorado, and the western United States has been grafted onto the continent. For most of the world's history, the land we call California did not exist. Where California has come to be there was only the deep blue sea reaching down to the spreading ocean floor. But the floor was moving into the Earth and being consumed under the ancient shoreline, creating volcanoes that rose to create new land.

On average, about 2 billion cubic meters of lava and ash are now being added to the continents by volcanoes each year. The rising material also brings valuable metals and precious stones to the surface. All the gold in California originated in this way, as did the famous copper deposits in *Cyprus*, the Greek word for "copper", rising with volcanic magma and spewing out on the surface with the lava flows.

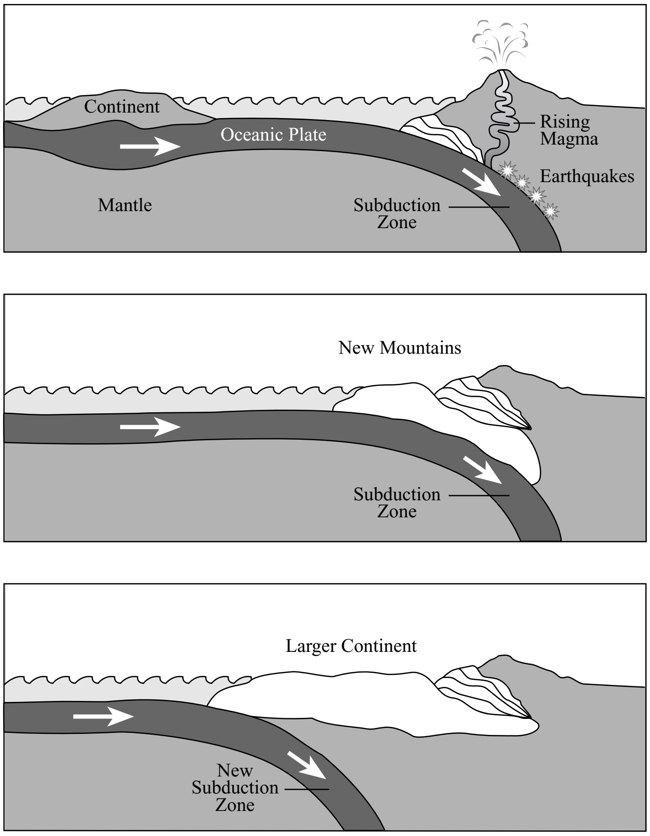


Fig. 4.14 Converging plates on Earth Magma, volcanoes and earthquakes are generated at a subduction zone (*top*) where a dense oceanic plate is pushed under a lighter continental plate. When continents on two moving plates meet head on, new mountains are generated (*center*). In some situations, the advancing plate may become disrupted and the plate motion may stop. The two continents then become welded together forming a larger one, and a new subduction zone can be formed elsewhere (*bottom*).

Diamonds were also forged in the crucible of the Earth's hot interior. The crushing pressures and blistering heat far within the mantle worked in unison to squeeze carbon into diamonds, and some of them have risen from the deep, entrained in explosive volcanic eruptions, even in the middle of continents.

Eventually, a moving continent reaches an open trench and jams it shut (Fig. 4.14), like trying to shove an eggplant down a garbage disposal. Continents are too light and thick to be subducted, and when they arrive at a trench the suture is closed up.

When two continents meet, they buckle upward to form a range of mountains and help to hold the land above the sea. Both land and oceanic sediment, built up over many millions of years, are tossed into the sky. The magnificent Himalayan range was formed this way (Fig. 4.15), when the Indo-Australian plate, with India firmly embedded, ran into the Eurasian plate, like a head-on collision of two cars. Slowly, the Himalayas shot up as India rammed into Asia, carrying the fossilized remains of ancient sea-creatures with them. Today the plate that carries India continues to slide beneath the Eurasian plate, widening the Indian Ocean and pushing the mighty Himalayas upward.

The European Alps have been fashioned in a similar way to the Himalayas, when the African Plate moved Italy up from Africa and collided with the Eurasian plate along Switzerland's former ocean shore. Today the African Plate continues pressing Italy northward and raising the Alps.

Like adolescents, the relatively young mountains in the Alps and Himalayas just can't stop growing, but there is a limit to how high they can stand. Gravity opposes the upward forces, and erosion wears away mountain summits as they are being pushed up from below.

Erosion provides the second major force, in addition to plate tectonics, for sculpting the Earth's face. As massive as it is, a range of mountains cannot resist eventual destruction by the erosion of wind, rain, flowing water, and ice. Old mountain ranges, such as the Appalachians in the United States, once stood as high as today's Himalayas, but they have eroded into gentle undulations and rounded knobs.

This erosion acting by itself would have worn away the continental mountains in a few hundred million years, and oceans would therefore have long ago covered the entire globe. But the mountains are constantly being rebuilt while enlarging the continents. Many of the world's present-day continents have indeed been assembled as former continents welded together.

The edges of plates are not the only place that land rises above the sea. Some oceanic islands are located thousands of kilometers from the nearest plate boundaries. Chains or strings of such isolated islands are attributed to hot spots, an idea introduced by the prolific J. Tuzo Wilson back in 1963. The hot spots are rising plumes of magma, or molten rock, anchored far beneath the ocean and deep within the mantle, even as far down as the core–mantle boundary. The relatively small, long-lasting and exceptionally hot regions provide a persistent source of magma capable of penetrating the mantle and piercing an overriding lithosphere plate, like a fixed blowtorch might melt holes in a steel plate moving by. As the plate glides slowly overhead at the rate of a few meters every century, it can leave a trail of islands that have risen out of the sea.

The Hawaiian Islands were formed in such a way, as the Pacific Plate moved over a deep, stationary hot spot at the slow rate of 0.13 meters per year (Fig. 4.16). Stretching to the north and west of the big island of Hawaii, they form a string of smaller islands, including Oahu and Midway, and submerged volcanoes, or seamounts, altogether extending about 6000 kilometers in length. Every

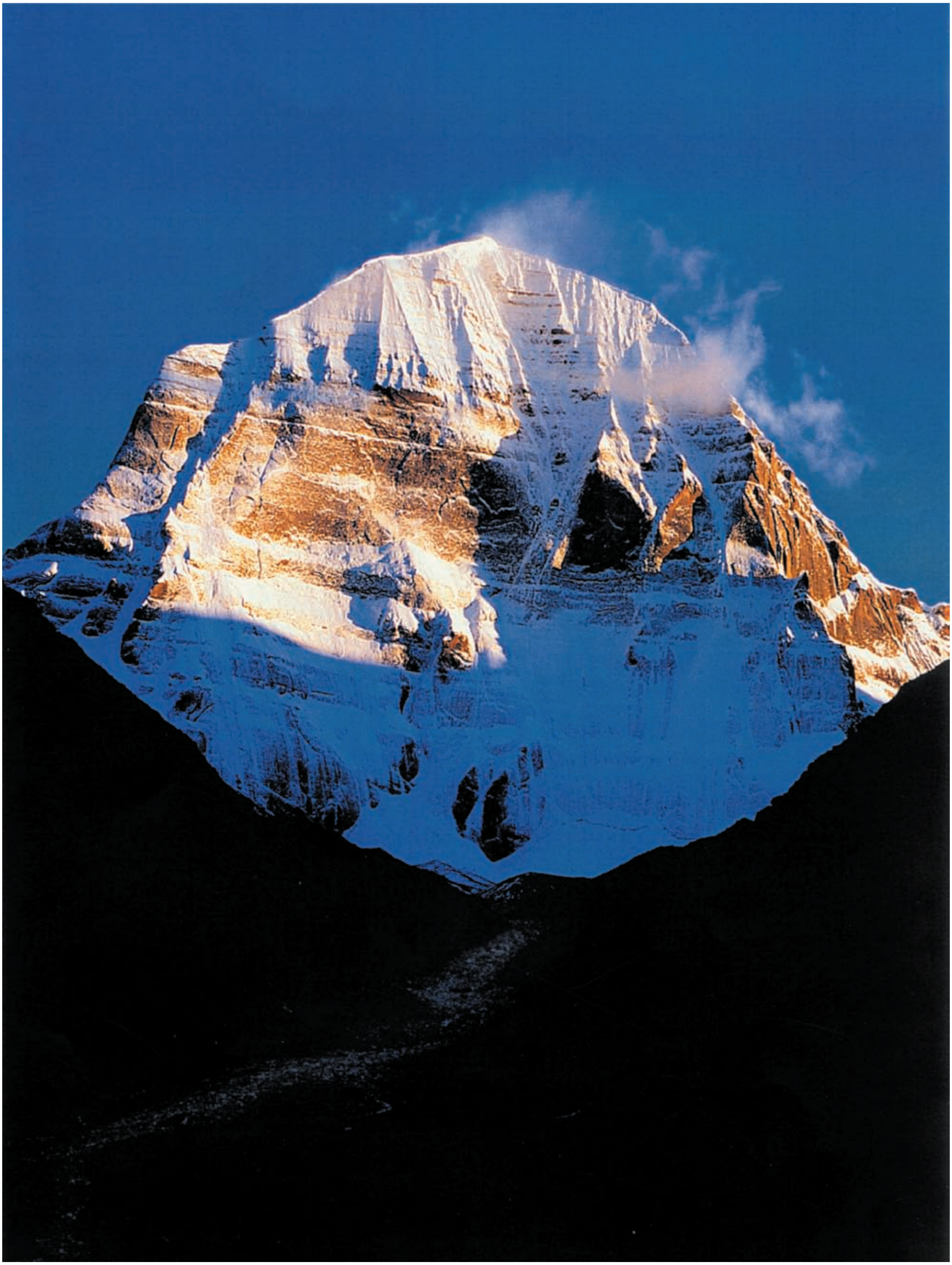


Fig. 4.15 Mount Kailash This mountain is sacred to both Hindus and Buddhists. Tibetans know it as Kang Rinpoche (The Precious Snow Mountain). It was formed by the collision of two former continents, now welded together in a seam known as the Himalayan mountain range. Buddhists consider it to be the palace Chakasamvara (Wheel of Supreme Bliss) and Hindus consider it to be the dwelling place of Shiva. There are also many nearby caves where famous hermits like Milarepa meditated for years. Pilgrims have been visiting and circumambulating Mount Kailash, in western Tibet, for thousands of years. (Courtesy of Matthieu Ricard, Shechen Monastery, Katmandu, Nepal.)

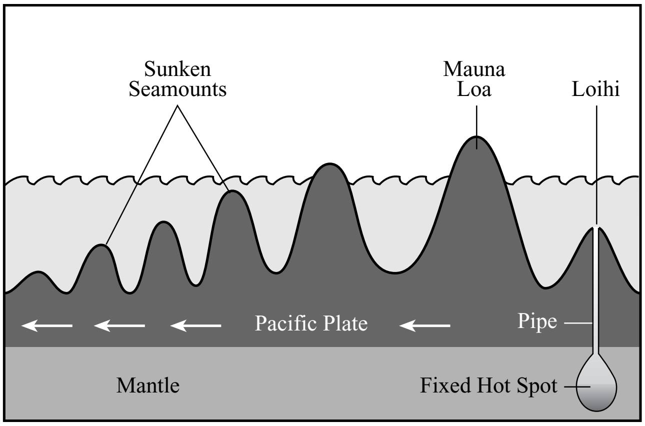


Fig. 4.16 Hot spot forms the Hawaiian Islands A hot spot that is anchored deep within the Earth has recently fed molten lava through a long pipe to Mauna Loa on the big island of Hawaii. The moving Pacific Plate has carried three other volcanic islands away from the hot spot. As the plate moves on, wind and water erode the peaks, reducing these now-extinct volcanoes to sunken islands known as seamounts. An underwater volcano, named Loihi, is now forming over the hot spot; it should rise above the ocean to become another Hawaiian island in about 50 000 years.

one of these islands and seamounts was formed in the exact place where Hawaii now stands. The plume pushed the first Hawaiian island up above the ocean surface in this location about 70 million years ago.

Kiluea, the world's largest active volcano, is still rumbling because Hawaii has yet to move completely off the hot spot. At the same time, the underwater volcano, Loihi, is being formed as the Pacific Plate moves steadily on, continuing its relentless journey over the hot spot (Fig. 4.16). In about 50 000 years Loihi should grow high enough to form the next Hawaiian island.

But why aren't the Hawaiian Islands just one long extended island, eroded away on the oldest end and standing tallest at the youngest end? Although the lower parts of the hot-spot plumes are shaped like the thin stem of a wine glass, their tops flare out into mushroom-shaped reservoirs of molten rock that pools under the lithosphere and overriding crust, slowly melting them and gathering enough strength for penetration. It takes thousands of years to break on through to the other side, just as a welder's torch takes a while to burst through a steel plate. The hot rock breaks though the lithosphere and overlying ocean floor sporadically rather than in a continuous stream, forming a succession of oceanic islands.

Most of the hot spots lie under oceans and give birth to island chains, but some of them penetrate the mantle under the continents. Such a hot spot has created the hot springs, boiling mud and geysers of the Yellowstone National Park in Wyoming. As the North American Plate moved above this hot spot, it created a long line of volcanoes that are now extinct, with ages ranging from 0.6 to 6 million years old. A semi-dormant volcano now rests under Yellowstone.

If a plate carrying a continent comes to rest over a hot spot, the heat and pressure from the upwelling magma will weaken and stretch the overlying material. And when the continental crust is stretched beyond its limits, cracks or rifts will form in it. The magma rises and squeezes through the widening cracks, forming volcanoes. If the upwelling is short-lived, the result is merely a rift scar, such as the Rhine Valley. If it persists, the rift widens and a continent can literally be split in two. In time, the gap reaches a coastline, permitting seawater to flow in and a new ocean is created (Fig. 4.17).

Hot spots are now tearing Africa apart at its seams. A Great Rift Valley stretches from Ethiopia to Tanzania; as it widens the continent will break apart and the sea will eventually enter. The African and Arabian Plates have already pulled apart in another location, forming the Red Sea and the Gulf of Aden. They are developing into an ocean that may eventually rival the Atlantic Ocean in size (Fig. 4.18). At the same time, the Mediterranean Sea is narrowing as Africa moves toward Europe.

Thus a new dynamic picture of the Earth has emerged. Continents are growing in size by accumulating volcanic material at their edges, when oceanic plates plunge under them, or by colliding continental plates that can weld together and create some of the world's largest mountains, but wind, rain, flowing water and ice inevitably cut them down to size. Earthquakes are bringing vast destruction as two plates grind together; strings of volcanic islands are rising from the ocean's depths; converging continents are squeezing oceans out of existence while new ones open up where continents are splitting apart; and the ocean floor remains in eternal youth as new floor spills out of the mid-ocean ridges and old floor is pushed back into the Earth. As we shall next see, our atmosphere is also in a perpetual state of flux, forever changing in sometimes dangerous ways.

4.4 The Earth's changing atmosphere

Our Sun-layered atmosphere

Our thin atmosphere is pulled close to the Earth by its gravity, and suspended above the ground by molecular motion. Its height is about 1 percent of the planet's diameter. Compared to the size of the Earth, the thickness of the air is something like the width of a window on a big building.

Because air molecules are mainly far apart, our atmosphere is mostly empty space, and it can always be

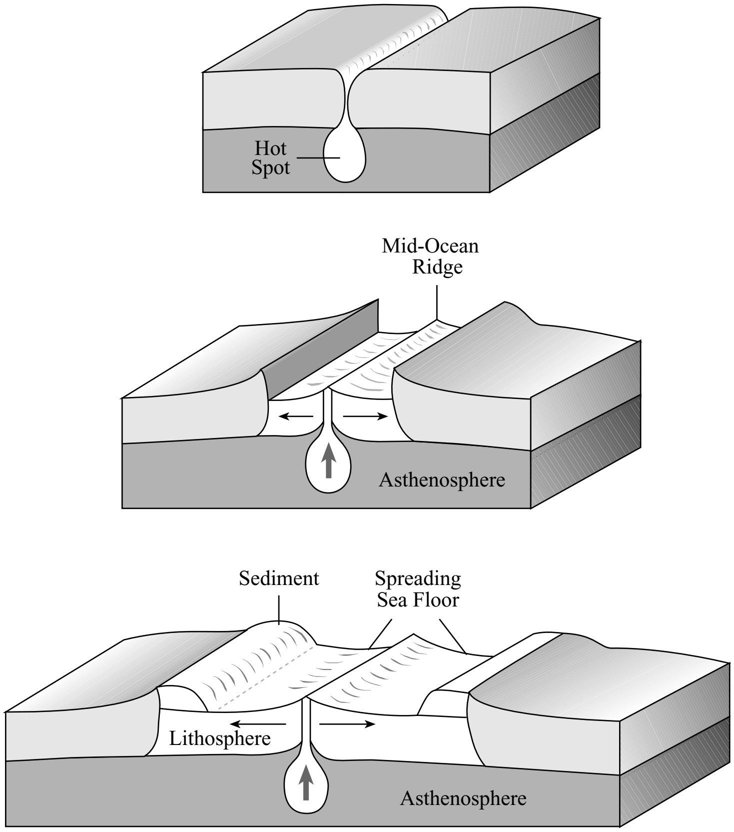


Fig. 4.17 The rifting of a continent A continental rift begins when molten lava rises up from deep in the Earth's interior and splits a continent open. As the fissure grows and widens, a future ocean floor spreads away from the ridge. Water should eventually flow into the cavity, making a new ocean.

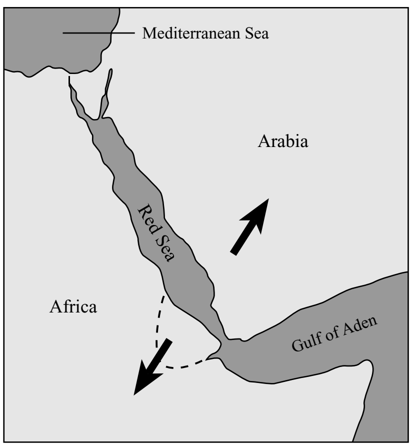


Fig. 4.18 An infant ocean An ocean is being born where the Arabian peninsula and the African continent are moving apart, a process that began about 20 million years ago. In a few hundred million years, the Red Sea could be as wide as the Atlantic Ocean is now.

squeezed into a smaller volume. The atmosphere near the ground is compacted to its greatest density and pressure by the weight of the overlying air. At greater heights there is less air pushing down from above, so the compression is less and the density and pressure of the air falls off into the near vacuum of space.

Not only does the atmospheric pressure decrease as we go upward; the temperature of the air also changes. It decreases steadily with increasing height in the lowest region of our atmosphere, called the troposphere from the Greek *tropo* for "turning". The troposphere extends from the Earth's surface to about 12 kilometers above sea level.

All of the Earth's weather takes place in the troposphere, a thin planetary skin just one-thousandth the diameter of the Earth. Yet about 80 percent of the atmosphere's total mass and almost all its water are concentrated in the troposphere.

The temperature falls at increasing heights in the troposphere because this layer of our atmosphere is heated from the warm ground below, and because the air expands in the lower pressure at higher altitudes and becomes cooler. The average air temperature drops below the freezing point of water, at 273 kelvin, about 1 kilometer above the Earth's surface, and bottoms out at roughly 10 times this height.

But the temperature is not a simple fall-off with height. It falls and rises in two full cycles as we move off into space (Fig. 4.19). The temperature increases are produced by the Sun's invisible radiation.

Different types of radiation differ in their wavelength, though they propagate at the same constant speed – the velocity of light. They range from long radio waves, about

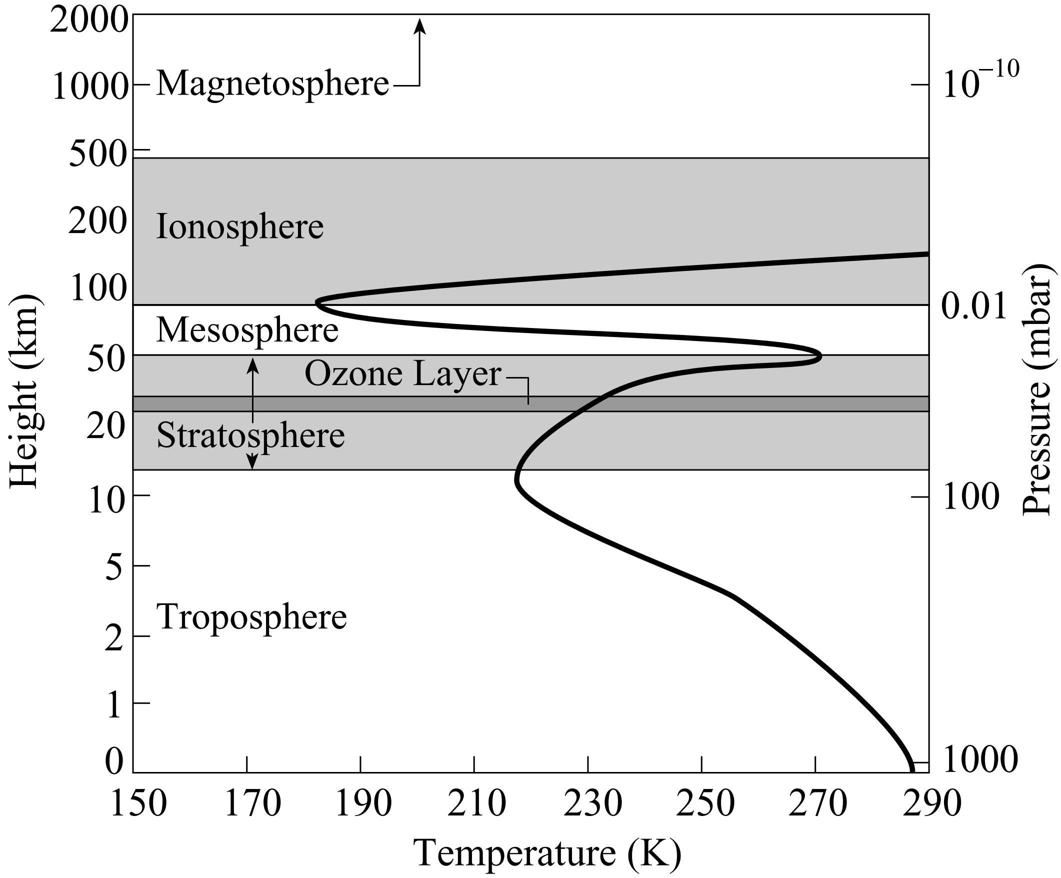


Fig. 4.19 Earth's layered atmosphere The pressure of our atmosphere (*right scale*) decreases with altitude (*left scale*; note that both the pressure and altitude scales are logarithmic). This is because fewer particles are able to overcome the Earth's gravitational pull and reach higher altitudes. The temperature (*bottom scale*) also decreases steadily with height in the ground-hugging troposphere, but the temperature increases in two higher regions that are heated by the Sun. They are the stratosphere, with its critical ozone layer, and the ionosphere. The stratosphere is mainly heated by ultraviolet radiation from the Sun, and the ionosphere is created and modulated by the Sun's X-ray and extreme ultraviolet radiation.

a meter in length, to short X-rays whose wavelengths are roughly a billionth, or 10^{-9}, of a meter. The radiation at most of these wavelengths goes unseen by humans. Our eyes are only sensitive to a narrow range of visible colors, with wavelengths between 4 ten-millionths and 7 ten-millionths, or 4×10^{-7} and 7×10^{-7}, meters. Ultraviolet radiation is on the short-wavelength side of blue light, with wavelengths of about 2 ten-millionths, or 2×10^{-7}, meters. The wavelengths of X-rays are about a hundred times shorter than the ultraviolet rays.

The most intense radiation from the Sun is emitted at visible wavelengths, and our atmosphere permits it to reach the ground. That is the colored sunlight that our eyes respond to. The Sun emits lesser amounts of invisible, short-wavelength radiation, which is partially or totally absorbed in the atmosphere.

Even though the total amount of invisible solar radiation is substantially less than the visible emission, the individual short-wavelength rays are more energetic. That is why we get sunburns from the ultraviolet radiation that manages to get through the atmosphere, and why we need to be protected from the Sun when climbing at high altitudes where the air is thinner and more ultraviolet penetrates the atmosphere. The greater energy of radiation at shorter wavelengths also explains why X-rays, generated by machines, can see through your skin and muscles to detect your bones.

When absorbed in our air, the invisible short-wavelength radiation from the Sun transfers its energy to the atoms and molecules there, causing the temperature to rise. There is, for example, a gradual increase in temperature just above the troposphere, within the next atmospheric layer named the stratosphere (Fig. 4.19). This layer is located between 10 and 50 kilometers above the Earth's surface. Its name is coined from the words "stratum" and "sphere".

The mesosphere, from the Greek *meso* for "intermediate", lies just above the stratosphere. The temperature declines rapidly with increasing height in the mesosphere, reaching the lowest levels in the entire atmosphere.

The temperature then begins to rise again with altitude in the ionosphere, a permanent spherical shell of electrons and ions, reaching temperatures that are hotter than the ground. The ionosphere is created and heated by absorbing the extreme ultraviolet and X-ray portions of the Sun's energy. This radiation tears electrons off the atoms and molecules in the upper atmosphere, thereby creating ions and free electrons that are not attached to atoms.

The ionosphere was postulated in 1902 to explain Guglielmo Marconi's (1874–1937) transatlantic radio communications. Since radio waves travel in straight lines and cannot pass through the solid Earth, they get around the planet's curvature by reflection from electrons in the ionosphere.

The Sun's invisible rays make the Earth's ozone layer

Solar ultraviolet radiation is largely absorbed in the cold and barren stratosphere, where it helps make ozone. When ultraviolet rays strike a molecule of the ordinary diatomic oxygen that we breathe, denoted by O_2, they split it into its two component oxygen atoms, or two O. Some of the freed oxygen atoms then bump into, and become attached with, an oxygen molecule, creating an ozone molecule, abbreviated O_3, that has three oxygen atoms instead of two. The Sun's ultraviolet rays thereby produce a globe-circling layer of ozone in the stratosphere.

Solar ultraviolet radiation also heats the statosphere, making the molecules of ozone move faster. The main reason for the decreasing temperatures in the overlying mesosphere is the falling ozone concentration and decreased absorption of solar ultraviolet.

Although the ozone is present to the extent of only about 10 parts per million, the ozone layer is critical to life below. It protects us by absorbing most of the Sun's ultraviolet emission and keeping its destructive rays from reaching the ground. If there were no ozone shield, plants, animals and humans could not even exist on land.

The amount of ozone in the stratosphere resembles the level of water in a leaky bucket. When water is poured into the bucket, it rises until the amount of water poured in each minute equals the amount leaking out. A steady state has then been reached, and the amount of water in the bucket stops rising. It will stay at the same level as long as you keep pouring water in at the same rate. However, if you pour the water in at a different rate, or punch a few more holes in the bucket, the steady-state level of water in the bucket changes.

Solar ultraviolet radiation supplies ozone to the stratosphere from above, like pouring water into a bucket, at a rate that varies slightly with the changing ultraviolet output of the Sun. We have recently been punching holes in the ozone layer from below, with chemicals used in everyday lives.

Synthetic chemicals are destroying the Earth's ozone layer

Man-made chemicals, called chloroflurocarbons, are consuming the protective ozone layer, eating holes in it and making it thinner. They are synthetic chemicals, entirely of human origin with no counterparts in nature. The name of the chemicals is a giveaway to their composition. Each molecule has been constructed in company laboratories by linking atoms of chlorine, fluorine and carbon.

The shorthand CFC notation abbreviates some of them. A number sometimes follows, providing a complex description of the number of atoms in each molecule, the most widely used being CFC-11 and CFC-12.

Beginning in 1930, the biggest producer of CFCs, the Du Pont Company, manufactured and marketed them under the name Freons. They have been widely used in refrigerators, plastic foams, spray-can propellants, automobile air-conditioning systems, and the cleaning of circuit boards used in televisions and computers.

The hardy substances don't interact chemically to form other ones. They are so inert and stable that once entering the atmosphere the CFC molecules can survive for more than a century, permitting them to drift and waft up into the ozone layer in the stratosphere. Although more than 20 million tons of CFCs have been released into the air, their combined concentration isn't very significant, only about one CFC molecule for every two billion molecules in the air. Yet even these seemingly insignificant amounts can have enormous impact.

In 1974, Mario J. Molina (1943–) and F. Sherwood Rowland (1927–), then at the University of California at Irvine, showed that the chlorine in the CFCs could destroy enormous amounts of ozone. Once arriving in the stratosphere, the Sun's ultraviolet rays will split chlorine atoms out of the CFCs, and the liberated chlorine sets off a self-sustaining chain reaction that destroys the ozone. A single chlorine atom will react with an ozone molecule, taking one oxygen atom to form chlorine monoxide; the ozone is thereby returned to a normal oxygen molecule. Moreover, when the chlorine monoxide encounters a free oxygen atom, the chlorine is set free to strike again. Each

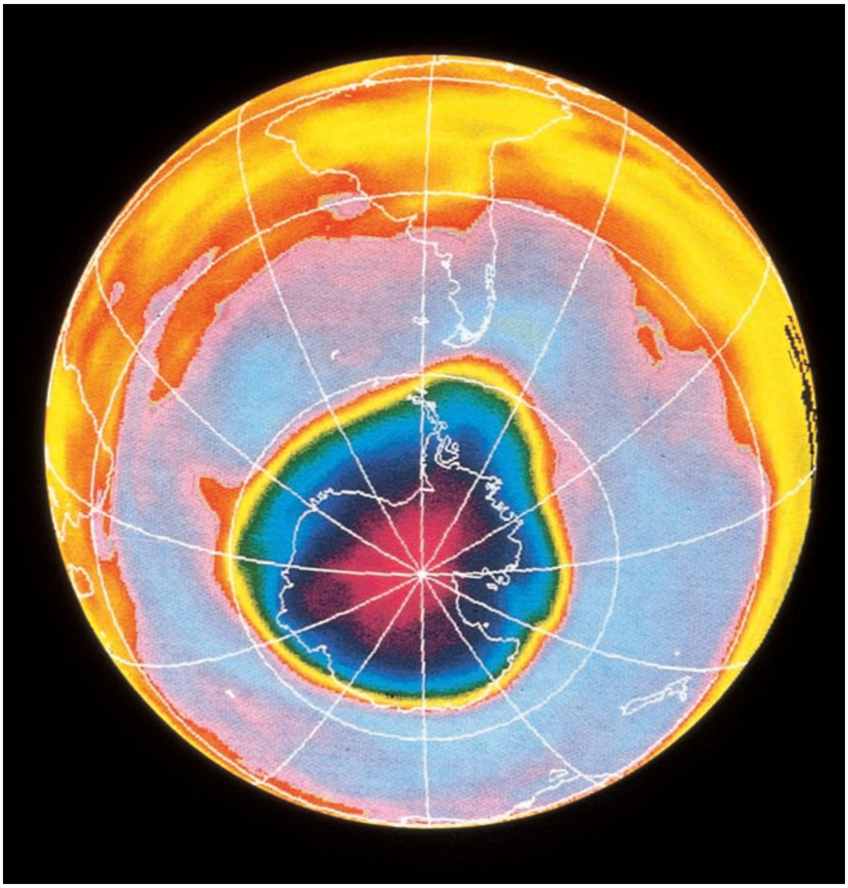


Fig. 4.20 Hole in the sky A satellite map showing an exceptionally low concentration of ozone, called the ozone hole, which forms above the South Pole in the polar spring (September–November). In October 1990 it had an area larger than the Antarctic continent, shown in outline below the hole. Eventually spring warming breaks up the polar vortex and disperses the ozone-poor air over the rest of the planet. (Courtesy of NASA.)

chlorine atom thus acts as a catalyst, destroying about 10 000 ozone molecules before it finally combines permanently with hydrogen in the air.

Molina and Rowland were awarded the Nobel Prize in Chemistry in 1995 for their "contribution to our salvation from a global environmental problem that could have catastrophic consequences". They shared the prize with the German chemist Paul Crutzen (1933–), who showed how the rate of ozone depletion could be accelerated by other chemical reactions in the atmosphere.

The ozone layer is itself invisible. But you can determine its ozone content by measuring the amount of solar ultraviolet radiation getting through the layer and reaching the ground. When there is more ozone, greater amounts of ultraviolet are absorbed in the stratosphere and less reaches the ground, and when the ozone layer is depleted, more of the Sun's ultraviolet rays strike the Earth's surface.

The British scientist G. M. B. (Gordon Miller Bourne) Dobson (1889–1976) pioneered measurements of the air's ozone content more than half a century ago. When his instrument was installed at Halley Bay, Antarctica, in 1957–58, Dobson found that the ozone abundance in polar spring (September–November) was noticeably less than that above other parts of the world.

Other British scientists continuously monitored the southern polar skies for 27 years; always detecting a springtime loss that became steadily larger as the years went on. By 1985 the ozone loss above Antarctica had nearly doubled when compared to the earlier measurements in the 1960s, and it extended all the way to the tip of South America, where another British monitoring station detected the ozone depletion. A continent-sized hole had opened up in the sky – the ozone hole (Fig. 4.20).

This unexpected discovery astounded space-age scientists who had not detected any ozone hole using satellites that had been monitoring the ozone layer from above. Their computers had been programmed to automatically reject large ozone depletions, apparently because their models did not predict such huge losses. So the now-famous ozone hole had been discarded as an anomaly, perhaps caused by an instrumental error. After reanalyzing the satellite data, the scientists confirmed the existence of an ozone hole in the local springtime above the South Pole.

We now know that strong winds concentrate ozone-destroying chemicals, the CFCs, within a vast towering vortex above Antarctica, resembling the eye of an immense hurricane. Each year the gaping hole opens up during Antarctic spring when the sunlight triggers ozone-destroying chemical reactions; the hole starts to close up in the early polar fall when the long sunless winter begins. Ozone-depleted air is dispersed globally, and the ozone is slowly restored, filling the hole until the cycle repeats in the following year.

Doing something about ozone depletion

The sudden and frightening discovery of an enormous ozone hole in 1985 sparked public awareness of the fragile ozone layer. In the meantime, the scientific community had been actively investigating Molina and Rowland's theory that the CFCs could be destroying the ozone layer. Although global models of the expected ozone depletion initially led to widely varying estimates of the potential threat, affecting the scientists' credibility and dampening public concern, a coordinated international investigation eventually led to a unified assessment of the problem.

A group of approximately 150 scientific experts reported in 1986 that atmospheric accumulations of CFC-11 and CFC-12 had nearly doubled from 1975 to 1985. The continued release of the synthetic chemicals at the 1980 rate could, they said, deplete the ozone layer by about 9 percent on a global average by the last half of the 21st century, with even greater seasonal and latitudinal differences. As a result, higher levels of dangerous ultraviolet radiation could reach heavily populated regions of the northern hemisphere.

Who cares if chemicals are punching a few holes in the sky and letting a little more sunlight reach the ground? The United States Environmental Protection Agency (EPA) cared. In 1986 it published a report of the many serious consequences of ozone depletion. A thinner ozone layer lets more solar ultraviolet radiation through to the ground, where it can produce severe biological harm. The most energetic ultraviolet rays will reduce the effectiveness of the human immune system, increasing human vulnerability to infections and cancer.

The EPA estimated that there could be over 150 million new cases of skin cancer in the United States alone among people currently alive or born by the year 2075, resulting in over 3 million deaths. The dangerous ultraviolet would also produce eye cataracts, distorting the vision of about 18 million people in the same population and blinding many of them. Added to this was the potential of widespread genetic damage to crops and forests, if nothing was done to stem the production of ozone-destroying chemicals.

Faced with the evidence of vanishing ozone, the global increase of atmospheric CFCs, and the prospect of widespread skin cancer and eye cataracts, international diplomats forged an accord in 1987 to limit and eventually ban the production of the substances that deplete the ozone layer. The treaty, known as the *Montreal Protocol*, has led to substantial reductions in ozone destroyers.

The *Montreal Protocol* was the first international agreement to protect the global environment. The treaty also marked the first time that the governments of the industrial nations agreed to help developing countries with environmentally safe substances and technology. It was further hoped that the precedent would pave the way for international agreement on global warming.

The *Montreal Protocal* was gradually strengthened over the years. A variety of ozone-depleting substances were added to those already banned, as amendments to the initial agreement made at meetings held in London, Copenhagen, Montreal and Beijing in 1992, 1994, 1999 and 2002, respectively, and the number of participating countries grew to nearly 200, including the vast majority of both the producers and consumers of the dangerous substances. This rapid and comprehensive accord was undoubtedly eased by the development of substitutes for CFCs in refrigerators, air conditioners, foaming, and cleaning solvents. In fact, the biggest producer, Du Pont, unilaterally stopped making the chemicals even before the *Protocol* required it.

Although production of ozone-destroying substances has been substantially curtailed under international agreement, more than 20 million tons of them have already been dumped into the atmosphere, and this damage cannot be undone. Because of their long lifetime and slow diffusion into the stratosphere, the synthetic chemicals that are already in the air will keep on destroying the ozone layer for about a century and full recovery will probably not occur until about 2070.

Scientists continue to use satellites to monitor the circulation, composition and temperature of the stratosphere, while also keeping a close eye on the Sun's varying ultraviolet output, which modulates ozone production. Computer models that forecast the future of ozone depletion take into account the Sun's varying ultraviolet radiation, and current measurements of the amounts of different ozone-destroying substances in the stratosphere, including the identification of new gases that might dominate ozone depletion, such a nitrous oxide. In 2006, for example, an analysis of 25 years of ozone observations at different altitudes in the stratosphere indicated that the Earth's protective ozone layer outside of polar regions stopped thinning about 10 years after the *Montreal Protocol* was first enacted.

The computer projections indicate that the Earth's ozone hole ought to eventually close, but it is going to take a while. The models suggest that the ozone hole will remain open longer than initially expected, forecasting that it will not substantially shrink until about 2020. An uneven ozone recovery is now anticipated due to the way global warming is changing the circulation of stratospheric air masses from the tropics to the poles, with over-recovery at mid-latitudes and depletion remaining over the tropics beyond 2100. This therefore brings us to the related scientific, and political, topic of global warming by heat-trapping gases dumped into the atmosphere by humans.

Global warming by human emission of heat-trapping, "greenhouse" gases

Our planet's surface is now comfortably warm because the atmosphere traps some of the Sun's heat and keeps it near the surface. The thin blanket of gas acts like a one-way filter, allowing visible sunlight through to warm the Earth's lands and oceans, but preventing the escape of some of the heat into the sink of space. Much of the ground's heat is reradiated out toward space in the form of longer infrared waves that are less energetic than visible ones and thus do not pass through the atmosphere's gas as easily as sunlight.

Some of the infrared heat radiation is absorbed in the air by "greenhouse gases", such as carbon dioxide, and some of the trapped heat is reradiated downward to warm the planet's surface and the air immediately above it.

The greenhouse effect is literally a matter of life and death. If the Earth had no atmosphere, it would be directly heated by the Sun's light to only 255 kelvin, which is well below the freezing point of water at 273 kelvin. Fortunately for life on Earth, the greenhouse gases in the air warm the planet to as much as 288 kelvin, and this extra heat can keep the oceans, lakes and streams from turning into ice. This "natural" greenhouse warming comes from minor ingredients of our atmosphere, such as water vapor and carbon dioxide, and it is perfectly normal and beneficial.

Of course, you can have too much of a good thing, like lying in the Sun all day in the summer or eating too much ice-cream too fast. For hundreds of years, humans have been filling the sky with carbon dioxide. The invisible waste gas is dumped into the air by burning fossil fuels – coal, oil and natural gas. When these materials are burned, their carbon atoms, denoted C, enter the air and combine with oxygen atoms, O, or oxygen molecules, O_2, to make carbon dioxide, CO_2.

About a century ago, in the middle of the industrial revolution, it was mainly coal that fueled factory boilers and warmed city houses, releasing carbon into the atmosphere to make carbon dioxide. Since the gas is colorless and odorless, it could not be seen or smelled, but it was detected indirectly by the noxious fumes emitted as a byproduct of burning high-sulfur coal. It has blackened entire cities,

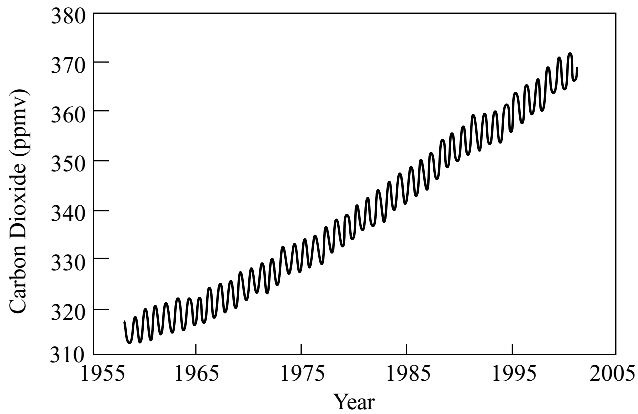


Fig. 4.21 Rise in Earth's atmospheric carbon dioxide The average monthly concentration of atmospheric carbon dioxide (CO_2) in parts per million by volume (ppmv) of dry air plotted against time in years observed since 1958 at the Mauna Loa Observatory, Hawaii. It shows that atmospheric amounts of the principal waste gas of industrial societies, carbon dioxide, have risen steadily for more than forty years. The up and down fluctuations superimposed on the systematic increase reflect a seasonal rise and fall in the absorption of carbon dioxide by trees and other vegetation. Summertime lows are caused by the uptake of carbon dioxide by plants, and the winter highs occur when the plants' leaves fall and some of the gas is returned to the air. (Courtesy of Dave Keeling and Tim Whorf, Scripps Institution of Oceanography.)

such as London, described in 1854 by Charles Dickens (1812–1870) in *Hard Times*. His account is an eerie foreboding of the dark, polluted skies that have recently been found in Beijing and Mexico City.

In the 20th century, the perfection of the internal combustion engine and the mass production of automobiles made oil one of the most important sources of atmospheric carbon, and therefore of carbon dioxide. Around the globe, cars, sports utility vehicles and trucks have been releasing huge amounts of the potentially dangerous heat-trapping gas into the air.

Every time we drive a car, use electricity from coal-fired power plants, or heat our homes with oil or natural gas, we release carbon into the lower atmosphere. The burning of forests, whose trees hold much carbon dioxide, has also contributed.

Just a few decades ago, no one knew if any of the carbon dioxide stayed in the atmosphere or if it was all being absorbed in the oceans. Then in 1958 Charles D. Keeling (1928–2005) began measurements of its abundance in the clean air at the Mauna Loa Observatory in Hawaii, showing that the amount of carbon dioxide in the atmosphere has been increasing non-stop at an accelerating rate over half a century (Fig. 4.21). Superimposed on the relatively small annual fluctuations, due to the growth and loss of plant leaves and associated carbon dioxide intake, there was a large systematic increase over the entire period of observation. Year by year the total measured concentration of carbon dioxide grew, as inexorably as the expansion of the world's population and human industry.

Moreover, studies of ice deposits in Antarctica indicate that the amount of CO_2 has been increasing at an exponential rate ever since the beginning of the industrial revolution in the mid-18th century. Air bubbles that are trapped in the ice act like time capsules, conserving the atmosphere of the past. The air was sealed in the bubbles when the ice was laid down, and can be extracted from cores drilled deep within the layered ice deposits. They indicate that the atmospheric concentration of carbon dioxide has increased 31 percent during the past two and a half centuries, a mere blink in the eye of cosmic time.

The atmosphere now contains almost 800 billion tons of carbon dioxide. Humans continue to release about 7 billion tons of it each year. In other words, each person on Earth is, on average, dumping about a ton of carbon dioxide into the air every year. Once added to the air, carbon dioxide spreads throughout the entire atmosphere. The sea will absorb about half of that carbon dioxide, but it will take decades and even centuries to disappear. So future generations will have to contend with our present activities.

Roger Revelle (1909–1991) and Hans E. Suess (1909–1993) realized the threat decades ago. They argued that the oceans might not readily absorb all of the carbon dioxide being released into the air, and that the amount of atmospheric carbon dioxide would steadily increase as the fuel and power requirements of our worldwide civilization continued to rise. With prophetic insight, they wrote in 1957 that the increase might alter our weather and climate.

The increase in carbon dioxide isn't the entire story. During the past several decades other heat-trapping gases have been accumulating noticeably in the atmosphere, such as methane (CH_4) and nitrous oxide (N_2O). Even though the total emissions of these molecules are quite small when compared with those of carbon dioxide, they are much more efficient at trapping infrared heat radiation. As a result, they can together contribute about as much global warming as carbon dioxide alone.

Methane is the same natural gas that we use at home for cooking and heating. When found in swamps, methane is known as marsh gas; it sometimes ignites spontaneously, producing flickering blue flares called will-o'-the-wisps. Some of it also escapes from coalmines, natural gas wells and leaky pipelines.

Most of the atmospheric methane does not, however, come from gas wells. It is produced by agricultural activities such as growing rice and raising cattle. The gas is emitted by bacteria that thrive in oxygen-free places like rice paddies and the stomachs of cows. Since pre-industrial times, the atmospheric concentration of methane has increased more than 110 percent. Although methane is about 200 times less abundant than carbon dioxide, each incremental molecule of methane has about 20 times the heat-trapping power as each additional molecule of carbon dioxide.

Nitrous oxide, or laughing gas, is also building up in the air, although not as rapidly as methane. The current rate of increase is about 0.2 percent a year, primarily as the result of nitrogen-based fertilizers but also from the burning of fossil fuels in cars and power plants.

The chlorofluorocarbons (CFCs) are very effective heat-trapping molecules. The addition of one CFC molecule to the air can have the same greenhouse effect as the addition of 10 000 molecules of carbon dioxide to the present atmosphere. Fortunately, the warming effect of these industrial chemicals may soon be leveling off since they have been banned on the basis of their ozone-destroying capability.

Contrary to popular misconception, however, ozone depletion and global warming are not the same thing. The CFC molecules that destroy ozone also trap heat, but the thinning of the ozone layer does not by itself make the Earth's surface hotter.

Possible consequences of Earth's rising temperature

The inexorable, irreversible build-up of carbon dioxide and other heat-trapping gases is raising temperatures across the Earth, giving it a rising fever. The Swedish chemist Svante Arrhenius (1859–1927) saw it coming, setting out in 1896 to find out what would happen if the amount of carbon dioxide were changed from the amounts then in the air. He concluded that a doubling of the atmospheric carbon dioxide would boost the Earth's temperature by a few degrees, and that the temperature would drop by almost the same amount if the gas decreased by half. Although obtained without the use of modern satellite observations or extensive computer models, Arrhenius' estimate of the global warming by doubling the amount of carbon dioxide in our air is comparable to modern estimates. He also pointed out that industrial activity was then noticeably increasing the amount of atmospheric carbon dioxide, and that humans were therefore altering the temperature of the globe.

There can be no doubt that the temperatures are already rising. The evidence comes from direct measurements of rising surface air temperatures and subsurface ocean temperatures, as well as retreating glaciers, increases in average global sea levels, and changes to many physical and biological systems. Taken alone, these events are no proof of global warming, but in combination they provide strong evidence for a warmer climate.

As the water in the sea gets warmer, it will expand as most substances do when heated. The sea will then ascend to higher levels, in much the same way that heating the fluid in a thermometer causes the fluid to expand and rise up. This is because warm water or other fluids occupy a greater volume than cold ones. Measurements indicate that the global sea level increased somewhere between 10 and 25 centimeters during the 20th century. However, you could not have noticed the change, for the sea level was only rising between 1.0 and 2.5 millimeters per year.

The melting of ice that now covers land, such as mountain glaciers, also contributes to the sea-level rise and most likely results from global warming. By the end of the 20th century, glaciers were retreating throughout the world, and those in Alaska were typically becoming about a meter thinner every four or five years. The loss of ice in Greenland and Antarctica is now increasing at an accelerated pace, and the melting of Arctic ice may eventually shut down the Gulf Stream, which carries tropical heat away from the mid-Atlantic.

Melting glacial ice releases water into streams and rivers, which add to the sea. Such melt-waters from mountain glaciers boosted the sea level between 2 and 5 centimeters in the 20th century.

Contrary to popular belief, the melting of floating icebergs will not raise the level of the surrounding sea. When ice cubes in your drink at home melt, they similarly do not cause any change in its level, for the melted ice produces the same volume of water as it displaces.

No one can see much advantage in the rising seas, which are one of the most certain effects of the warming projected during the coming decades. There is just one indirect advantage – the meltdown of polar ice, that contributes to the sea rising, could result in a permanent ice-free passage in the Arctic Ocean, providing a new shipping route between Europe and Asia.

The climate experts predict a rise in sea level of between 0.09 and 0.9 meters (3 inches to 3 feet) over the next 100 years if nothing is done to curtail the emission of greenhouse gases. The resultant flooding will seriously disrupt coastal areas where more than a quarter of the world's population now lives. In the worst-case increase, Venice and Alexandria will be inundated, as will many cities on the Atlantic and Gulf coasts of the United States, including Boston and New York City. Residents in South Florida will not have to worry about the sweltering heat; their homes will be flooded with seawater. Thirty million people in Bangladesh could be displaced by a 0.9-meter increase in sea level, and the rising waters would most likely force the evacuation of 70 million Chinese.

Salt water could move several kilometers inland at the mouths of rivers, invading coastal drinking-water systems. The Nile, Yangtse, Mekong and Mississippi deltas are all at risk. Island nations will suffer severe flooding or completely disappear under the rising waters; they include the Bahamas, many of the Caribbean islands, Cyprus and Malta in the Mediterranean, and several archipelagos around the Pacific Ocean (Fig. 4.22).

Even a modest rise in sea level will wipe many of the world's beaches out of existence. Flooding isn't the problem; it's the removal of sand by waves. Even a 0.3-meter (1-foot) rise in sea level creates wave action that might erode away up to 50 meters (150 feet) of some beaches. So people who live by the seashore had better sell their homes, and you can forget winter vacations in parts of Florida and the Caribbean islands.

The modest increase in temperatures at mid-northern latitudes, where most people live, will be welcome. There will be longer summers, shorter winters and warmer nights. Residents of cities like Boston will suffer fewer colds, experience fewer heart attacks from shoveling snow, and spend less on heating, snowplowing and road salting. On the other hand, summer air-conditioning will cost more, the winter ski slopes may turn to slush, and the colorful fall foliage could disappear as future generations of trees move away from the heat to the north.

Even if global warming is at the upper end of the predictions in 100 years, many humans should be able to adapt without much difficulty. After all, the predicted rise in temperature is less than the average daily temperature difference between New York City and Atlanta, Georgia, or between Paris and Naples. And those who live in the colder, northern locales are already used to a seasonal temperature increase between winter and summer that can be three times greater than the largest predicted heating over the next 100 years. Moreover, life has thrived during past periods when the planet was substantially warmer than it is now, and the Arctic was free of ice.

So the good thing is that humans are adaptable. But the bad thing is also that humans are adaptable. As long as the climate changes occur slowly, we can adapt without realizing what is happening, but some very uncomfortable things can happen at the top part of the expected warming in 100 years.

There is an applicable proverb about frogs. If you put a frog in boiling water, it will jump out and save itself. But if you gradually increase the heat, with the frog in the water, it will die.

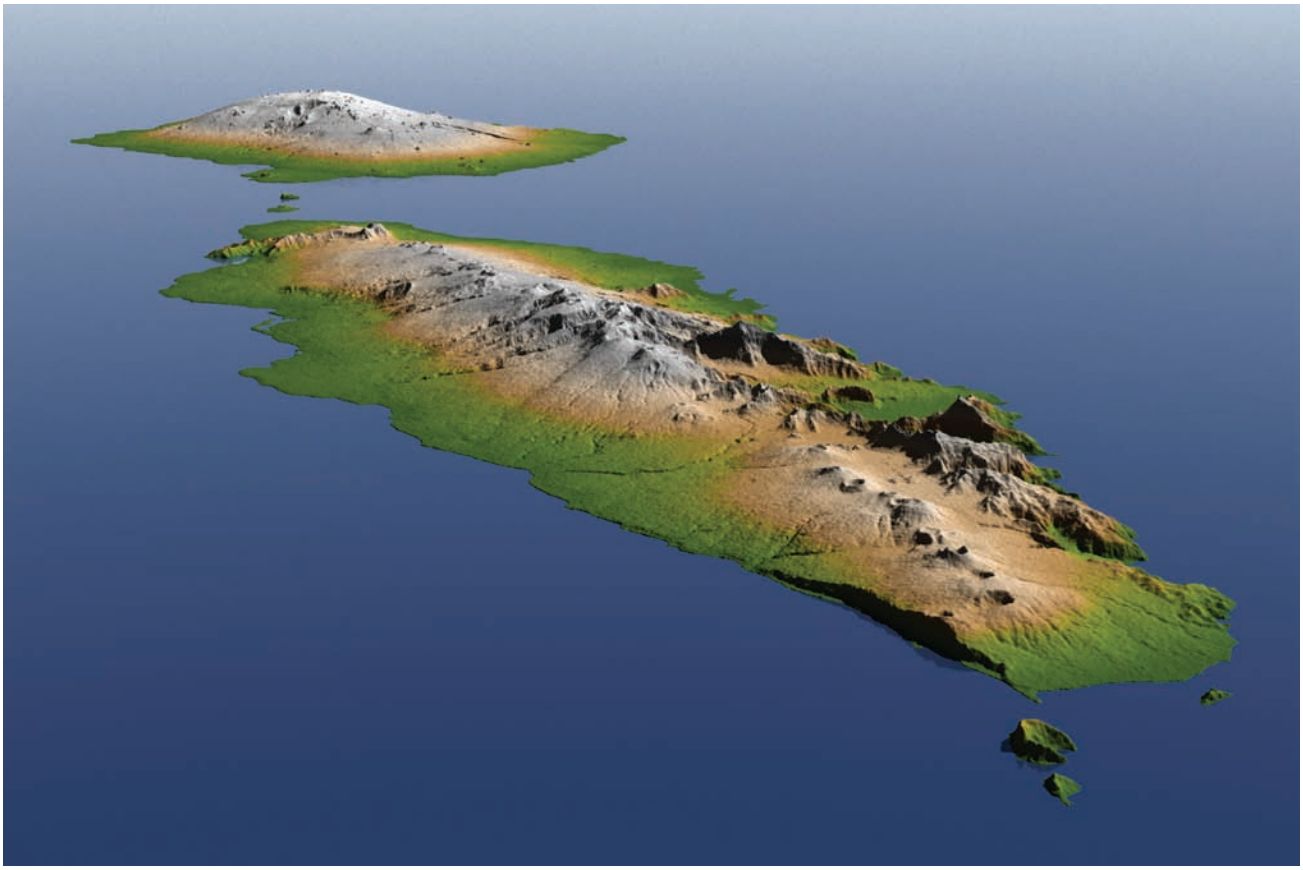


Fig. 4.22 Independent State of Samoa The topography of Savai'i (*background*) and Upolu (*foreground*), the two large islands of the Independent State of Samoa. The highest volcano rises 1.9 kilometers. The low-lying coastal lands were inundated on 29 September 2009 by a tsunami generated by a major earthquake located about 200 kilometers to the south. Villages located on the coasts are also in danger of flooding by rising seas expected from future global warming. Color-coding is directly related to topographic height, with green at the lower elevations, rising through yellow and tan, to white at the highest elevations. Bright and dark shades display topographic slope in the northeast–southwest direction. The Shuttle Radar Topography Mission aboard the *Space Shuttle Endeavor* acquired the elevation data used in this image on 11 February 2000. (Courtesy of NASA/JPL/NGA.)

Very hot will be decidedly unwelcome in many places. Within deserts, entire cities will be immobilized under the heat. The wealthy will move out of Palm Springs, and Las Vegas could become a ghost town. Residents in many other large cities should experience severe heat waves, making them feel like the world is melting down in a pool of sweat. As the climate becomes hotter and drier, drought will probably become more severe in areas prone to it, and supplies of freshwater will dwindle. More frequent bouts of extreme weather will also be expected, with widespread flooding and intense hurricanes.

Agriculture in some regions will be better than other regions. The longer growing season and increasing carbon dioxide will foster plant growth, making much of the developed world greener. Agriculture will likely become more productive in Canada, northern Europe, Russia and the northern United States. As droughts turn some mid-American farms to dust, both agricultural production and population centers in the United States will shift north, and the same thing will probably happen in Europe.

The world's poorest countries are, on the other hand, highly vulnerable to agricultural disaster, for they are already located in arid and semi-arid regions. A further rise in temperature will almost certainly reduce crop yield in south Asia and sub-Sahara Africa, where expanding deserts will additionally claim more land.

As environmental conditions change over time, plants and animals will migrate, as they have throughout geological history – moving up and down in latitude as the globe warms and cools. The recent increase in temperatures has already caused some species to move north, and the accelerated heating could wipe out many of them in the future. Some plants and animals might not be able to move fast enough to keep pace with the rapid rate of temperature change, and climate-sensitive habitats could

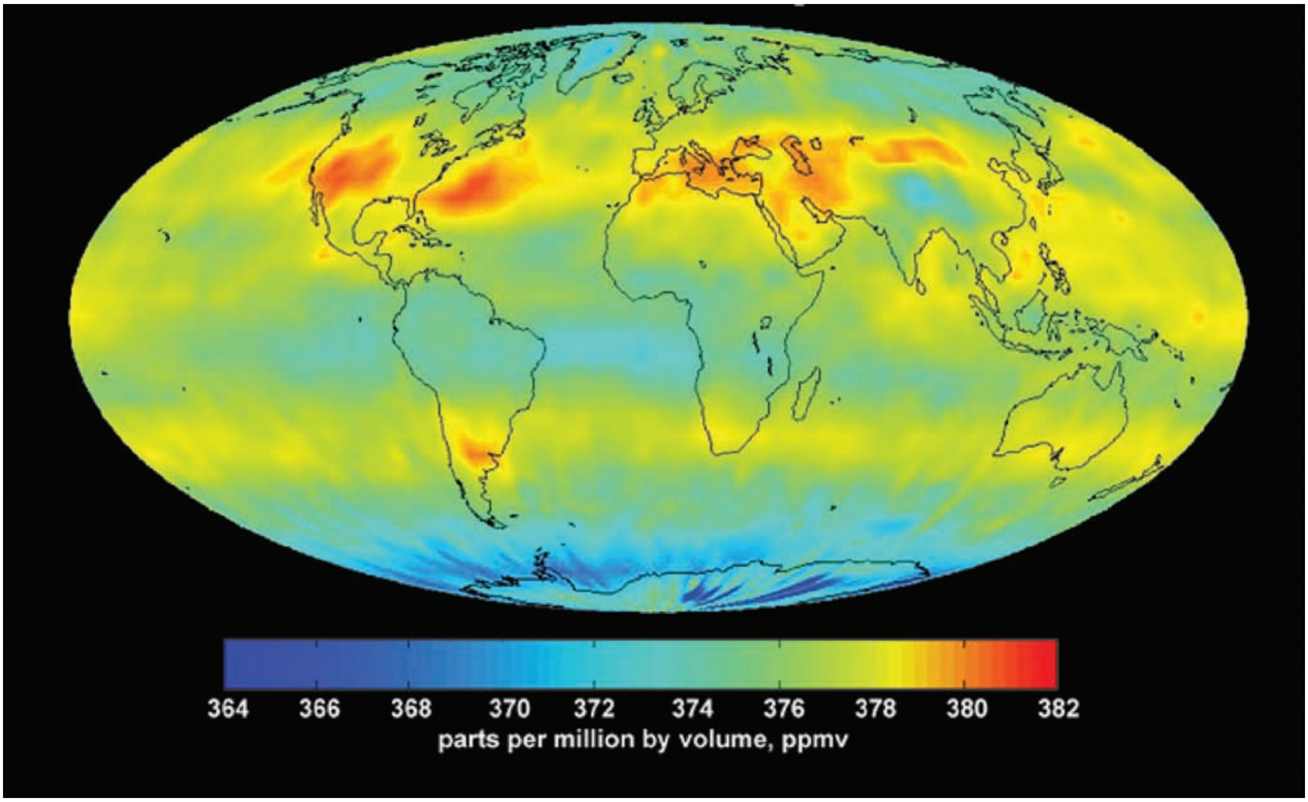


Fig. 4.23 Carbon dioxide in the Earth's troposphere The concentration of carbon dioxide about 8 kilometers above the Earth's surface in July 2003. The regional distribution can still be seen by the time that the gases reach this height in the mid-troposphere, with large concentrations above parts of the United States and China. This image was acquired from the Atmospheric Infrared Sounder Experiment aboard the *Aqua* spacecraft. (Courtesy of NASA/JPL.)

be destroyed altogether, hastening the extinction of some species.

And there are other catastrophes that might result from significant global warming in the future. Hurricanes will become stronger and wetter; water supplies will be reduced and forest fires will become more common; more species will become extinct; drought will be intensified within the interiors of many continents; the American Midwest might become a colossal dust bowl; and power companies will be unable to air-condition sweltering cities.

The last factor in our catalog of likely consequences of severe global warming is increased health risk – a topic close to the hearts of most people and nearly every politician. Many diseases might spread dramatically as the temperatures head upward, especially those carried by mosquitoes. As the world warms, mosquitoes will move north, into regions where the winter cold used to kill them, bringing malaria, dengue fever, yellow fever and encephalitis with them. Extremely hot weather may also directly terminate the lives of a lot of people, particularly among the very old and very young, the poor and weak, and those with cardiovascular and respiratory disease.

Doing something about global warming

Although the most severe consequences of global warming are not likely to be noticed by you or your children, we've already initiated changes that will affect future generations. Once in the atmosphere, about half of the carbon dioxide stays there for centuries, so our grandchildren and their children will have to contend with the consequences of our present actions. The invisible waste gases that we have already dumped in the air will slowly change the climate of the Earth regardless of future actions.

The scientists are doing all they can to improve our understanding of global warming. They continue to monitor signs of the rising temperatures, such as the meltdown of Antarctica, retreating glaciers and rising sea levels. Instruments aboard orbiting satellites are now monitoring the global distribution and temporal changes of heat-trapping gases in the atmosphere (Figs. 4.23, 4.24). And climate scientists continue to refine their models, using measurements of increasing accuracy to predict the future in intricate computer calculations.

When you strip away the rhetoric, the experts are uncertain about the exact future severity of climate change and

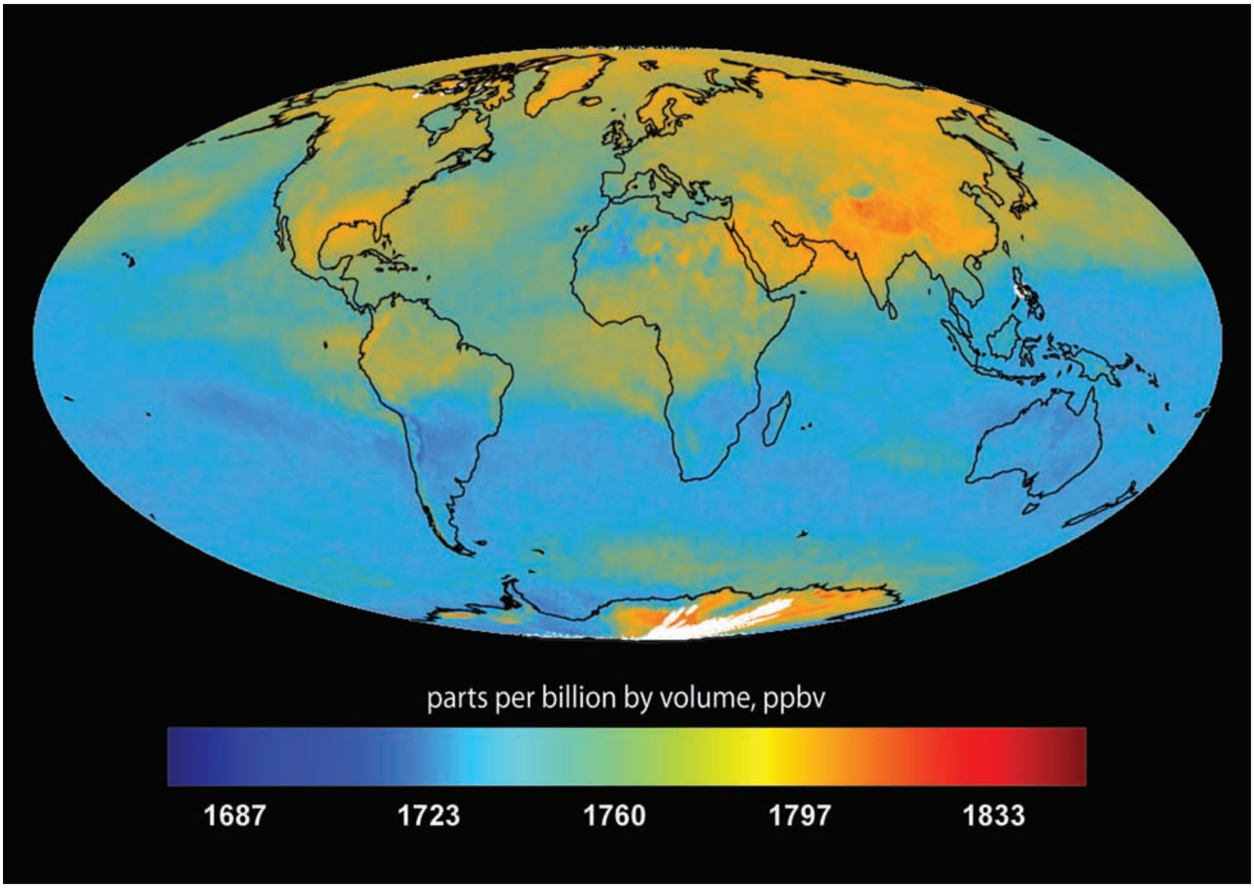


Fig. 4.24 Methane in the Earth's troposphere The global distribution of methane in the lower parts of the Earth's atmosphere, known as the troposphere. This image was taken from the Atmospheric Infrared Sounder Experiment aboard the *Aqua* spacecraft in August 2005. Such images will help identify sources of this heat-trapping greenhouse gas, as well as its seasonal and multi-year variation and its transport around the globe. (Courtesy of NASA/JPL.)

even less about the future physical impact in particular countries or regions. After decades of research, the model builders cannot say precisely what will happen to the climate as the result of the atmospheric build-up of heat-trapping gases. They just don't know enough about the atmosphere, clouds or the oceans to predict accurately the future global climate. Despite the lack of precision, the general trends have nevertheless become obvious, with uncontestable observations of global warming.

The majority of scientists are therefore now in agreement that we ought to do something to curtail human emissions of carbon dioxide and other heat-trapping gases. The chances of serious consequences are high enough to take action to stop their probable effects. In 2005, for example, the world's most influential scientific academies warned national leaders that they can no longer ignore the clear and increasing threat posed by global warming, and that the scientific understanding of climate change was then sufficiently clear to justify nations taking prompt action. The unprecedented joint statement included the heads of the scientific academies of Brazil, Canada, China, France, Germany, India, Italy, Japan, Russia, the United Kingdom, and the United States.

Then in 2007 the Nobel Peace Prize was awarded, in two equal parts, to the Intergovernmental Panel on Climate Change (IPCC), and Albert Arnold "Al" Gore Jr. (1948–) for their efforts to build up and disseminate greater knowledge about man-made climate change and to lay the foundations for the measures that are needed to counteract such change. The Nobel presentation speech recognized the importance of understanding and containing man-made global warming that will decisively impact our existence on Earth.

Most scientists support prudent steps to curb the continued build-up of heat-trapping gases, even asserting that the evidence warrants a sense of urgency. Yet, whether we like it or not, global warming has become politicized, the subject of a contentious debate. It has entered the arena of world politics, a shadowy realm of diplomacy, economic interests, political alliances, and national security.

So what's being done about the problem? In December 1997, representatives of the world's nations met in Kyoto, Japan, to establish, for the first time, specific legally binding targets and timetables for the emission of heat-trapping gases. The treaty, called the *Kyoto Protocol*, calls for mandatory reductions in the emissions of greenhouse gases such as carbon dioxide and methane. But it has not been signed by the United States, which contributes about 18 percent of the total emissions with just 4 percent of the world's population. And developing nations are not bound by the treaty restrictions even though their emissions are expected to surpass even the unrestrained emissions of the richer nations in a few decades. China, for example, has already overtaken the United States as the world's largest carbon-dioxide emitter on an annual basis, and the combined increase in greenhouse gases contributed by these two countries will outstrip any reductions agreed to by other countries.

More recently, heads of state and government of more than a hundred nations gathered in Copenhagen, Denmark, between 7 December and 18 December 2009 to seek a consensus on an international strategy for fighting global warming and associated climate change. Despite the serious intent of global leaders and meaningful proposals for reductions of carbon-dioxide emissions, if a binding agreement was made, the United Nations' sponsored Copenhagen Summit fell short of even modest expectations.

No legally binding treaty was ratified by any nation. A draft *Copenhagen Accord* asks countries to submit emission targets, but it was not adopted or passed unanimously, and it did not contain any legally binding commitments for reducing carbon-dioxide emissions.

The main difficulty in arriving at a more meaningful agreement lies in a long-standing division between the rich and poor nations. The United States, whose emissions of carbon dioxide over the past 200 years has helped cause global warming, has long refused to accept any binding limits on its greenhouse-gas emissions. Fast-developing China, a "poor" nation, is also strongly opposed to mandatory ceilings on the emissions. For China, and also India, the main concern is economic growth, which delivers stability and prosperity, and keeps the government in power. Since that growth involves burning more coal, which releases large amounts of carbon dioxide into the atmosphere, these countries are going to find it difficult to substantially reduce their emissions. Coal provides roughly half of the electricity in the United States and most of that in China, and the inexpensive fuel will help move the populations of China and India from poverty to middle-class prosperity.

Both the United States and China probably find it unacceptable to approve a treaty that might cause serious economic damage. Wealthy countries fear that mandatory limits on the emissions of heat-trapping gases could cause a recession in their prosperous economies, while the poor ones are afraid that such limits will destroy the economic growth needed for the very survival of their people. And since neither side in the debate will compromise, a consensus is impossible.

To put it another way, no one likes to be told what to do, and voluntary steps to improve the climate will be opposed by the vested interests of governments, corporations and other powerful institutions. As a result, no comprehensive, ratified, international global-warming treaty exists, at least so far. As always, the greatest hope is in the young, who are contributing to a mass movement on behalf of climate action.

Something important may nevertheless result from the future interactions of the calm, compromising diplomats and politicians. Most major nations – including the United States, the 27 nations of the European Union, China, India, Japan and Brazil, for example, restated, in January 2010, their earlier pledges to curb emission of heat-trapping gases by 2020, some by promising absolute cuts, others by reducing the rate of increase.

As the traditional Negro jubilee spiritual goes:

He's got the whole world in his hands,
The whole wide world in his hands,

but the "he" is not a supernatural deity – it's us. We humans have modified the atmosphere, warming the globe, and we are starting to do something about it. But it is bound to be only a temporary fix.

In the long run nature will take over the weather and climate once again. A hundred million years ago, when the dinosaurs roamed the Earth, there were no ice caps and tropical plants flourished near the South Pole. Deep cold nearly turned the Earth into a ball of ice about 10 000 years ago, when the planet was in the depths of an ice age. In just a few million years from now, entire continents and oceans can be destroyed or created new, changing the flow of air and ocean currents and altering global weather patterns. And even if it is pretty warm right now, the die is cast for the next ice age and the glaciers will come again.

Ice and fire

During the past million years the Earth has undergone a series of warm and cold periods. During the cold periods, called ice ages, huge ice sheets build up on the continents and in the polar seas. The growing layer of continental ice flows towards the equator, scouring and covering large areas of land. Then the climate warms and the ice retreats.

Each glacial ice age lasts about 100 000 years. There is a relatively short interval of unusual warmth between the ice ages that lasts 10 000 or 20 000 years. During such an interglacial interval, the world's climate becomes more pleasant and serene. We now live in such a warm time, called the Holocene period, which has enabled human civilization to flourish.

The recurrent ice ages and warm intervals are caused by variations in the amount and distribution of sunlight reaching the Earth, but not by any intrinsic fluctuations in the amount of light radiated by the Sun itself. Three astronomical cycles combine to alter the angles and distance at which sunlight strikes the far northern latitudes of Earth, triggering the ice ages. This explanation was fully developed by the Serbian astrophysicist Milutin Milankovitch (1879–1958) from 1920 to 1941, so the astronomical cycles are now sometimes called the Milankovitch cycles.

When there is less sunlight being received in far northern latitudes, the winter temperatures are colder there, and the summer temperatures are milder. So less polar ice melts in the summer, and over time the winter snows are compressed into ice to make the glaciers grow.

The varying gravitational forces produced by the other planets, whose distances from Earth change, produce a rhythmic stretching of the Earth's orbit. These planetary perturbations periodically change the shape of the Earth's orbit from circular to slightly elliptical and back again, over a period of 100 000 years. As its path becomes more elongated, the Earth's distance from the Sun varies more during each year, intensifying the seasons in one hemisphere and moderating them in the other.

Shorter cycles are due to repetitive changes in the wobble and tilt of the Earth's rotational axis, which vary over 23 000 and 41 000 years respectively. The greater the tilt, the more intense the seasons in both hemispheres, with hotter summers and colder winters.

Successive layers of frozen atmosphere have been laid down in Greenland and Antarctica, providing a natural archive of the Earth's past climate over the past 420 000 years. Bubbles of air trapped in falling snowflakes and entombed in ice are deposited every year, building up on top of each other like layers of sediment. When extracted in deep ice cores, they reveal secrets about the ancient climate. Such cores strongly support the idea that changes in the Earth's orbit and spin axis cause variations in the intensity and distribution of sunlight arriving at Earth, which in turn initiate natural climate changes and trigger the ebb and flow of glacial ice.

The current Holocene interglacial, which has already lasted 11 000 years, may not continue for more than a few thousand years, and we could then enter an ice age. The next time it happens, the advancing glaciers might bury Copenhagen, Detroit and Montreal under mountains of ice, and because of the drop in sea level people might then walk from England to France, from Siberia to Alaska, and from New Guinea to Australia.

Perhaps global warming by human activity will help counteract a coming ice age – no one knows for sure. And there is no way out in the long run, for the Sun will inevitably fry the Earth. Well-accepted models of stellar evolution indicate that the Sun began its life about 4.5 billion years ago shining with about 70 percent of the brightness it has today, and that it has been slowly increasing in brightness ever since. As the Sun continues to brighten, the planet will eventually become a burned-out cinder, a dead and sterile place.

Astronomers calculate that the Sun will become hot enough in 3 billion years to evaporate the oceans away, and 4 billion years thereafter the Sun will balloon into a giant star, engulfing the planet Mercury and melting the Earth's surface. Thus, our very remote descendants are destined to an end in fire, consumed by the Sun that once nurtured us.

So our long-term prospects aren't all that great, and we might as well concentrate on protecting, improving and experiencing the magnificent world that we are so privileged to inhabit. And to get back to more immediate concerns, the Sun provides dangers whenever humans or their spacecraft venture into space.

4.5 Space weather

The Sun is the ultimate power source. It warms the ground we walk on, lights our days, sustains life, and provides directly or indirectly most of the energy on our planet. And it is solar heat that powers the winds and cycles water from sea to rain, the source of our weather and arbiter of our climate. Nowadays, and in all former times, it is the Sun-driven seasons that dominate weather on Earth.

Once it was realized that the space between the Sun and Earth is not empty, and just more rarefied than our transparent atmosphere, it was natural to suppose that the Sun also powered space weather. The term refers to conditions on the Sun, in the Sun's winds, and near the Earth that can affect space-borne and ground-based technological systems and human life and health.

As our civilization deploys ever more sophisticated technology, it becomes increasingly at the mercy of storms in space. Its gusts and squalls, the cosmic equivalent of terrestrial blizzards or hurricanes, are related to explosive outbursts on the Sun, and to dynamic processes in interplanetary space, in near-Earth space and in the magnetosphere.

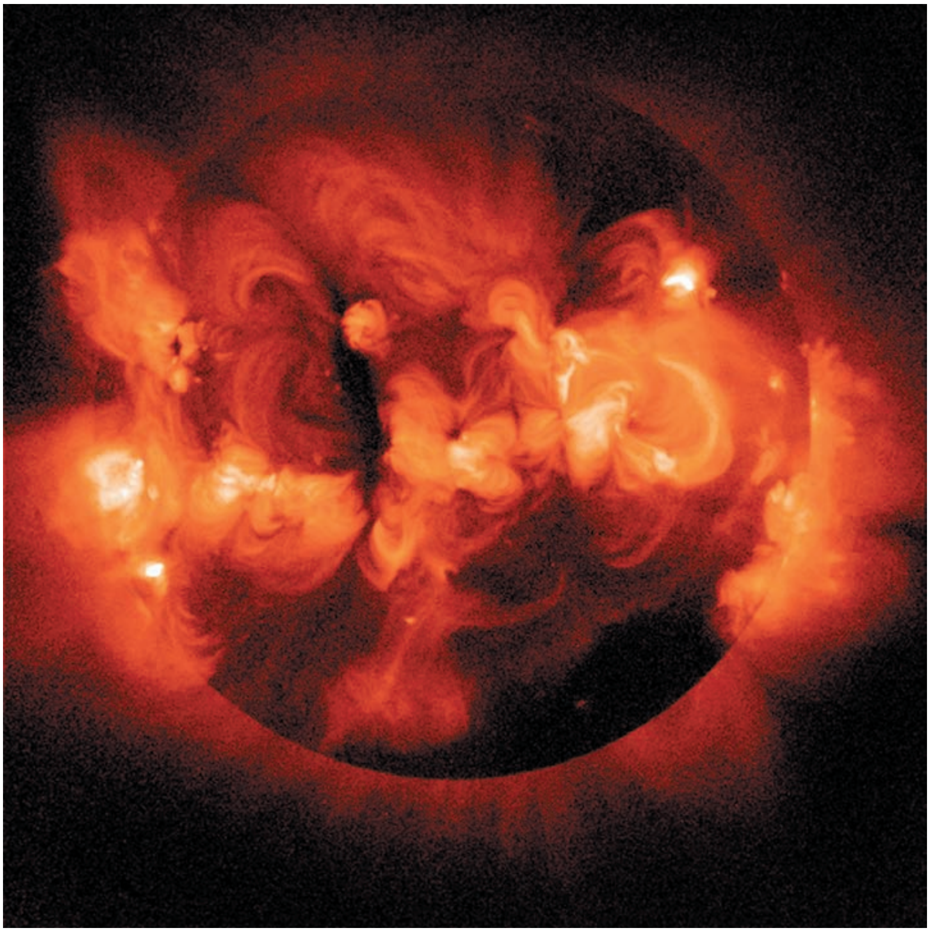


Fig. 4.25 The Sun in X-rays The bright glow seen in this X-ray image of the Sun is produced by ionized gas at a temperature of a few million kelvin. It shows magnetic coronal loops that thread the corona and hold the hot gas in place. The brightest features are called active regions and correspond to the sites of the most intense magnetic field strength. The Soft X-ray Telescope (SXT) aboard the Japanese *Yohkoh* satellite recorded this image of the Sun's corona on 1 February 1992, near a maximum of the 11-year cycle of solar magnetic activity. Subsequent SXT images, taken about five years later near activity minimum, show a remarkable dimming of the corona when the active regions associated with sunspots have almost disappeared, and the Sun's magnetic field has changed from a complex structure to a simpler configuration. (Courtesy of Gregory L. Slater, Gary A. Linford, and Lawrence Shing, NASA, ISAS, the Lockheed-Martin Solar and Astrophysics Laboratory, the National Astronomical Observatory of Japan, and the University of Tokyo.)

Down here on the ground, we are shielded from much of this space weather by the Earth's atmosphere and magnetic fields, keeping us from bodily harm. But out in deep space there is no place to hide, and both humans and satellites are vulnerable. Energetic protons accelerated by explosions on the Sun can cripple spacecraft and seriously endanger unprotected astronauts that venture into outer space. Sun storms can also disrupt global radio communications and disable satellites used for navigation, military reconnaissance or surveillance, and communication, from cell phones to pagers, with considerable economic, safety and security consequences. This technology has become part of our everyday lives, enhancing our vulnerability to space weather and increasing the importance of understanding and predicting it.

High-flying humans at risk from solar explosions

Powerful bursts of radiation, as well as energetic charged particles and magnetic fields, are being hurled into interplanetary space by solar explosions. These outbursts come in two main varieties: solar flares, and coronal mass ejections (CMEs). Both kinds of solar activity are powered by the Sun's magnetic energy, and they both vary in step with the Sun's 11-year cycle of magnetic activity. Solar flares and CMEs are more frequent and tend to be more powerful during the maximum in the activity cycle.

All of the solar flares, and most of the fastest coronal mass ejections, with the largest amount of energy,

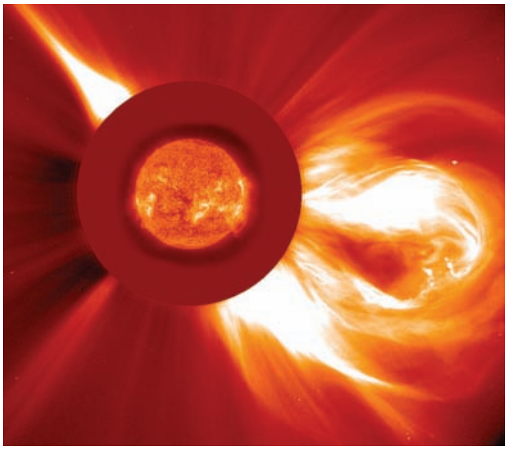


Fig. 4.26 Mass ejection from the Sun A huge coronal mass ejection is seen in this coronagraph image, taken on 5 December 2003 with the Large Angle Spectrometric COronagraph (LASCO) on the *SOlar and Heliospheric Observatory* (*SOHO*). The solid red circle corresponds to the occulting disk of the coronagraph that blocks intense sunlight and permits the corona to be seen. An image of the singly ionized helium, denoted He II, emission of the Sun, taken at about the same time, has been appropriately scaled and superimposed at the center of the LASCO image. The full disk helium image was taken at a wavelength of 30.4 nanometers, corresponding to a temperature of about 60 000 kelvin, using the Extreme-ultraviolet Imaging Telescope, or EIT for short, aboard *SOHO*. (Courtesy of the *SOHO* LASCO and EIT consortia. *SOHO* is a project of international cooperation between ESA and NASA.)

originate from places of intense magnetism on the Sun. Known as solar active regions, they contain closed magnetic fields that are rooted in sunspots and whose loops constrain the intense X-ray-emitting gas of the quiescent, non-exploding Sun (Fig. 4.25).

Solar flares are brief catastrophic outbursts that flood the solar system with intense radiation and high-speed electrons and protons. In just a few minutes they can release an explosive energy of up to 10^{25} joule, equivalent to 20 million 100-megaton terrestrial nuclear bombs, raising the temperature of Earth-sized regions on the Sun to tens of millions of kelvin. The other type of solar explosive activity, the CMEs, expand away from the Sun at speeds of hundreds of kilometers per second, becoming larger than the Sun and removing up to 50 billion tons, or 5×10^{13} kilograms, of the Sun's atmosphere (Fig. 4.26).

At any given phase of the solar cycle, intense solar flares are as much as 100 times more frequent than mass ejections, but the CMEs energize particles on a grand scale that covers large regions in interplanetary space (Fig. 4.27). They move straight out of the Sun and flatten everything in their path, like a gigantic falling tree or a car out of control. Fast CMEs plow into the slower-moving solar wind and act like a piston that drives shock waves ahead of them, accelerating electrons and protons as they go, like ocean waves propelling surfers.

High-energy protons from a solar flare or coronal mass ejection can easily pierce a spacesuit, causing damage to human cells and tissues. The explosive solar emissions can endanger the health and even the lives of astronauts when they venture into outer space to construct a space station, repair a spacecraft, or walk on the Moon or Mars.

Solar astronomers, and employees of national space-weather forecast centers, therefore keep careful watch over the Sun during space missions, to warn of possible solar activity occurring at just the wrong place and time. Space flight controllers can then postpone space walks during solar storms, keeping astronauts within the heavily shielded recesses of a satellite or space station. The astronauts would also be told to curtail any strolls on the Moon or Mars, and to move inside underground storm shelters.

Failing to communicate on Earth

Eight minutes after the outburst of an energetic flare on the Sun, a strong blast of X-rays and extreme ultraviolet radiation reaches the Earth and radically alters the structure of the planet's upper atmosphere, the ionosphere, by producing an increase in the amount of free electrons that are no longer attached to atoms. Even during moderately intense flares, long-distance radio communications can be temporarily silenced over the Earth's entire sunlit hemisphere. The radio blackouts are particularly troublesome for the commercial airline industry, which uses radio transmissions for weather, air traffic and location information; the United States Air Force and Navy are also concerned about this solar threat to radio communications.

The Air Force operates a global system of ground-based radio and optical telescopes and taps into the output of national, space-borne X-ray telescopes and particle detectors in order to continuously monitor the Sun for intense flares that might severely disrupt military communications and satellite surveillance.

Space-weather interference with radio communication can be avoided by using short-wavelength, ultra-high-frequency signals that pass right through the ionosphere to satellites that can relay the transmissions to other locations. But the telecommunications industry is also threatened by the loss of their satellites due to disabling solar outbursts.

Earth-orbiting satellites in danger

Solar energetic particles arising from solar flares or coronal mass ejections can degrade, disrupt or destroy a satellite.

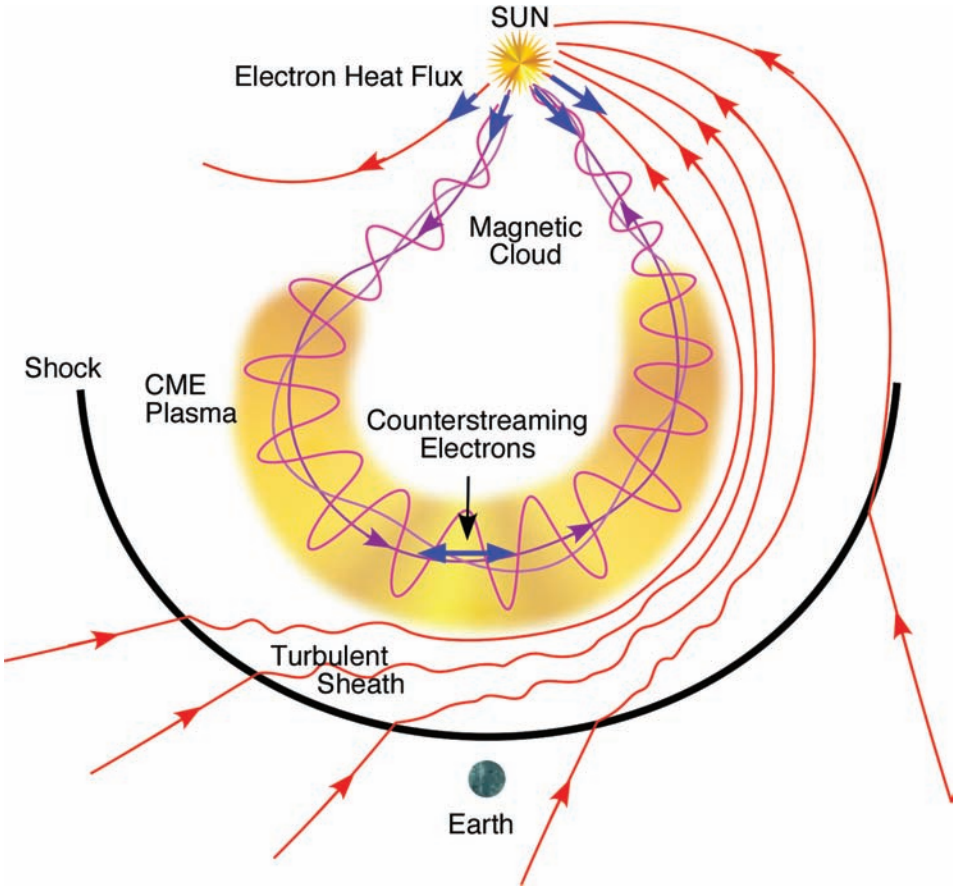


Fig 4.27 Magnetic cloud from the Sun hits the Earth When a coronal mass ejection travels into interplanetary space, it can create a huge magnetic cloud containing bidirectional, or counterstreaming, beams of electrons that flow in opposite directions within the magnetic loops that are rooted at both ends in the Sun. The magnetic cloud also drives an upstream shock ahead of it. Magnetic clouds are only present in a subset of observed interplanetary coronal mass ejections. (Courtesy of Deborah Eddy and Thomas Zurbuchen.)

And there are now roughly 1000 of them in daily use by governments, corporations and ordinary citizens. Geosynchronous satellites, which orbit the Earth at the same rate that the planet spins, stay above the same place on Earth to relay and beam down signals used for aviation and marine navigation, cellular phones, global positioning systems, national defense, and internet commerce and data transmission. Other satellites whip around the planet, scanning air, land and sea for environmental change, weather forecasting and military reconnaissance. All of these spacecraft can be temporarily or permanently disabled by solar energetic particle events, causing engineers to design spacecraft with greater shielding and increased redundancy in their components.

Geosynchronous satellites, for example, are endangered by the coronal mass ejections that cause intense geomagnetic storms. These satellites orbit our planet once every 24 hours at an altitude of 35 786 kilometers, or a distance of about 6.6 Earth radii from the planet's center, and thus remain at constant longitude above the Earth. A powerful coronal mass ejection can compress the magnetosphere from its usual location at about 10 Earth radii to below the satellites' geostationary orbits, exposing them to the full brunt of the gusty solar wind and its charged, energized ingredients.

Infrequent, anomalously large eruptions on the Sun can hurl very energetic protons toward the Earth and elsewhere in space, interfering with satellite instruments (Fig. 4.28). The solar protons can easily enter a spacecraft to produce single-event upsets in electronic components by ionizing a track along parts of their circuits. The ionized tracks can occur in transistors and memory devices, producing erroneous commands and crippling their microelectronics. Such single-event upsets have already destroyed at least one weather satellite and disabled several communications satellites.

Space weapons can also wipe out a satellite; so if you didn't know the Sun was at fault, you might think someone was trying to shoot down the satellites. But error-correcting software has been developed to decrease damage by

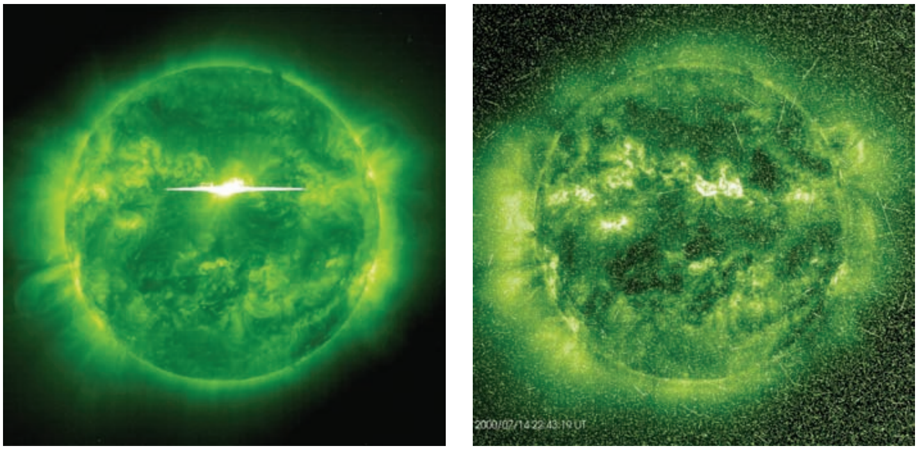


Fig. 4.28 Solar flare produces threatening space storm of energetic particles A powerful solar flare (*left*), occurring at 10 hours 24 minutes Universal Time on Bastille day, 14 July 2000, unleashed high-energy protons that began striking the *SOHO* spacecraft near Earth about 8 minutes later, continuing for many hours (*right*). Both images were taken from the Extreme-ultraviolet Imaging Telescope (EIT) on the *SOlar and Heliospheric Observatory* (*SOHO*). (Courtesy of the *SOHO* EIT consortium. *SOHO* is a project of international cooperation between ESA and NASA.)

single-event upsets to military satellite operations. Commercial satellites, which are less expensive than military ones to build, have less protection and are more vulnerable.

Solar threats to electrical power systems on Earth

While altering the Earth's magnetic field, a colliding coronal mass ejection can produce strong electric currents in nearby space. If these currents connect to long-distance power lines on the ground, they can blow circuit breakers, overheat and melt the windings of transformers, and cause massive failures of electrical distribution systems. They can plunge major urban centers, like New York City or Montreal, into complete darkness, causing social chaos and threatening safety. The threat is greatest in high-latitude regions where the currents are strongest, such as Canada, the northern United States and Scandinavia.

Forecasting space weather

Our technological society has become increasingly vulnerable to explosions on the Sun. They emit energetic particles, intense radiation, powerful magnetic fields and strong shocks that can have enormous practical implications when directed toward Earth. The solar emissions can disrupt navigation and communication systems, pose significant hazards to humans in space, destroy Earth-orbiting satellites, and create power surges that can black out entire cities. Recognizing our vulnerability, national centers and defense agencies continuously monitor the Sun from ground and space to forecast threatening activity. An example is the Space Environment Center (SEC) of the United States National Oceanic and Atmospheric Administration. It collects and distributes space weather data, using satellites and ground-based telescopes to monitor the Sun and interplanetary space.

With adequate warning, operators can power down sensitive electronics on navigation and positional satellites, putting them to sleep until the danger passes. Airplane pilots and cellular telephone customers can be warned of potential communication failures. The launch of manned space flight missions can be postponed, and walks outside spacecraft or on the Moon or Mars might be delayed. Utility companies can reduce load in anticipation of induced currents on power lines, in that way trading a temporary "brown-out" for a potentially disastrous "black-out".

What everyone wants to know is how strong the storm is and when it is going to hit us. Most of the more energetic coronal mass ejections come from magnetic explosions in active regions with sunspots, producing a flare in tandem with the ejections. So a good place to begin our space weather forecasts is to know when a threatening active region, with its sunspots and strong magnetic fields, is on the Sun.

Active regions appear more frequently near the maximum of the 11-year sunspot cycle, as do solar flares and coronal mass ejections. So long-term solar activity can be forecast in a general way using this cycle.

On a shorter timescale of weeks, we can use helioseismology to detect large solar active regions on the hidden backside of the Sun. The technique of helioseismology uses observations of solar pulsations to infer the

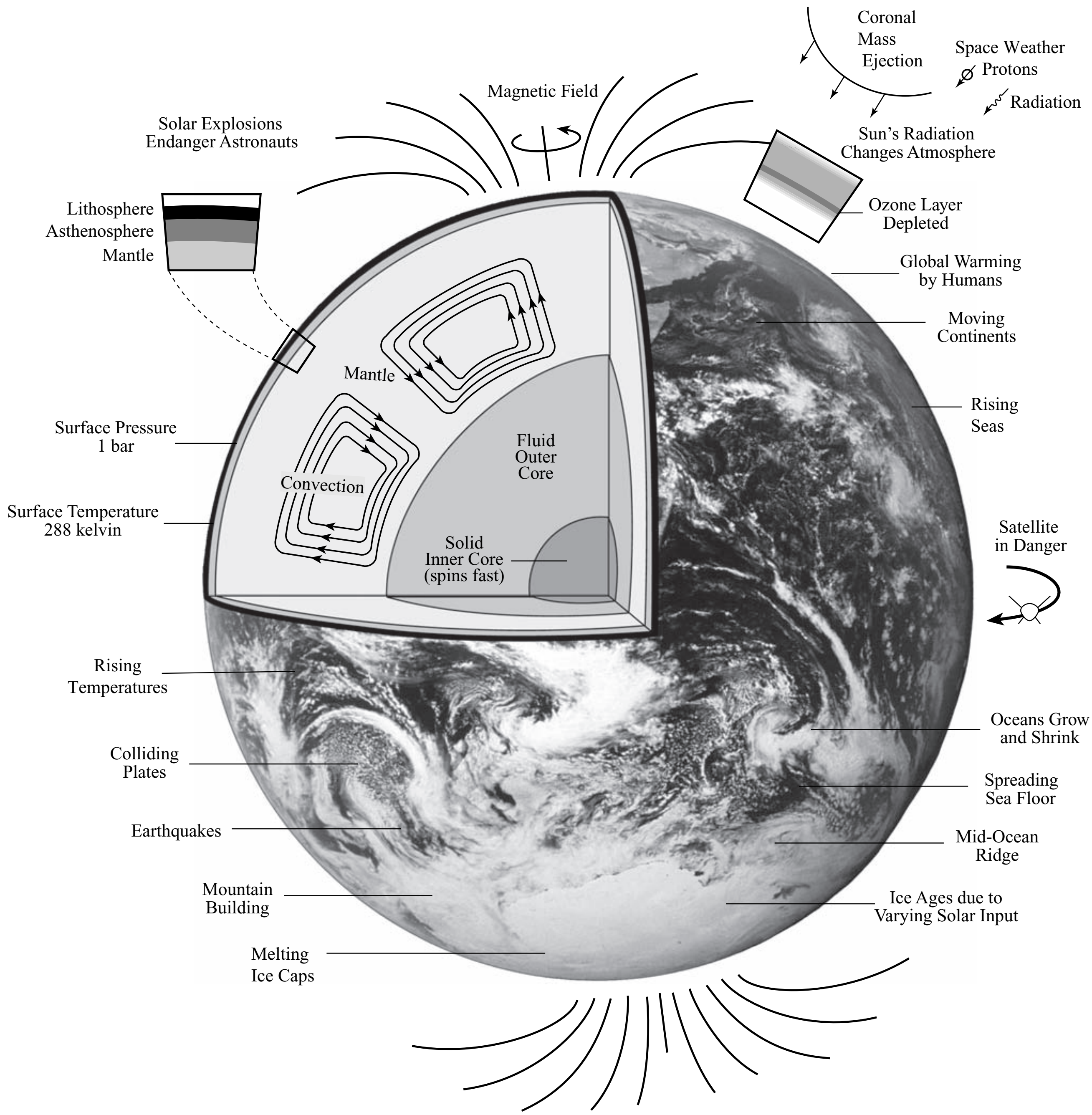


Fig. 4.29 Summary diagram

trajectories of sound waves within the Sun, including those that see through it. Since the solar equator rotates with a period of 27 days, when viewed from Earth, the detection of a magnetically complex and strong active region on the far side of the Sun can give more than a week's warning before it swings into view to threaten the Earth. Daily images of the unseen, far side of the Sun are available on the web at http://soi.stanford.edu/data/full_farside and http://gong.nso.edu. Like winter storms on Earth, some of the effects of space weather can be predicted days in advance. A coronal mass ejection arrives at the Earth one to four days after leaving the Sun, and solar astronomers can watch them happen days in advance. Solar flares are another matter. As soon as you can see a solar flare on the Sun, its radiation and fastest particles have already reached us, taking just 8 minutes to travel from the Sun to Earth. Dangerous but less energetic particles might take an hour to get here.

The ultimate goal of space weather forecasters is to predict when the Sun is about to unleash its pent-up energy, before a solar flare or coronal mass ejection occurs. One promising technique is to watch to see when the Sun's magnetism has become twisted into a stressed situation, for it may then be about to explode. Observing the X-ray emission of an active region to determine when the magnetic fields are sheared and twisted can do this. The signature of an immanent explosion might be found deeper down, under the visible disk of the Sun. The techniques of local helioseismology have demonstrated that the strength of flares from active regions is correlated with the amount of circulating flows beneath them.

However, some regions that exhibit magnetic shear and twist never erupt, so contorted magnetism may be a necessary but not sufficient condition for solar flares or coronal mass ejections. And the Sun's sudden and unexpected outbursts often remain as unpredictable as most human passions. They just keep on happening, and even seem to be necessary to purge the Sun of pent-up frustration and to relieve it of twisted, contorted magnetism.

And to be honest, scientists have not solved the question of what exactly initiates a solar flare or coronal mass ejection, igniting the explosion from stressed coronal magnetic fields. They think the storms might be triggered when magnetized coronal loops are pressed together, driven by motions beneath them, meeting to touch each other, merging to break open the magnetic fields and release free magnetic energy. But no one has identified a signature that allows prediction of exactly when such an outburst might occur. So far, we only have signs of a possible solar storm; it's something like seeing that dark storm cloud but not knowing if it's going to rain.

5 The Earth's Moon: stepping stone to the planets

- When the Moon moves into the Earth's shadow, the full Moon turns blood red; when the Earth travels into the Moon's shadow it can become dark during the day.
- The full Moon looks bigger near the horizon than directly overhead, but its changing size is an illusion.
- The Moon spins on its axis with the same period in which it revolves around the Earth, at 27.3 days, keeping its far side forever hidden to Earth-bound observers.
- The near side of the Moon contains light, rugged, cratered regions called highlands and dark smooth lava flows dubbed maria; the far side of the Moon is mostly highlands and has very few maria.
- For more than two centuries, lunar craters were attributed to volcanoes on the Moon, but they are now widely known to be due to the explosive impact of interplanetary projectiles, known as meteors when in space and meteorites upon hitting the surface of a moon or planet.
- More than thirty years ago, twelve humans roamed the surface of the Moon and brought back nearly half a ton of rocks.
- Because the Moon has almost no atmosphere, its sky remains pitch black in broad daylight and there is no sound or weather on the Moon.
- Two modest spacecraft, named *Clementine* and *Lunar Prospector*, chalked up an impressive list of accomplishments in the 1990s, including evidence for a lunar core and for water ice at the poles of the Moon.
- Rocks returned from the Moon contain no significant amounts of water, but there is evidence for small quantities of water in some places such as the permanently shaded regions at the lunar poles. Comets may have deposited the water.
- Space agencies from China, Europe, India, Japan, and the United States have all sent spacecraft to the Moon in the early 21st century, obtaining detailed information about the altitude, geological, chemical and gravity characteristics of the lunar surface and sending their spacecraft into controlled impact with the Moon.

- High-resolution maps acquired from lunar orbit are being used to specify potential landing sites and resources for future human exploration of the Moon.
- Humans might return to the Moon to create unique astronomical observatories, and establish a permanent base and way station for trips to Mars; but that is not likely to happen in the near future.
- Moonquakes, which are much weaker than earthquakes, indicate that the Moon has a small dense core, probably surrounded by a partially molten zone. The core has been confirmed by gravity measurements from the orbiting *Lunar Prospector* spacecraft, and laser-ranging measurements have confirmed the molten zone.
- There is no life on the Moon, and there apparently never was any.
- Earth rocks and Moon rocks are similar in their mix of light and heavy oxygen isotopes, but the Moon rocks contain relatively little iron and few volatile elements common on Earth.
- Impact basins excavated by cosmic collision produce as much topographical relief on the Moon as there is on the Earth due to ongoing tectonic processes.
- Vast blocks of the lunar surface are magnetized, but they do not combine into an overall global dipole like the Earth's magnetism. Some of the ancient lunar magnetism has been concentrated on the other side of the Moon from large impact basins.
- Radioactive dating indicates that the oldest rocks returned from the Moon are about 4.6 billion years old, which is about the same age as the Earth.
- During its early youth, between 4.4 and 4.6 billion years ago, a global sea of molten rock covered the Moon, but now a layer of fine, powdery Moon dust covers it.
- A heavy bombardment cratered the highlands until about 3.9 billion years ago, when the large impact basins were formed; lunar volcanism subsequently filled these basins to create the maria between 3.2 and 3.9 billion years ago.
- Most of the features we now see on the Moon have been there for more than 3 billion years.
- The Moon's gravity draws the Earth's oceans into the shape of an egg, causing two high tides as the planet's rotation carries the continents past the two tidal bulges each day.
- The Moon acts as a brake on the Earth's rotation, causing the length of the day to steadily increase and the Moon to move away from the Earth.
- The Moon provides a steadying influence to the Earth's seasonal climatic variation, anchoring and limiting the tilt of the planet's rotation axis.
- The Moon was most likely born during the ancient collision of a Mars-sized body with the young Earth; the giant impact dislodged material that would become the Moon that we know.

5.1 Fundamentals

The Earth has one Moon, which is a natural satellite. To distinguish it from the moons of other planets, we denote our Moon with a capital M and also sometimes call it the Earth's Moon. The mass, size, density and other physical properties of our Moon are given in Table 5.1.

The Earth's Moon is a unique satellite within the solar system, the largest relative to its planet. Mars is the only other terrestrial planet to have a moon, and its two satellites are very small. The giant planets have extensive satellite systems, but these moons are usually composed of low-density rock-ice mixtures unlike our high-density rocky Moon.

Table 5.1 Physical properties of the Moon[a]

Mass	7.348×10^{22} kilograms = 0.0123 M_E
Mean radius	1737.5 kilometers = 0.2725 R_E
Bulk density	3344 kilograms per cubic meter
Sidereal rotation period	27.322 days = fixed star to fixed star
Sidereal orbital period	27.322 days = fixed star to fixed star
Synodic month	29.53 days = new Moon to new Moon
Mean distance from Earth	3.844×10^8 meters
Increase in mean distance	0.0382 ± 0.0007 meters per year
Mean orbital speed	1023 meters per second
Angular radius at mean distance (geocentric)	15 minutes 32.6 seconds of arc
Angular radius at mean distance (topocentric)	15 minutes 48.3 seconds of arc
Age	4.55×10^9 years

[a] Here M_E and R_E respectively denote the mass and radius of the Earth. The Earth to Moon mass ratio is 81.300 587.

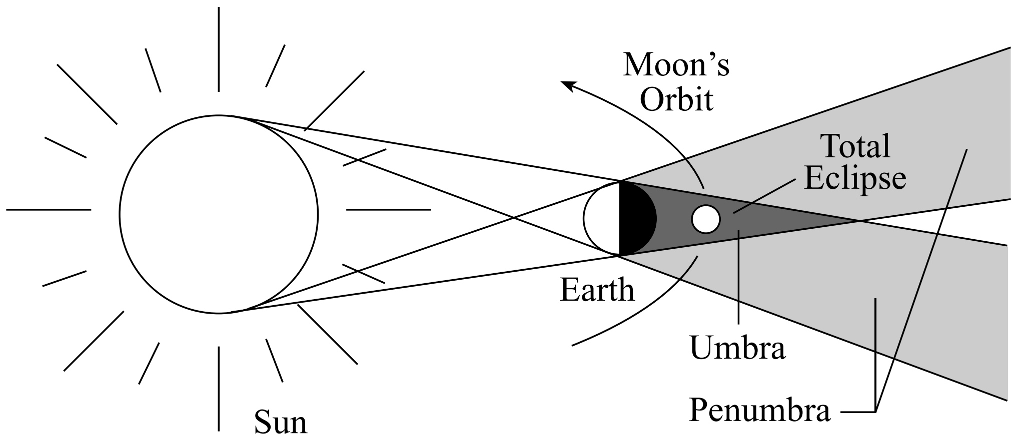


Fig. 5.1 Lunar eclipse During a lunar eclipse the initially full Moon passes through the Earth's shadow. A total lunar eclipse occurs when the entire Moon moves into the umbra. Because no portion of the Sun's visible disk can be seen from the umbra, it is the darkest part of the Earth's shadow. Only part of the Sun's disk is blocked out in the larger penumbra. A partial lunar eclipse occurs when the Moon's orbit takes it only partially through the umbra or only through the penumbra.

5.2 Eclipses of the Moon and the Sun

Once or twice in a typical year, the Moon's orbital motion carries it through the Earth's shadow. This is an eclipse of the Moon, when the Sun's illumination of the Moon has been removed. The word *eclipse* is derived from the Greek term for "abandonment".

A lunar eclipse can be seen from half of the Earth. There are two regions in the Earth's shadow at the time of a lunar eclipse: the *umbral* region where there is no direct sunlight, and the *penumbral* region where the Sun's light is partially shadowed (Fig. 5.1). The umbral shadow is darker, and it is in the shape of a narrow cone pointing away from the Earth.

The full Moon turns a deep red when in the umbral shadow of the Earth (Fig. 5.2). Ancient Hebrew writers often used this appearance as a metaphor to describe the end of the world. For instance, the prophet Joel declared that the Lord:

> will shew wonders in the heavens and in the Earth, blood, and fire, and pillars of smoke. The Sun shall be turned to darkness, and the Moon into blood . . .
>
> (Joel 2:30, 31; KJV)

And then there are the lyrics to some of Bob Dylan's (1941–) songs that include "the Moon rising like wildfire", and "when there was blood on the Moon".

A total eclipse of the Sun occurs when the Moon passes between the Earth and the Sun, and the Moon's shadow falls on the Earth. In an incredible cosmic coincidence, the Moon is just the right size and distance to blot out the visible solar disk when properly aligned and viewed from

Fig. 5.2 The blood-red Moon If the Earth had no atmosphere, the Moon would disappear in darkness during a total lunar eclipse. As shown here, the Moon actually becomes dark red for an hour or so. This is because the Moon is illuminated by sunlight that is bent part way around the Earth and is reddened in passing through the Earth's atmosphere, just as the Sun is reddened at sunset. If the Earth is heavily clouded, the sunlight is obstructed and the Moon is particularly dark during a lunar eclipse. (Courtesy of Eric Mandon, Observatoire Populaire de Rouen.)

the Earth. In other words, the apparent angular diameter of the Moon and the visible solar disk are almost exactly the same, about 30 minutes of arc, so that under favorable circumstances the Moon's shadow can reach the Earth and cut off the light of the Sun.

The outer atmosphere of the Sun, known as its corona, becomes momentarily visible to the unaided eye when the Moon blocks out the Sun's bright disk and it becomes dark during the day. The corona is then seen at the limb, or apparent edge, of the Sun, against the blackened sky as a faint, shimmering halo of pearl-white light (Fig. 5.3). But be careful if you watch an eclipse, for the light of the corona is still very hazardous to human eyes and should not be viewed directly.

Since the Moon and the Earth move along different orbits whose planes are inclined to each other (Fig. 5.4),

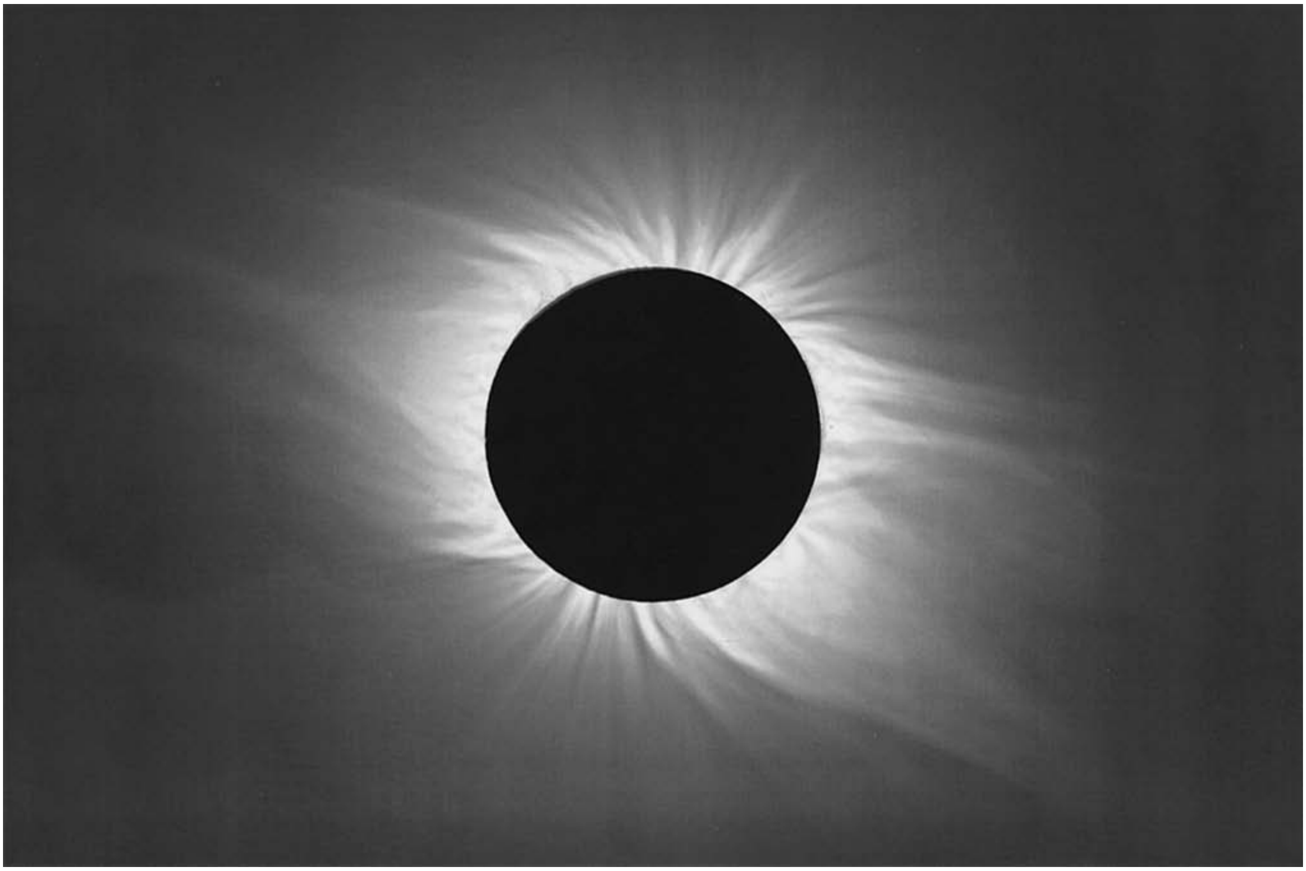


Fig. 5.3 Gossamer corona The Sun's corona as photographed during the total solar eclipse of 26 February 1998, observed from Oranjestad, Aruba. To extract this much coronal detail, several individual images, made with different exposure times, were combined and processed electronically in a computer. The resultant composite image shows the solar corona approximately as it appears to the human eye during totality. Note the fine rays and helmet streamers that extend far from the Sun and correspond to a wide range of brightness. (Courtesy of Fred Espenak.)

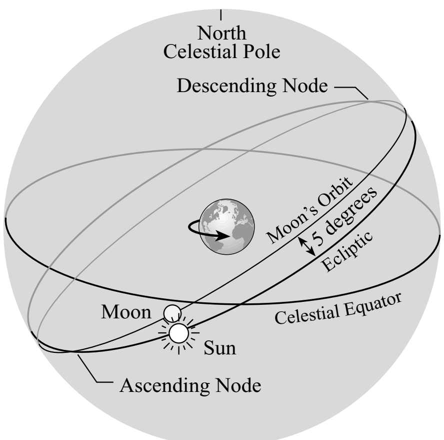


Fig 5.4 Celestial paths of the Moon and Sun The orbit of the Earth's Moon is tilted 5 degrees to the Sun's route across the sky, the ecliptic, allowing these paths to cross at two nodes. These are the only points at which eclipses can occur. During a lunar eclipse the Moon and Sun are located at opposing nodes, so that the Moon can move into the Earth's shadow cast by the Sun. A solar eclipse occurs when the Moon and Sun cross paths at the same node.

a total eclipse of the Sun does not happen very often. The Moon only passes between the Earth and the Sun about three times every decade on average. Even then, a total eclipse occurs along a relatively narrow region of the Earth's surface, where the tip of the Moon's shadow touches the Earth (Fig. 5.5). At other nearby places on the Earth, the Sun will be partially eclipsed, and at more remote locations you cannot see any eclipse of the Sun.

The Moon's orbital motion carries its shadow rapidly eastward across the ground at about 1600 kilometers per hour. As a result, the longest total eclipse of the Sun observed at a fixed point on the ground lasts just under eight minutes.

If the Moon is at a distant part of its orbit at the time of solar eclipse, the Moon appears smaller than the Sun, and the tip of the Moon's shadow does not quite reach the Earth. The bright ring of the Sun's disk is then seen around the edge of the Moon. This is an annular eclipse, and it has none of the darkness and excitement of a total eclipse.

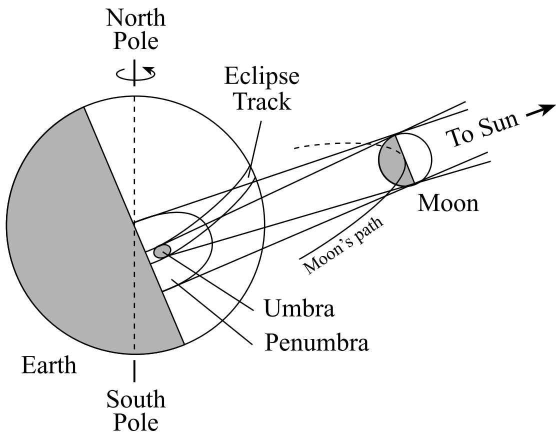


Fig. 5.5 Solar eclipse During a solar eclipse, the Moon casts its shadow upon the Earth. No portion of the Sun's photosphere can be seen from the umbra region of the Moon's shadow (*small gray spot*); but the Sun's light is only partially blocked in the penumbra region (*larger half circle*). A total solar eclipse, observable only from the umbra region, traces a narrow path across the Earth's surface.

5.3 The Moon's face

When a full Moon rises or sets, it is a captivating sight. It looks huge, dwarfing everything in the foreground (Fig. 5.6), and as the song goes, "when the Moon hits your eye like a big pizza pie, that's amore". But appearances can be deceiving. The Moon is no bigger when it is close to the horizon than when it is high in the sky. Its changing apparent size may be an illusion caused by comparing the Moon to other objects when it is viewed along the ground.

This so-called Moon illusion seems to arise from the way that the brain deals with apparent distance, not size. When people view the Moon near the horizon, there are large foreground objects, such as trees, buildings and hills, for comparison, so the Moon looks very far away and huge. When the Moon is overhead, alone in an otherwise empty sky, there are no other objects to gauge its distance; the Moon then appears to be closer and we think it is smaller than at the horizon (Fig. 5.7). Our perception of the dome of the background sky may also play a role in the Moon illusion.

Artists often portray the Moon's face in all of its round fullness, and the full Moon is the subject of all kinds of myths and superstitions (Focus 5.1).

Although it appears bright in contrast to the night sky, the Moon's face is as dark as the asphalt on highways and darker than most rocks on Earth. The fraction of incoming sunlight that is reflected from the lunar surface is known as its *albedo*, and values range from 5 to 10 percent for the darkest regions and 12 to 18 percent for the brightest ones. On average, the lunar surface reflects just 12 percent

Fig. 5.6 An enormous Moon In this awesome picture, a man and child seem enveloped by the Earth's Moon, which looks huge in comparison to the tree in the foreground. When the Moon is overhead, alone in an otherwise empty sky, there are no other objects to gauge its distance; we then think the Moon is smaller than at the horizon. (Courtesy of der Foto-Treff.)

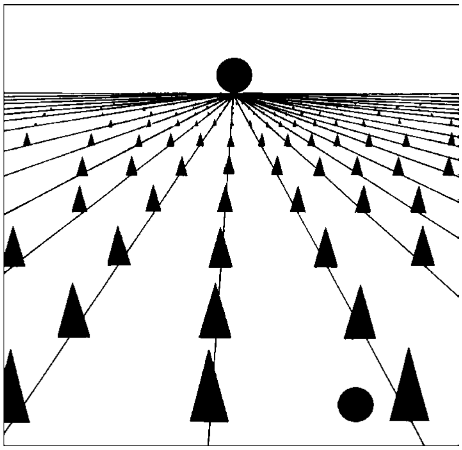


Fig. 5.7 Moon illusion We make decisions about size because of our perceptions of distance. The two black disks in this figure are the same size, but we see the bottom one as smaller because we think it is closer. The top disk seems larger because it appears to be farther away. The Moon on the horizon is similarly thought to be huge because comparisons with objects on the ground make us think it is far away – also see Fig. 5.6. When people look straight up at the Moon, in an otherwise empty sky, they no longer have land clues to compute the Moon's distance and it is perceived as being closer and smaller.

of the sunlight that strikes it, which makes it one of the least shiny objects in the solar system.

Earth-bound observers always see the same side of the Moon. We call this the near side of the Moon, in contrast to the far side, not visible from Earth. The far side of the Moon is not its dark side, for it is illuminated by sunlight in the same way as the near side.

Gravitational interaction between the Earth and its Moon tie the two together, and lock the Moon's rotation into synchronism with its orbital motion. Our planet's greater gravitational pull on the near side of the Moon brakes the Moon's rotation and holds it in place like an invisible string, so the Moon's sidereal rotational period is precisely equal to its sidereal orbital period of 27.322 days, the time it takes for the Moon to return to a given position among the stars. In other words, the Earth's gravity has synchronized the Moon's rotation with its orbital motion, so the Moon rotates on its axis once each orbit. This condition, in which the spin of one body is precisely equal to, or synchronized with, its revolution around another body, is known as a synchronous orbit.

You can demonstrate synchronous rotation by holding a ball at arm's length and slowly turning around. As your body completes one rotation you always see the same side of the ball, but the ball has completed one rotation while revolving once about your body. But you can't

Focus 5.1 Full Moons

Ancient Greeks thought marriages consummated during a full Moon would be prosperous and happy. In England, a distinction was made between lunacy and insanity; the former happened only during a full Moon, while the latter was permanent. Yet there is no scientific evidence that people become abnormally crazy at the time of full Moon. The Navajo Indians believed that a woman is more likely to give birth during a full Moon because of its pull on the amniotic fluid. It has even been rumored that a male child is more likely to be conceived when aided by the extra gravitational pull of a full Moon. Of course, the pull of lunar gravity depends only on the Moon's mass and distance, and has no direct connection with the amount of sunlight that we see illuminating it.

A full Moon is considered unlucky on Sunday; the Sun's day, but lucky on Monday, whose name is derived from Moon Day. The phrase "once in a blue Moon" refers to the second full Moon in a single month, which typically occurs every few years. The reason for the rarity is that the 29.5306 days between full Moons is just slightly shorter than the average month of 30.4369 days in the average year of 365.425 days.

A blue Moon is also the fourth full Moon in a season, which normally has three, and by the way, the last time a month elapsed without at least one full Moon was in February 1866, an event that will not repeat itself for 2.5 million years.

At the time of harvest Moon, the full Moon rises at sunset, providing extra time for farmers to harvest their crops. According to folklore, the harvest Moon also appears bigger than other full Moons, but this is because it stays close to the horizon. A full Moon that illuminates the landscape all through the night is known as a hunter's Moon. All full Moons look bigger near the ground than directly overhead, a visual effect known as the Moon illusion.

The yellow color of a rising full Moon is due to scattering of light in the great thickness of air near the direction of the horizon; the haze and humidity of summer air can provide an orange color. It has even been suggested that the term "honeymoon" derives from the amber-colored full Moons of June, but the origin of the word "honeymoon" dates back to the 16th century, when the first month of marriage was said to be the sweetest.

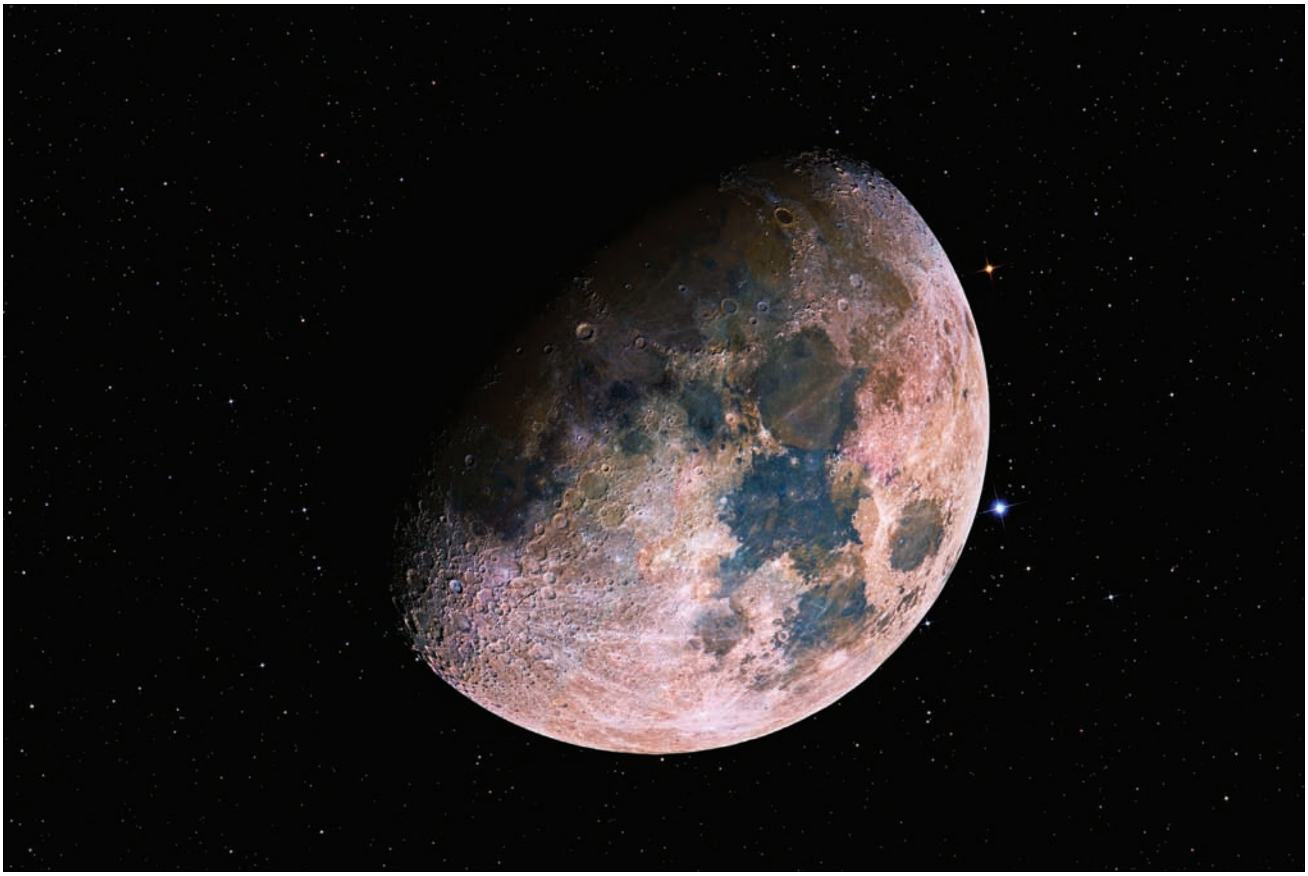


Fig. 5.8 The gibbous Moon The gleaming light of a gibbous Moon is shown in this digital combination of several high-resolution Earth-based images and a representative background star field. Though not visible to the eye, even with a telescope, the color differences at various places on the lunar surface are real. They correspond to regions with different chemical compositions that have been carefully mapped using spectrometers in satellites orbiting the Moon. (Courtesy of Noel Carboni.)

watch someone else do this; you have to demonstrate it to yourself.

Although the same side of the Moon always faces the Earth, this doesn't mean that one side of the Moon is always dark. Like the Earth, the Moon gets its light from the Sun, and sunlight always illuminates one half of the Moon. As the Moon orbits the Earth we see varying amounts of its illuminated near side. When we see a full Moon, the near side is in sunlight and the far side is dark. And when a new Moon is seen from Earth, the near side is dark and the far side is in full sunlight. In between new and full Moon we see either a crescent Moon (concave) or a gibbous Moon (bulging out from the sunlit side; Fig. 5.8). Thus, although there is a "dark side of the Moon", it is not equivalent to the "far side".

The Moon is the only planetary body that can be distinguished with the unaided eye as a globe, and even without a telescope you can tell that its surface is not uniform. Its face contains large irregular features of light and dark material (Fig. 5.9), familiarly known as the "Man in the Moon".

It wasn't until Galileo Galilei (1564–1642) turned his primitive telescope to the Moon that it became clear that our satellite is rugged and mountainous like the Earth. When he looked closely at the division between light and shadow – day and night – on the globe of the Moon, Galileo discovered that the dividing line was ragged and that he could see high mountain peaks casting long pointed shadows. When sunlight strikes the lunar surface obliquely, every mountain, hill or valley is sharply delineated.

Since he had clear evidence that the Moon was not the perfectly smooth crystalline sphere that had been proclaimed in the writings of Aristotle (384–322 BC), Galileo could write in 1610 that:

> the surface of the Moon is not perfectly smooth, free from inequalities and exactly spherical, as a large body of philosophers considers with regard to the Moon and other [heavenly] bodies, but on the contrary, it is full of inequalities, uneven, full of hollows and protuberances. It is like the surface of the Earth itself, which is varied everywhere by lofty mountains and deep valleys.

Fig. 5.9 The full Moon Our Moon glows by light it reflects from the Sun, orbiting the Earth about once a month, where the "mon" in month is short for moon. The gleaming light of the full Moon, shown here, has beckoned humanity since ancient times, pulling us closer to the heavens and out into the cosmos. It lies suspended in space, always apart yet inextricably linked to the Earth. Terrestrial observers always see the same near side of the Moon. This view of the near-side full Moon enhances the contrast between the dark maria and the bright craters. The dark circular Mare Imbrium (Sea of Rains) is prominent in the northwest (*upper left*), immediately above the bright rays of craters Copernicus and Kepler (*middle left*). The dark circular Mare Serenitatis (Sea of Serenity) lies to the east (*right*) of Imbrium. (Photo courtesy of UCO/Lick Observatory.)

But do not be deceived by these Earthly comparisons. The conspicuous mountain ranges on the Moon were thrown up about 4 billion years ago as rims of impact basins gouged out by immense cosmic collisions, and the lunar mountains have nothing to do with the plate tectonics that created the much younger terrestrial mountains.

The Moon's rough terrain is mostly confined to the brighter regions that Galileo called *terrae*, Latin for "lands"; they are now known as the highlands because they are higher than the dark regions (Fig. 5.10, left).

Galileo also discovered that the dark patches are smooth and level, resembling seas seen from a distance. He called them *maria*, the Latin word for "seas"; *mare*, pronounced "MAHrey", is the singular for "sea" (Fig. 5.10, right). However, we now know there are no substantial amounts of water in the maria. The dark maria cover about 17 percent of the lunar surface. When spacecraft were sent past the Moon to look at its averted face, they found that the far side contains very few maria (Fig. 5.11). Altogether, The heavily cratered highlands cover more than 80 percent of the Moon's total surface.

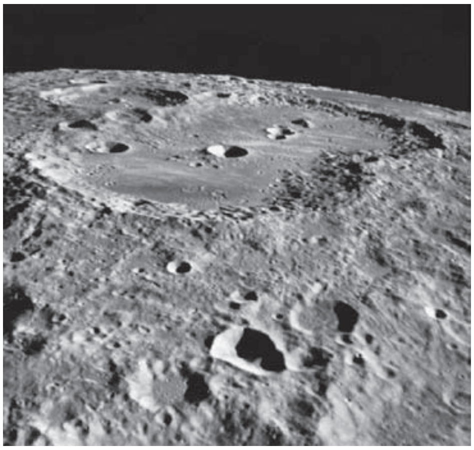
Chemical examination of rock samples returned from the Moon has shown that the maria are ancient volcanic outflows composed of dark lava. This material flowed out from inside the Moon to fill large impact basins that were formed at about the same time as the lunar highlands (Fig. 5.12; Table 5.2). One of them, the Imbrium Basin that contains Mare Imbrium, now forms the "eyesocket" in the face of the "Man on the Moon"; it has a diameter of 1500 kilometers.

Craters form one of the most striking features of the Moon's landscape (Fig. 5.13). There are at least 30 000 of them with a diameter greater than one kilometer. The word *crater* is derived from the Greek word for "cup or bowl", and it is a good description of the bowl-shaped depressions. They are just beyond the limit of visibility with the unaided eye, but a pair of binoculars will reveal a few of the larger ones. When seen through a telescope, the bright highlands are resolved into an enormous number of overlapping craters that have been visible to generations of telescope-using observers.

At around the time of full Moon, a pair of binoculars will also show bright streaks that radiate from several craters like the spokes of a wheel. These are the lunar rays, and the debris of crater formation produced them. Some of the rays go more than one-quarter of the way around the Moon (Fig. 5.14).

The ubiquitous craters were thought to be of volcanic origin throughout the 19th century and well into the 20th century, but they are now widely known to be due to the explosive impact of interplanetary projectiles (Focus 5.2). Although the maria are filled with ancient volcanic outpourings of molten rock, it spread out rapidly and did not build up in one place. So there are no large volcanoes or calderas on the Moon.

Unlike the Earth, where erosion and tectonic processes tend to obscure the effects of impact and to destroy its ancient surface rocks, the surface of the Moon preserves a pristine record of an ancient bombardment extending back several billion years. Even the youngest rocks on the Moon are as old as some of the oldest rocks found on Earth, about 3.2 billion years. Many of the planets and other satellites in the solar system bear the scars of a similar ancient rain of debris, providing a common element in their history.



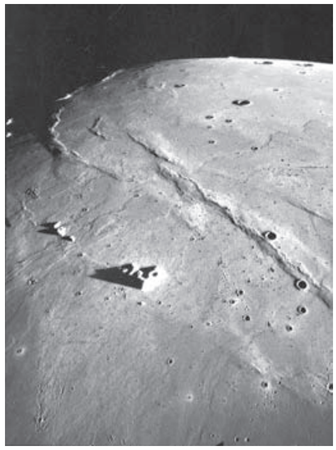


Fig. 5.10 Rough highlands and smooth maria on the Moon (*Left*) The heavily cratered lunar highlands are illustrated in this *Apollo 17* image of the Van de Graaff crater. It is located on the far side of the Moon, at the northeast edge of Mare Ingenii. This unusual crater formation seems to be composed of two merged craters with no intervening rim, and the surrounding region is saturated with craters upon craters formed during an intense bombardment of the Moon about 3.9 billion years ago.
(*Right*) Lunar volcanism is seen frozen into place in this *Apollo 15* image of Mare Imbrium, the Sea of Rains. The oblique perspective enhances individual lava flows and the relief of lunar maria ridges. The lunar maria contain relatively few craters when compared with the lunar highlands. The maria formed a secondary crust on the Moon, when lava filled the large impact basins over a period of several hundred million years ending around 3.2 billion years ago. The fluid spread rapidly, creating thin extensive sheets rather than piling up to form volcanoes. (Courtesy of NASA.)

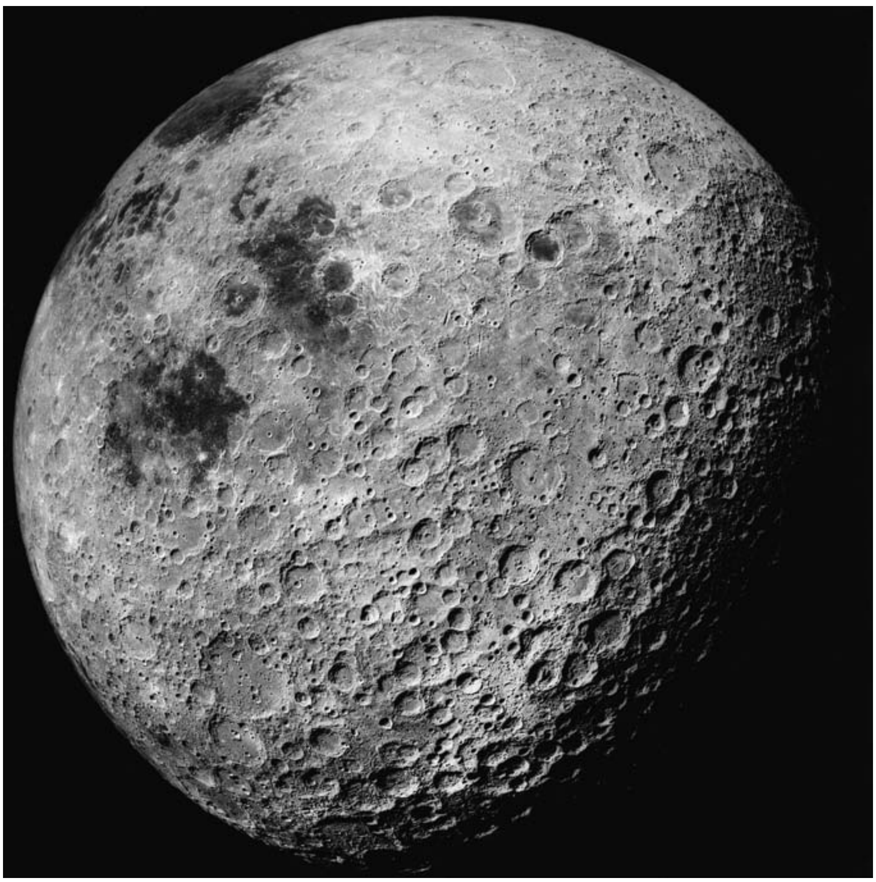


Fig. 5.11 Far side of the Moon Locked into synchronous rotation by tidal interaction with the Earth, our Moon always presents its familiar near side to us. The far side, which remains invisible from Earth, is seen from Moon-orbiting spacecraft. This image, taken from *Apollo 16*, shows the eastern edge of the near side (*left*) and the rough, heavily cratered far side of the Moon, which contains fewer smooth, dark lunar maria than the near side. This is most likely because the far-side crust is thicker than the near-side crust, so molten material, or magma, have greater difficulty in flowing to the surface to form smooth maria on the far side. (Courtesy of NASA.)

Table 5.2 Large impact basins and maria on the Moon[a]

Maria (Latin)	Seas (English)	Basin diameter (kilometers)
Oceanus Procellarum	Ocean of Storms	3200
Mare Imbrium	Sea of Rains	1500
Mare Crisium	Sea of Crises	1060
Mare Orientale	Eastern Sea	930
Mare Serenitatis	Sea of Serenity	880
Mare Nectaris	Sea of Nectar	860
Mare Smythii	Smyth's Sea	840
Mare Humorum	Sea of Moisture	820
Mare Tranquillitatis	Sea of Tranquility	775
Mare Nubium	Sea of Clouds	690
Mare Fecunditatis	Sea of Fertility	690

[a] Dating of rocks returned from the Moon indicate that the Imbrium, Serenitatis and Nectaris impacts occurred 3.85, 3.87 and 3.92 billion years ago.

Fig. 5.12 Orientale impact basin on the Moon The Orientale impact basin, shown here in a *Lunar Orbiter IV* image, is nearly 1000 kilometers across. It was probably created during an intense bombardment of the Moon, about 3.9 billion years ago. The collision caused ripples in the lunar crust, resulting in three concentric circular rings, standing a few kilometers high. Molten basaltic lava from the Moon's interior subsequently flowed into the impact site, most likely about 3.2 billion years ago, creating the dark, smooth floor in the center of the basin. Much smaller, sharper, younger craters have impacted this ancient basin in more recent times. Located on the extreme western edge of the Moon, the Orientale basin is difficult to see from the Earth. (Courtesy of NASA.)

Fig. 5.13 Lunar craters Copernicus and Reinhold Bright ejected material radiates outward from the crater Copernicus near the lunar horizon. It is one of the youngest lunar craters on the near side of the Moon, with an estimated age of 800 million years and a diameter of 93 kilometers. The craters in the foreground are Reinhold A and B. (Courtesy of NASA.)

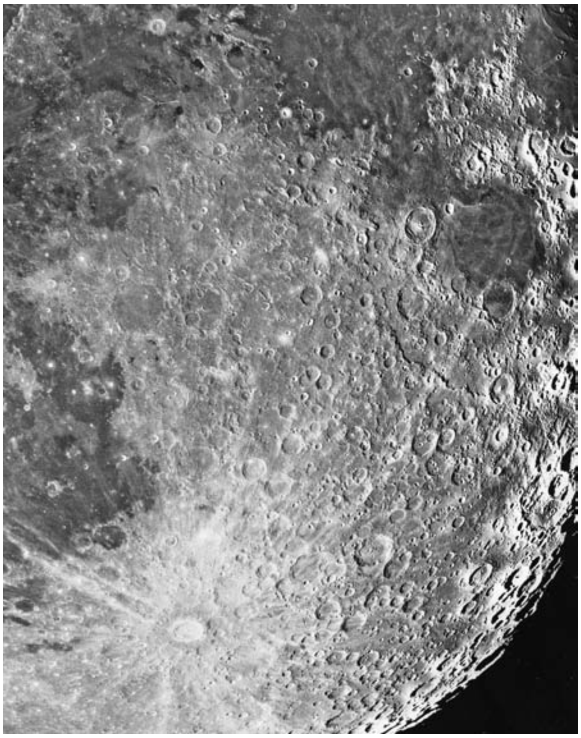


Fig. 5.14 Lunar rays White rays splash out across the Moon from crater Tycho (*bottom center of this image*). Tycho is a large, young crater with a diameter of 85 kilometers and an age of 108 million years. Only relatively recent craters retain their white rays, for those of older craters are darkened and worn away by continued meteorite impact. The dark, flat circular feature in the upper right is Mare Nectaris (Sea of Nectar). This clear image was produced using the unsharp masking technique that permits high contrast and fine resolution. (Anglo-Australian Telescope © 1976. Photo prepared by David F. Malin.)

Focus 5.2 Lunar craters – volcanoes or bombs?

Early interpretations of the lunar craters suggested they were formed by volcanic activity. At the time, volcanic craters were the only Earth craters known, and no impact craters had been recognized on Earth. In addition, for more than two centuries reputable astronomers reported seeing smoke and even fire coming from volcanic eruptions on the Moon.

Gradually, the evidence began to favor the idea that the craters are formed by meteoritic impact. It was found that the floors of most craters are slightly depressed below the surrounding level, in contrast to volcanic caldera that appear at the summits of volcanoes. In addition, large lunar craters contain flat floors and central peaks that are not often found in volcanic craters on the Earth. The central peaks of large craters on the Moon are created by the rebound of the underlying lunar surface following the collision of a big meteorite.

The round shape of nearly all the craters on the Moon can also be explained by the impact hypothesis. Because they strike at great speed, projectiles from space disintegrate in the explosive impact, producing a circular crater regardless of the direction at which the projectile struck. The lunar craters were also seen to resemble impact craters because the amount of material piled in the rims is nearly equal to the material excavated from the interior.

Ralph B. Baldwin (1912–2010) marshaled additional evidence for the explosive origin of lunar craters in his influential book *The Face of the Moon*, first published in 1949. He connected the relationship between the depth and diameter of craters on the Moon to the one describing the shell and bomb craters created during World War II, additionally noting that these man-made explosions have a circular form regardless of the angle of impact. Baldwin also argued that the dark, smooth maria occupy huge basins that were gouged out by rare, powerful impacts that punctured holes in the thin lunar crust, permitting lava to well out into them from the molten lunar interior.

The impact origin for lunar craters was confirmed when rocks were returned from the Moon. Samples from the highland craters and larger basins are conglomerates of pre-existing rocks that have been welded together by impact. Dating of these rocks indicate that the battered highland crust is a museum of impact scars created during an ancient bombardment about 4 billion years ago.

Although the lunar maria were filled during ancient episodes of volcanism, the Moon has apparently been volcanically inactive for 3 billion years. The supposed volcanic outbursts that were reported by several astronomers may have been explosive flashes of light generated during the impact of small meter-sized meteorites on the dark side of the Moon.

5.4 *Apollo* expeditions to the Moon

Race to the Moon

The Soviet Union triggered the Space Age about half a century ago, by launching the first artificial Earth satellite, *Prosteyshiy Sputnik*, the simplest satellite, in 1957, hurling the *Luna 3* spacecraft past the invisible far side of the Moon two years later, and sending the first human, cosmonaut Yuri A. Gagarin (1934–1968), into Earth orbit aboard the *Vostok 1* capsule two years after that.

Soviet officials cited there early accomplishments as evidence that communism is a superior form of social and economic organization. And the United States feared that a missile gap existed between it and its adversary, which seemed to verify the threat that the Soviet Union posed to world peace.

Stimulated by the worldwide excitement generated by the first human flight in space, the visionary President, John Fitzgerald Kennedy (1917–1963), decided that the United States had to defeat the Soviets at their own game, and deliver an American to the Moon.

Thus, on 25 May 1961, just six weeks after the Gagarin flight, Kennedy delivered his now-famous address to a joint session of Congress, including the declaration: "I believe that this nation should commit itself to achieving the goal, before the decade is out, of landing a man on the Moon and returning him safely to Earth". The president's call to action struck a responsive chord in the American public and was galvanized under the newly created National Aeronautics and Space Administration (NASA). In less than nine months, on 20 February 1962, John H. Glenn Jr. (1921–) became the first American to orbit the Earth, and the race to the Moon was in full tilt.

For several years the two superpowers traded accolades. The Russians sent the first woman – Valentina Tereshkova (1937–) – into space, and they were the first to orbit three men in the same spacecraft. On 18 March 1965 the Russian cosmonaut Aleksei A. Leonov (1934–) was the first to walk in space, from the *Vostkod 2* capsule, closely followed by the American astronaut Edward H. White (1930–1967) who took the first United States spacewalk on 3 June 1965 from the *Gemini 4* spacecraft.

During 1965 and 1966 the United States launched 10 successful flights of the two-man *Gemini* spacecraft, including the first rendezvous of two spacecraft, *Gemini 6* and *7*, and it was well prepared to embark on the *Apollo* program to land men on the Moon. It began with an ill-fated flight simulation on 27 January 1967, when faulty wiring ignited a flash electrical fire that asphyxiated and incinerated three astronauts on the ground. Yet, in just 22 months after this tragic setback, the manned *Apollo 8* spacecraft entered lunar orbit, all but ending the race to the Moon. It opened the way for the historic *Apollo 11* mission seven months later, when Neil Armstrong (1930–) planted the American flag on lunar soil. The achievement was a spectacular triumph, the ultimate space first in the global geopolitical competition with the Soviet Union.

From 1960 to 1970 the Soviet Union sent several unmanned spacecraft to the Moon, including roving vehicles and sample returns to Earth, but they had lost the race to put a man on the Moon. Personal rivalries, shifting political alliances and bureaucratic inefficiencies had apparently bred failures and delays. After the triumphs of *Apollo 8* and *11*, the Russian lunar program faded into oblivion, and they turned their attention to long-duration missions in Earth-orbiting space stations.

But to be honest, the driving factor in the race to the Moon was not scientific. It was the Cold War rivalry between the United States and the Soviet Union. And the American achievement was a spectacular political triumph. The dissipation of the Soviet Union's lead in space tarnished the image of Soviet competence and diminished their status in world affairs. In contrast, landing men on the Moon and conquering the frontier of space taught the American people that nothing is impossible if they set their sights high enough; with resolve and willpower you can accomplish anything, especially in a democratic nation that stresses individual freedom. And it was surely a contributing factor to the idea that success can result from technological superiority.

The *Apollo* program to land men on the Moon

Before the United States accomplished manned landings, three types of robot spacecraft were sent to reconnoiter and answer two main questions for the proposed lunar landing. The first concerned the danger of encountering rocky terrain, where it would be impossible to land without capsizing. The second was the prediction, by some astronomers, that a thick layer of dust covers the lunar surface, perhaps as deep as a kilometer, which would make travel impossible. In fact, the astronauts might sink into the dust, suffocate and vanish into the Moon, like sinking into quicksand on Earth. After all, the lunar surface has been battered, churned and worn down by a hail of meteorites over the eons, creating loose debris of rocks, pebbles, grains, soil and dust.

To start resolving these uncertainties, three *Ranger* spacecraft crashed into the Moon, transmitting television pictures back to Earth as they rapidly approached the lunar surface. Watching these pictures was a dizzying experience, and the transmission of the final frames was interrupted by the crash itself. These were followed by five *Lunar Orbiters* that mapped most of the Moon's surface to locate potential landing sites, missing only the polar regions. The final stage of preparation involved soft landings by the *Surveyors 1*, *3*, *5* and *7* that tested the detailed physical and chemical properties of the lunar surface and certified the safety of the initial *Apollo* landing sites. While the ground-control crews watched anxiously, the feet of the three-legged *Surveyor* robots sank only a few centimeters into the lunar soil, showing that there was no thick dust layer and people could walk on the Moon without sinking in over their heads.

The *Apollo* spacecraft was designed to carry three men into orbit around the Moon. A small, Spartan landing craft, the *Lunar Excursion Module* (*LEM*), would ferry two of the crewmen from lunar obit to the Moon's surface and then back to the mother ship, while the third astronaut remained orbiting the Moon in the larger *Command and Service Module* (*CSM*).

On 21 December 1968 three *Apollo 8* astronauts became the first humans to break free of the Earth's gravity. Although the crew would only orbit the Moon and not land on it, the unprecedented voyage provided the first sight of the Earth seen from afar – a radiant blue-and-white sphere rising beyond the battered face of the Moon in the dark void of space (Fig. 5.15). We then saw our home world in a new perspective, beautiful and vulnerable, a tiny, fragile oasis shimmering all alone in the vast deep chill of outer space. The sheer isolation of the Earth became plain to every person on the planet. It stimulated a worldwide awareness of the Earth as a unique and vulnerable place, fostering the ecology movement and helping us to get a better feeling for the planet's place in our lives and the Universe.

On 20 July 1969, the spindly-legged *Lunar Module Eagle* carried two *Apollo 11* astronauts to the lunar surface. While an estimated half-billion people watched, Neil Armstrong took the controls to avoid a hazardous crater, and radioed the first words from another world: "Houston, Tranquility Base here. The *Eagle* has landed."

With Buzz Aldrin (1930–) at his heels, Armstrong groped cautiously down the ladder to the surface. He stood firmly on the fine-grained surface, and an ancient dream

Fig. 5.15 Earthrise In 1968, the *Apollo 8* spacecraft carried the first humans on a journey around the Earth's Moon. When they reached the far side of the Moon, the crew looked back toward the Earth along the lunar horizon, watching our planet rise as the spacecraft continued its orbit around the Moon. They helped create a new image of the Earth as a blue and turquoise ball suspended alone in dark space, light and round and shimmering like a bubble, flecked with delicate white clouds. (Courtesy of NASA.)

had come true — man had set foot on another world and humans were no longer confined to their native planet.

As Armstrong put it: "That's one small step for man, one giant leap for mankind." Although he forgot the "a" in front of "man", everyone knew what Armstrong meant. Moments after his initial footstep, Aldrin gazed out at the Sea of Tranquility and said simply "magnificent desolation". The next day, the Italian newspapers put it more succinctly: "Fantastico!"

After the historic landing, it was time to return to Earth. The two astronauts flew the ascent stage of *Eagle* back to the Moon-orbiting *CSM*, where Michael Collins (1930–) was waiting to take them home (Fig. 5.16).

By the time of the *Apollo 17* mission in 1972, the lunar landings had become so commonplace that astronaut Eugene A. Cernan (1934–) muttered "Let's get this mother out of here" as he blasted off the Moon in his *Lunar Module Challenger*. That's more like Sir Edmund Hillary's (1919–2008) comment when descending from the summit of Mount Everest on 29 May 1953 – "We knocked the bastard off."

Still, the first human visit to the Moon was a historic occasion. Even now, there is a sense of participation, a feeling that our lives were enriched and made memorable by the landing.

In all, twelve humans have walked on the lunar surface, to gather samples, take photographs and make other scientific measurements (Table 5.3). All of the landing sites were on the near side and close to the lunar equator because these were the only places the astronauts could go safely (Fig. 5.17). Direct radio contact with Earth would be lost if they landed on the far side of the Moon. Sites near the equator were chosen to always be able to get astronauts back from the lunar surface quickly in case something bad happened on the Moon. A landing near the edge or limb of the Moon, as viewed from Earth, was ruled out if the spacecraft was to return to Earth in daylight. Within these constraints, the landing sites were chosen to provide samples

Table 5.3 *Apollo* missions to the Moon

Mission	Launch date[a]	Landing site	Accomplishments	Sample (kilograms)
Apollo 8	21 Dec. 1968	Lunar Orbiter	First humans to orbit Moon	
Apollo 10	18 May 1969	Lunar Orbiter	Test Lunar Excursion Module	
Apollo 11	16 July 1969	Mare Tranquillitatis	First human landing	21.6
Apollo 12	14 Nov. 1969	Oceanus Procellarum	First ALSEP[b]	34.3
Apollo 13	11 Apr. 1970	Flyby	Landing aborted	–
Apollo 14	31 Jan. 1971	Fra Mauro, highland	First highland landing	42.6
Apollo 15	26 July 1971	Hadley-Apennine	First lunar rover	77.3
Apollo 16	16 Apr. 1972	Descartes	Highland landing	95.7
Apollo 17	07 Dec. 1972	Taurus-Littrow	Last flight	100.5

[a] The spacecraft landed on the Moon four or five days after launch.
[b] ALSEP is an acronym for Apollo Lunar Surface Experiments Package.

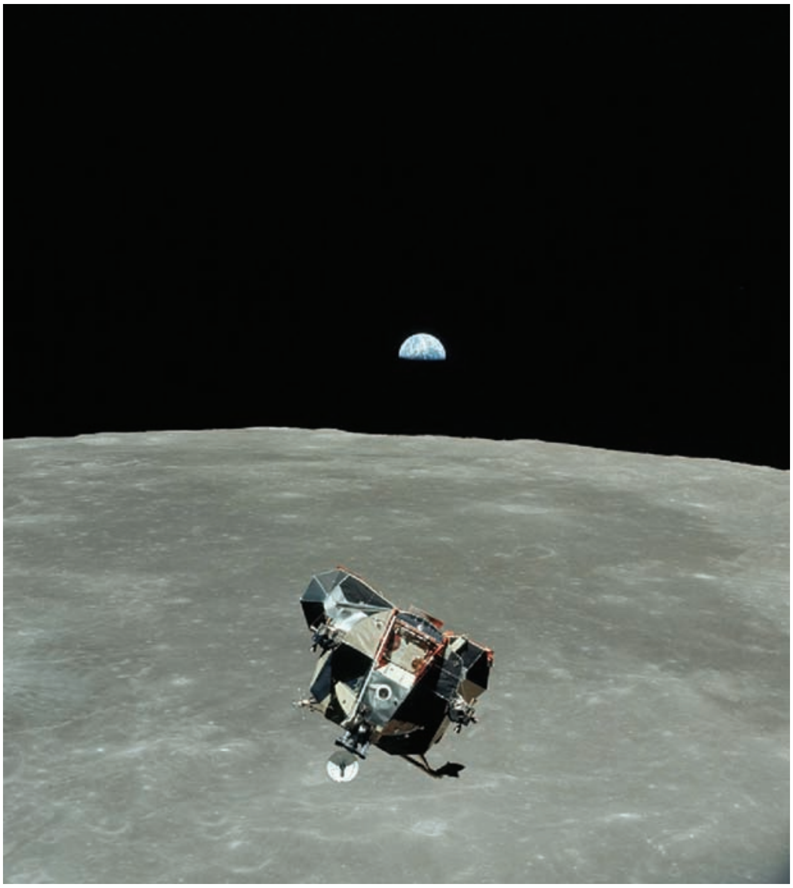


Fig. 5.16 Going home from the Moon The ascending *Lunar Module* of the *Apollo 11* mission, carrying Neil Armstrong (1930–) and Buzz Aldrin (1930–) back from the first human landing on the Moon. They were returning to the Moon-orbiting *Command and Service Module*, when its pilot Michael Collins (1930–) took this photograph, on 21 July 1969. (Courtesy of NASA.)

of a wide variety of terrain, from the smooth maria to the heavily cratered highlands.

Apollo 11 landed on the smooth plains of Mare Tranquillitatis, and *Apollo 12* settled down on a mare site near the edge of the vast Oceanus Procellarum. Rocks returned from these first missions confirmed the volcanic basalt nature of the maria and established their antiquity, with ages greater than 3 billion years. The *Apollo 14* landing site was located in highland terrain near the crater Fra Mauro, an area thought to be covered with debris thrown out by the impact that formed the Imbrium Basin. Dating of material obtained from this site indicated that the basin-forming impact occurred 3.85 billion years ago. *Apollo 15* was the first mission to employ a roving vehicle; it was sent to the Hadley-Apennine region containing both mare and highland units. The so-called Genesis

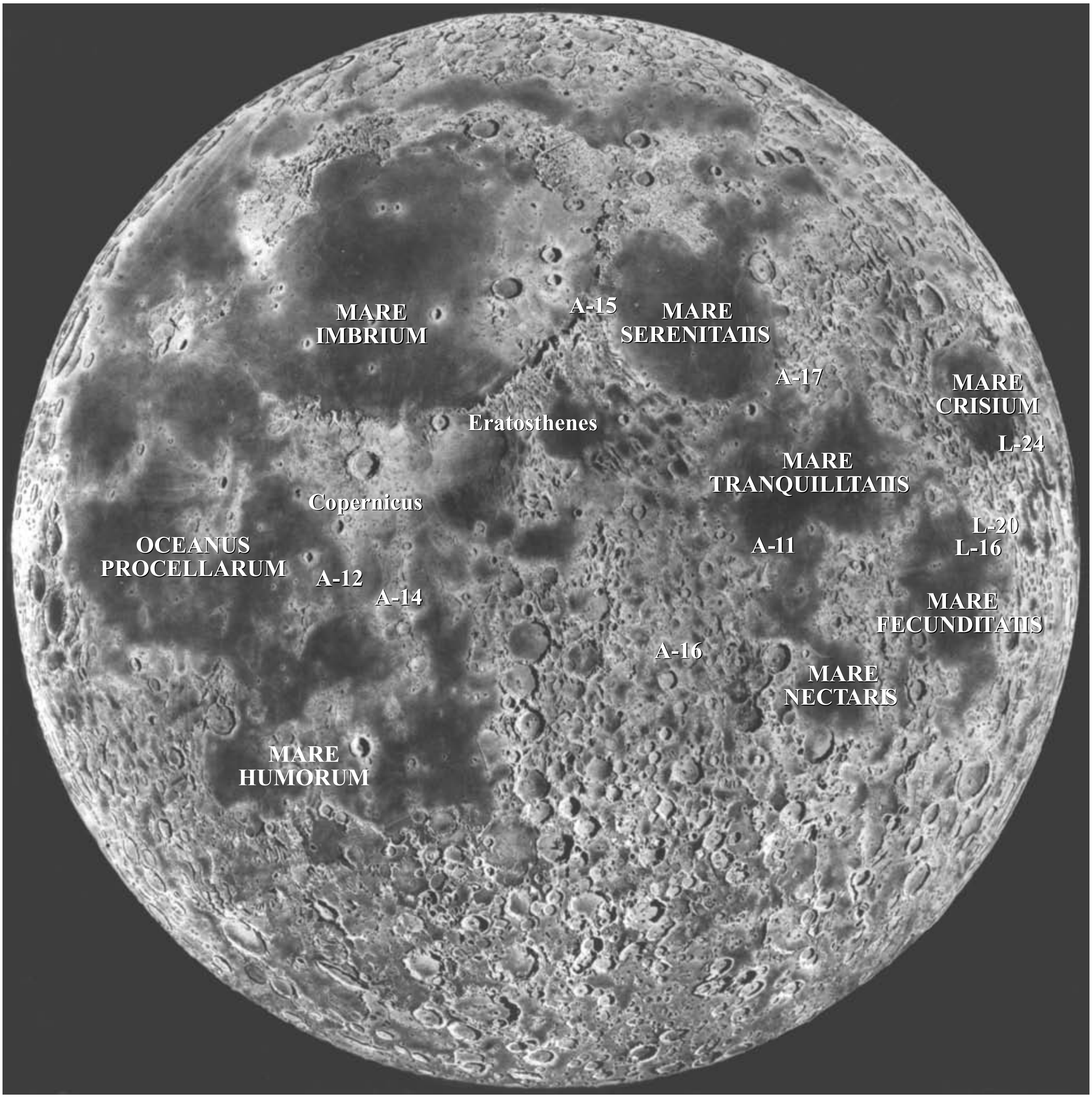


Fig. 5.17 Apollo landing sites on the Moon The six *Apollo* (A) landing sites were located in safe places near the equator on the near side of the Moon. Within this constraint, the sites were designed to obtain samples from a wide variety of terrain. *Apollo 11* and *12* respectively landed on Mare Tranquillitatis and Oceanus Procellarum. The spot chosen for *Apollo 14* was the Fra Mauro Formation, which is covered with material ejected during the ancient impact that created the Imbrium Basin. By landing at a point just inside the Apennine Mountains, the *Apollo 15* astronauts could sample highlands, maria and the Hadley Rille. The *Apollo 16* mission sampled the highlands near crater Descartes, while *Apollo 17* landed near Mare Serenitatis. The location of the three Soviet *Luna* (L) unmanned sample-return sites are also shown.

rock found during this mission is a primitive chunk of highland anorthosite dating back some 4.5 billion years. *Apollo 16* landed on the Descartes highlands near the rim of the Nectaris basin, blasted out 3.92 billion years ago, and *Apollo 17* was sent to the Taurus-Littrow site, at the edge of the Serenitatis basin, excavated 3.87 billion years ago.

What happened to *Apollo 13*? Its three astronauts almost lost their lives when an oxygen tank exploded aboard the spacecraft on the way to the Moon. Sealed

Focus 5.3 Black skies on the Moon

The lunar astronauts stepped into a stark but beautiful world. With no clouds, dust, moisture or haze to obscure the view, distant details stood out clearly against the deep black background. Because the Moon has no atmosphere to speak of, the sky was pitch black in broad daylight, there were no sounds to disturb the eternal stillness, and the Sun's true light could be seen, unfiltered by any air. And since there is no air to breathe on the Moon, visiting astronauts were bundled in oxygen cocoons known as spacesuits.

By way of comparison, incident sunlight contains all the colors, but the Earth's air molecules scatter blue light more strongly than red light, making the overhead sky appear blue. The Sun's rays being bent by the atmosphere cause the twilight zone between the night's darkness and sunrise or sunset on Earth. And when the Sun rises or sets, most of the blue light is scattered out before reaching us, so the light of the setting Sun is reddened; airborne dust helps this effect. In contrast, astronauts orbiting the Moon saw no twilight, and no colorful sunrise or sunset.

How do we know that the Moon has no appreciable atmosphere? The earliest evidence came from the abrupt vanishing of stars behind the edge of the Moon during lunar occultations. The word *occult* means "to hide". If the Moon had a thick atmosphere, the starlight would gradually dim during a lunar occultation, but this vanishing actually takes less than one second. It is in sharp contrast to the gradual fading of the Sun, stars, and planets when they set behind the horizon on Earth. The lack of a significant atmosphere also follows from the Moon's relatively small mass and gravity, together with its proximity to the warming Sun.

Instruments carried to the Moon by the *Apollo* astronauts identified the barest wisp of helium and argon atoms, and subsequent imaging observations from Earth revealed just a whisper of sodium and potassium atoms, emitting a detectable fluorescent glow when exposed to sunlight. But this is not a permanent atmosphere in the sense that ours is, for the lunar "exosphere" is continuously being created, lost and replaced every few hours or weeks, depending on the atom.

The Moon's atmosphere is 100 trillion, or 10^{14}, times more tenuous than the Earth's air, and so thin and rarefied that its constituent particles hardly ever hit each other. So the Moon has no atmosphere in any practical sense.

inside, the crew was in danger of dying by re-breathing their carbon dioxide. They survived by converting the tiny *Lunar Module*, with its intact, breathable air and fuel, into a lifeboat, canceling the planned lunar landing and heading home. *Apollo 13* was very nearly a catastrophe, which has been shown in the captivating 1995 Oscar-winning movie *Apollo 13*. If the tank had exploded earlier, there would not have been enough electric power and water to go around the Moon and get home again. And if it had occurred later, when the astronauts were on their way down to land on the Moon, there would not have been enough fuel left in the *Lunar Module* to go home.

The other lunar astronauts recorded an eerie wasteland below a blackened sky (Focus 5.3), battered and scarred with craters of all sizes and covered with dust. It clung to the astronauts' clothing and equipment and showed the sharp outline of their footprints (Fig. 5.18); but there were no clouds of dust above the airless surface. Walking on the lunar surface was like walking on plowed soil or wet sand, and most of the finer dust had evidently been plowed down into the Moon by the churning of the meteorites.

Armstrong and Aldrin never strayed more than a hundred meters from their lander, like timid children testing the water when entering a lake or sea for the first time. The astronauts of the next two missions (*Apollo 12* and

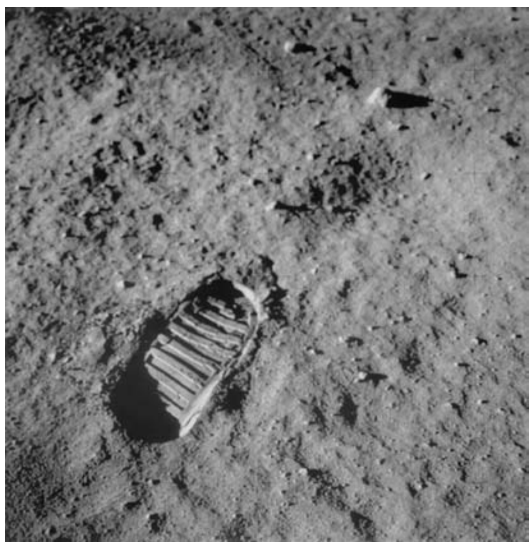


Fig. 5.18 Boot prints on the Moon On 20 July 1969, Neil Armstrong (1930–) became the first human to walk on the Moon. His boot print, shown here, reveals a thin layer of Moon dust, about 0.01 meters thick. Because there is no atmosphere or weather on the Moon, the footprint will probably remain for 1 or 2 million years. By that time, the constant rain of micrometeorites will have erased it. Altogether, twelve astronauts have left boot prints on the Moon. (Courtesy of NASA.)

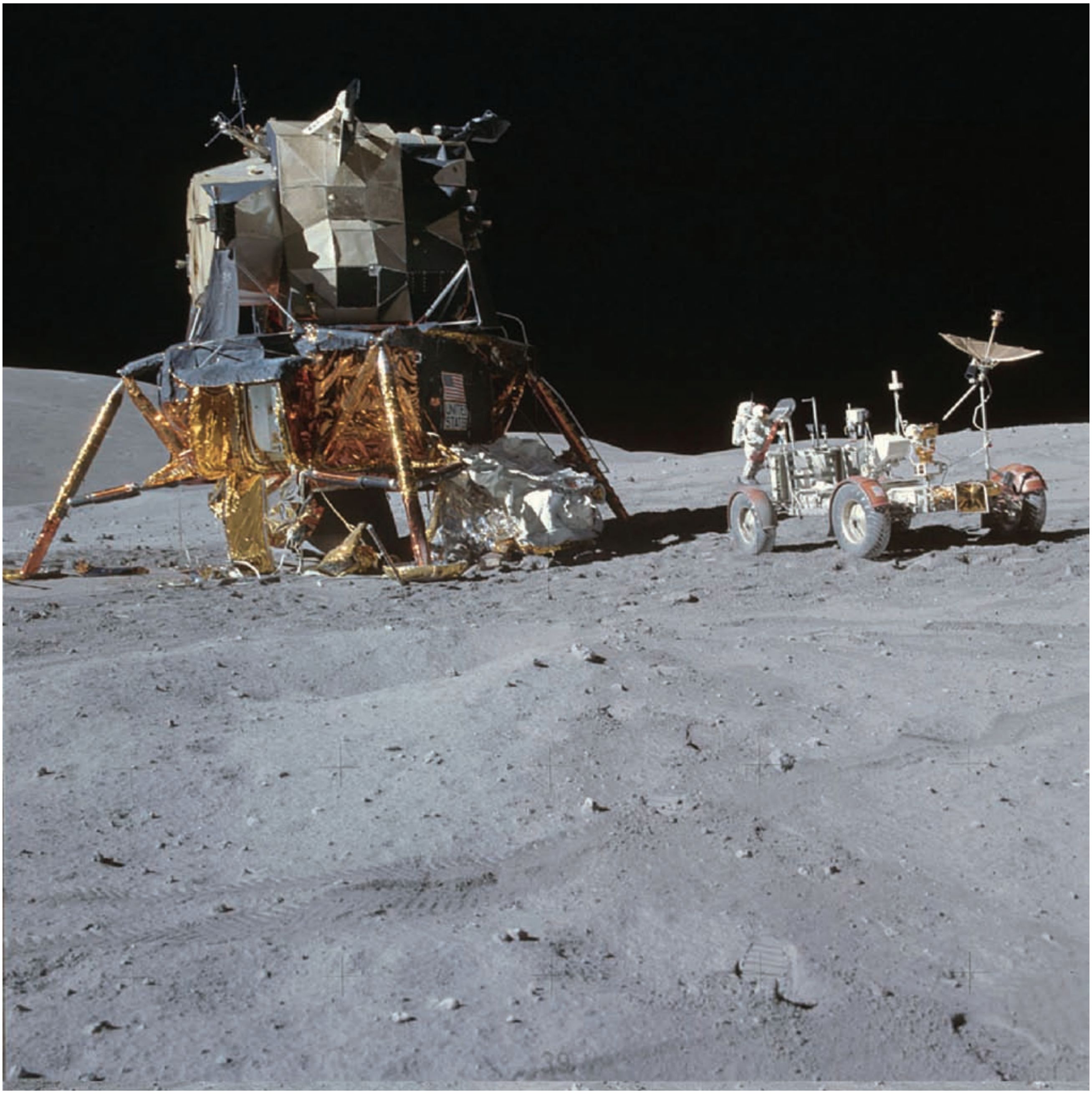


Fig. 5.19 Driving on the Moon *Lunar Rovers* were used to travel across the Moon's rugged terrain, gathering rocks from a wide variety of locations. In this image, *Apollo 15* astronaut James Irwin (1930–1991) prepares to take a *Lunar Rover* for a drive. The *Lunar Module* "Falcon" is at the left side of this image. The St. George Crater is located about 5 kilometers behind Irwin, and the lunar mountains Hadley Delta and Appennine Front are in the background at the left. The *Rovers* were left on the Moon. Free from wind, rain and rust, they will remain intact for millions of years; one might even imagine a returning astronaut using one that was discarded hundreds of years before. (Courtesy of NASA.)

14) had greater confidence and took longer moonwalks. During the last three missions (*Apollo 15, 16* and *17*) astronauts roamed as far as 7 kilometers from the landing site, visiting some of the most spectacular places on the Moon in a battery-powered car called the *Lunar Rover* (Figs. 5.19, 5.20). Still, the visits were always short, with the entire missions lasting between one and two weeks.

Unlike the early landings on the smooth lunar maria, the last three *Apollo* flights visited mountainous areas: the Appenines, the Descartes highlands and the Taurus Mountains. The tops of all the mountains were rounded off into gentle hills without sharp peaks or steep cliffs. Although it looked as if the Moon had been sandblasted smooth by eons of meteorite bombardment, the main reason for the

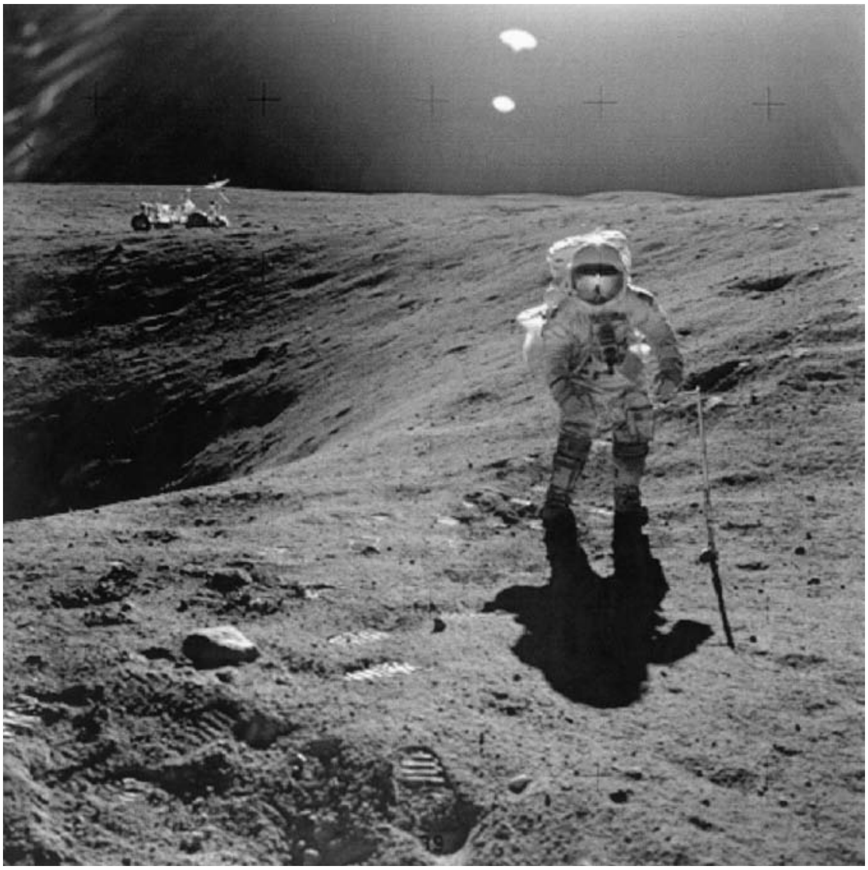


Fig. 5.20 Moonwalk Charles Duke (1935–) strolls across the lunar surface during the *Apollo 16* mission in April 1972. Small impacting particles have sandblasted the lunar surface, producing smoothed, undulating layers of fine dust and rounding the surfaces of lunar rocks. Larger meteorites have pounded and churned the surface, producing a layer of ground-up rocky debris. (Courtesy of NASA.)

gradual lunar slopes is that there is no water or ice erosion, as on Earth, to cut deep valleys and shape mountain crags, or tectonic activity to toss up crumpled mountain ranges.

The astronauts left behind the Apollo Lunar Surface Experiments Package (ALSEP). This nuclear-powered array of instruments included seismometers to monitor vibrations of moonquakes and meteorite impact, magnetometers to measure possible magnetic fields, and other instruments to analyze gases and charged particles streaming from the Sun to the Moon. The astronauts also brought lunar soil and rocks back home with them, altogether 382 kilograms and not an ounce of cheese (Fig. 5.21).

Mirrors were also left on the Moon, to reflect laser light fired from Earth. Every reflector contained 100 small mirrors, each in the shape of the three-sided corner of a box. These corner cubes reflect light directly back toward its point of origin. Observations of pulsed laser light, sent to the lunar mirrors and back, has permitted astronomers to measure the Moon's distance with an accuracy of two centimeters, or to better than one part in 10 billion, showing that the Moon is moving very slowly away from the Earth.

Sophisticated experiments were also performed from the *Command Module*, mapping the magnetic fields, chemical composition, surface radioactivity and terrain from a distance as the mother ship circled the Moon.

Returning to the orbiting craft, the astronauts jettisoned the landing *Lunar Module* and headed for Earth, arriving home about three days later. Biologists felt there was a chance that the astronauts, or the returned rock samples, might infect the human race with some deadly lunar virus. The astronauts from the first three lunar landing missions, *Apollo 11*, *12* and *14*, were therefore placed in quarantine for three weeks after their return. They remained in fine health, and the crews of the last three missions, *Apollo 15*, *16* and *17*, did not have to suffer through the quarantine.

The achievement of landing humans on the Moon was a spectacular American triumph in the cold-war confrontation with the Soviet Union, in an incredible, warlike mobilization of scientists, engineers, and technology with an optimistic, can-do spirit. Imaginative thinkers viewed it as the stepping stone to permanent lunar bases, giant space stations and the colonization of Mars. But the dreams quickly dissipated and the sense of mission disappeared. The "age of *Apollo*" was short-lived, and public interest quickly waned.

The goal had been reached, the crisis was over, and the enemy had been conquered. The *Apollo* program lost its political appeal in the face of growing public indifference, its budget shriveled, and the program was abruptly cut

Fig. 5.21 Moon rock Harrison Schmitt (1935–) about to walk behind Split Rock during the *Apollo 17* mission in December 1972. Eugene Cernan (1934–) had already scooped up samples from the debris on the front side of the boulder. The huge rock rolled down about a billion years ago, splitting into five pieces during the fall. The total length of the boulder, when reassembled, is about 20 meters. *Apollo 17* was the last of the six missions that landed humans on the Earth's Moon and returned them safely. It investigated the dark terrain at the Taurus-Littrow landing site, deployed explosives to be used with seismographs to examine the lunar interior, and returned rocks to the Earth. (Courtesy of NASA.)

short with the cancellation of the *Apollo 18*, *19* and *20* lunar missions, largely for political reasons.

The excitement must have also quickly dissipated for the men who visited the Moon. They had trained their whole lives to stay for just a few days on the Moon, waiting for years for that opportunity. A few moments of fame followed, but there was no encore. There wasn't anything comparable to do, and many of the astronauts had some hard adjusting to do after their return from the Moon.

5.5 Inside the Moon

Moonquakes

As the Earth has earthquakes, so the Moon has moonquakes which were first detected by the sensitive seismometers placed by the *Apollo* astronauts at four widely spaced locations on the lunar surface — at the *Apollo 12*, *14*, *15* and *16* landing sites. Because interfering winds, sea waves, and road traffic do not shake the Moon, the lunar seismometers can detect moonquakes that are relatively weak by terrestrial standards.

More than 12 500 seismic events were recorded over the eight years the seismometers were used. As expected, they were able to record occasional tremors of the Moon caused by the impact of small meteorites, as well as a shuddering from the deliberate crash landing of *Lunar Modules* near the end of some *Apollo* missions. But a great many more events were generated inside the Moon.

A few of these moonquakes, 28 in all, were shallow, emanating from the upper mantle, the largest about 5.0 on the Richter scale. But more unexpected were the numerous tiny moonquakes, several a day on average, which occurred further down, about halfway to the center of the Moon and deeper than any earthquake.

These deep moonquakes are much smaller than even mild earthquakes. If you stood directly over most moonquakes you would not even feel your feet shake. They almost never exceed a magnitude of 2 on the Richter scale. Although earthquakes of this magnitude are recorded on Earth, they are not felt by humans and produce no damage to buildings.

The moonquakes are not only gentler than earthquakes; they also have distinctly different behavior. While tremors on the Earth start suddenly and persist for only a few minutes, the moonquake waves build up gradually and continue for more than an hour, suggesting that the body of the Moon is an almost perfect medium for the propagation of seismic waves.

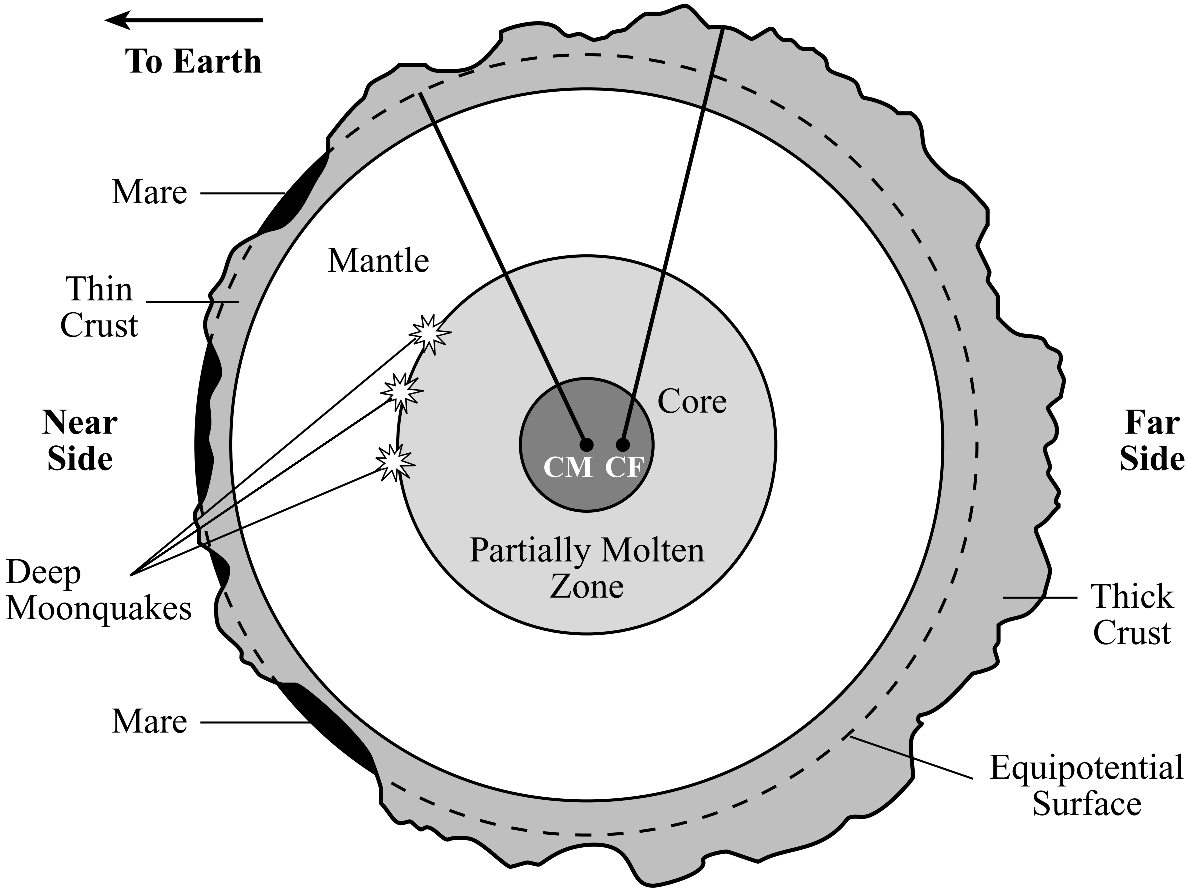


Fig. 5.22 Lunar interior A schematic cross-section of the Moon shows its internal structure. The lunar crust is thinner on the near side that faces the Earth, and thicker on the far side. Fractures in the thin crust have allowed magma to reach the surface on the near side, where the lava-filled maria are concentrated. The Moon has an iron-rich core with a radius of about 20 percent of the Moon's average radius of 1738 kilometers. A partially molten layer is believed to encircle the Moon's core, out to depths of about 1000 kilometers. The Moon's center of mass (CM) is offset by 2 kilometers from its center of figure (CF), so an equipotential surface, which experiences an equal gravitation force at all points, lies closer to the lunar surface on the hemisphere facing Earth. Therefore magmas originating at equipotential depths will have greater difficulty reaching the surface on the far side.

Moreover, nearly identical seismograph records were obtained again and again, indicating that certain regions, known as *nests*, are repeatedly generating moonquakes in the same way. The largest nest emitted 323 moonquakes during the period of observation.

The rate of the moonquakes, in general, seemed to rise and fall every 27 days, the time it takes the Moon to circle the Earth, suggesting that they were caused by the tidal pull of the Earth.

The seismograph records have permitted the construction of a model for the lunar interior, in much the way that geologists have modeled the inside of the Earth. The Moon is slightly asymmetrical in bulk form, with a thicker crust on the far side. Most of the volcanic maria occur on the near side where the crust is thinner (Fig. 5.22). On average, the crust on the near side of the Moon is about 50 kilometers thick, or less. It is only a few tens of kilometers thick beneath the mare basins. In contrast, the far side crust is believed to be about 15 kilometers thicker than the near side crust. As a result, the Moon's center of mass is offset from its geometric center by about 2 kilometers in the direction of the Earth.

A small lunar core

Scientists from the *Apollo* era were unable to agree whether the Moon has an iron-rich core; but they were certain that it had to be much smaller than the core of the Earth, which is 55 percent of the radius of the planet and 32 percent of the planet's mass. Data from the *Lunar Prospector* spacecraft, obtained in 1998–99, have been used to gauge the size of the lunar core. They indicate that the Moon's core has a radius of about 350 kilometers, or 20 percent of the satellite's radius, and only about 2 percent of the body's mass.

Radio telescopes on Earth were used to measure small Doppler-effect changes in the *Lunar Prospector*'s radio signal as the spacecraft moved toward or away from the Earth, thereby identifying slight variations in the craft's velocity as it orbited the Moon. Since these velocity changes are caused by the varying gravitational pull of the Moon, they could be used to construct a full gravity map of the near and far sides of the Moon, from pole to pole. The resulting map revealed the distribution of mass within the Moon, and showed that it has a small, dense, metallic core, about 350 kilometers in radius if mostly iron.

A second method studied the weak magnetic field induced within the Moon when it passes through the tail of the Earth's magnetosphere each month. This technique confirmed the presence of a lunar core of about the same size as that inferred from the gravity data.

The relatively small core of the Moon has profound implications for its origin. If the Moon and Earth coalesced independently, their cores might be expected to occupy a similar fraction of their volume; instead the Moon seems to have coalesced from material that has been blasted out of the young Earth. The giant impact might have taken place when the Earth was still forming. In that case, most of the Earth's iron might have already sunk to its core, but there could be enough iron-rich rock, expelled into space from the Earth and impacting object, to build a lunar core.

Internal, partially molten zone in the Moon

The *Apollo* seismic data suggested that the outer half of the Moon is cold and solid, but that it might be warm and partially molten in a lower zone. The moonquake waves lost energy if they went deeper than 1000 kilometers, about halfway to the center of the Moon at 1738 kilometers down. The *Apollo* scientists argued that the deep moonquakes might be generated at the boundary between the outer solid shell and the inner molten zone.

Evidence for a partially molten zone has been obtained from accurate measurements of the Moon's distance using laser ranging. The laser beams are sent to the Moon from telescopes on Earth and reflected from corner mirrors left on the Moon by the *Apollo* astronauts. Measurements of the round-trip travel time of a pulse of laser light yields twice the distance between the Earth and the Moon, by multiplying the time by the velocity of light, with an accuracy of 0.02 meters. After more than 30 years of such determinations, scientists have concluded that the Moon's surface moves in and out by as much as a tenth of a meter, or 10 centimeters, every 27 days, in response to the shifting gravitational tugs of the Earth. This elastic yielding suggests that the interior is pliable, with a larger, partially molten layer surrounding the core.

Mascons on the Moon

Precise radio tracking of *Apollo* spacecraft on the near side of the Moon showed that their orbits are gravitationally deflected toward the circular maria. The spacecraft acted as though the maria contained mass concentrations, abbreviated as *mascons*, which pulled at the spacecraft and changed their velocity when passing overhead. Virtually all the maria on the near side showed this unexpected feature, and the excess mass in each is about 10^{18} kilograms, or 1/70 000 the total mass of the Moon. Because radio tracking of the orbiting spacecraft was not possible when they passed to the far side, it was not known from the *Apollo* missions whether there are mascons on the far side of the Moon.

The *Lunar Prospector* spacecraft has been used to detect mascons from gravity data for the entire Moon, discovering several mass concentrations beneath the floors of large impact basins, including at least four on the far side. The nearside basins have been filled with mare lava, but the farside basins remain unfilled, so it isn't the lava that is providing the extra gravitational pull.

What are the mascons? The most likely explanation is that they represent an upward bulging of high-density mantle rocks that rose in the aftermath of basin-forming impacts. The impact that formed the largest basins has weakened the crust so much that the dense mantle has moved up beneath them, raising and fracturing the basin floors.

5.6 The lunar surface

Rocks from the *Apollo* missions

During the *Apollo* landings from 1969 to 1972 a dozen people roamed the Moon taking hundreds of rock samples, placing them in labeled bags, and returning 382 kilograms of Moon rocks to Earth in sealed containers. These specimens from another world have permitted scientists to decipher the composition of the lunar crust, and to reconstruct our satellite's history.

Since contact with the Earth's atmosphere would alter the composition of the lunar samples, they are kept in cabinets filled with a dry, oxygen-free atmosphere of nitrogen and are manipulated with long gloves sealed to the walls of the cabinets. When not under investigation, the rocks are kept in a massive steel-lined vault at the Lunar Receiving Laboratory of NASA's Johnson Space Center at Houston, Texas.

Scientists spent years examining the rocks brought back from the Moon in search of water, but none was found. For more than 40 years the Moon was therefore thought to be bone dry. As discussed in Section 2.4, orbiting lunar spacecraft nevertheless detected signs of surface water on the Moon in the late 1990s and early 2000s, and this probably led to a re-examination of rocks collected during the *Apollo* missions. An analysis of the lunar samples in 2008 showed evidence of very small amounts of water, at about 50 parts per million, in volcanic glasses formed on the Moon about 4 billion years ago. The chemical signatures of water on the Moon were confirmed in 2010, in the form of hydroxyl ions, each containing one atom of oxygen and one atom of hydrogen. The hydroxyl was locked up in mineral crystals as water-bearing magma cooled on the Moon's surface soon after it formed.

Although scientists have now found 100 times more water in the Moon's minerals than previous limits, the concentrations are very low and until recently impossible to detect. Moreover, the Moon was never soaking wet, and both its interior and surface are still considered drier than the driest desert on Earth.

The Moon is also lifeless. Extensive testing revealed no evidence for life, past or present, among the lunar samples. They contain no living organisms, fossils or native organic compounds. Thus the Moon is a desolate place, barren of life.

From its low mean mass density, we would expect the bulk of the Moon to be composed of silicates, or minerals in which atoms of silicon and oxygen are linked to other elements. Laboratory investigations of the lunar samples indicate that the lunar crust is indeed composed of such minerals, just as the Earth's crust is.

The surface of the Moon has been bombarded by meteorites for billions of years, breaking the lunar crust into rock fragments and fine-grained material known as the lunar *regolith*. It is the loose debris that has fallen back to the Moon after eons of meteorite bombardments. The regolith is the Moon's version of soil, which has organic connotations here on Earth, but the Moon's soil contains no organic material.

The regolith covers the entire lunar surface to depths as great as 20 meters. It is thickest in the highland regions that have been exposed to meteoritic bombardment longest. The regolith in the maria is 2 to 8 meters deep.

The Moon rocks are roughly divisible into three types: anorthosites, basalts and breccias, and they all exhibit important differences from terrestrial rocks. The lunar samples are all much older than most rocks found on Earth, and they are composed of material that has been previously melted (anorthosites), erupted through magma–lava outflow (basalts), and crushed by meteorite impacts (breccias).

The *anorthosites* are the oldest rocks ever found, dating back to 4.5 billion years ago. They are found in the light-colored lunar highlands, and contain a type of mineral known as plagioclase feldspar, commonly found on Earth, but with a difference. The lunar anorthosites were melted more than 4 billion years ago.

Basalts are dark lava that fills mare basins, forming a secondary crust. These thin volcanic veneers were created after heat from radioactive decay accumulated in the Moon, leading to the rise of magma and the eruption of basaltic lava about 3 billion years ago. The surface of Venus and the Earth's ocean floor are also secondary crusts formed in this way. But most of Venus was resurfaced about 750 million years ago, and the Earth's ocean floor is still being created.

The Moon contains none of the type of rocks that were generated deep inside the Earth during plate tectonic processes, such as the continental granites.

After the lunar rocks solidified, they were broken up, flung about and pulverized by meteorite impacts. Energetic impacts, powerful enough to excavate meter-sized craters, have compacted and welded the regolith into aggregates called *breccias*. They retain compositional information from the era in which they formed.

Scanning the surface of the Moon from spacecraft in the 1990s

The *Apollo* rock and soil samples came from only six sites on the near side, chosen mainly to be safe and easy to get to. A global perspective of the Moon's surface composition therefore had to wait until the *Clementine* spacecraft surveyed the unexplored regions on both the near and far sides in the 1990s.

After the *Apollo* missions, no one had even a single glimpse of the Moon's far side for nearly two decades, and then it was obtained by the *Galileo* spacecraft on its way to explore Jupiter's realm. In order to reach the giant planet, *Galileo* gained speed by twice swinging past the Earth, passing by the Moon in the process. In 1990 and 1992 the instruments on *Galileo* obtained images of the lunar limb and far side from vantage points not previously obtained.

Composites of *Galileo* images taken in three colors – violet, red and near infrared – have been used to depict compositional variations of the lunar surface (Fig. 5.23). They have been calibrated by *Apollo* sample returns that specify the chemistry at specific sites on the near side of the Moon. Some mare basalts are rich in titanium, while many others are relatively low in titanium but rich in iron and

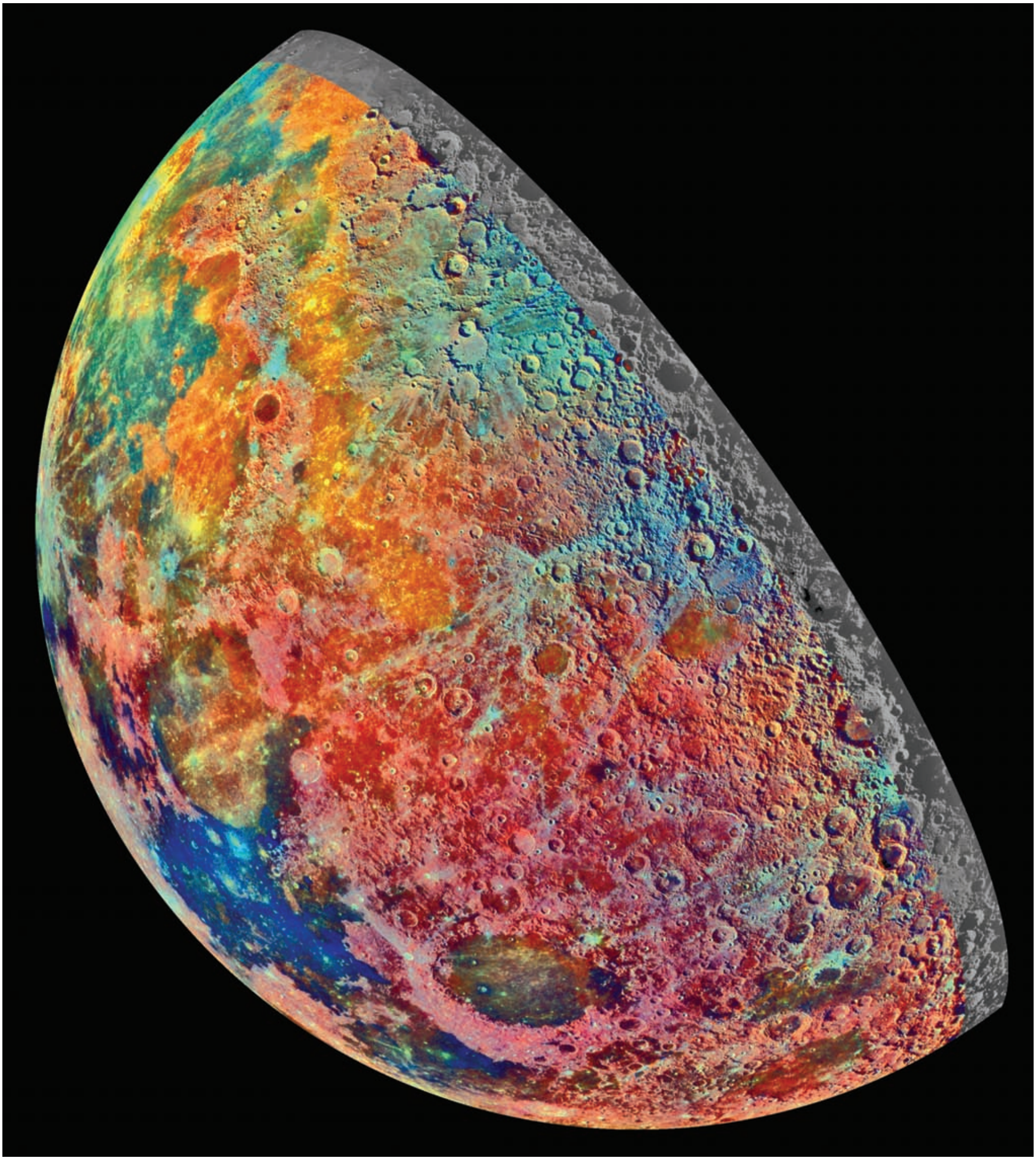


Fig. 5.23 Compositional variations on the Moon's surface This mosaic of images, taken through three spectral filters, shows exaggerated false-color differences in reflected sunlight in order to specify composition differences on the lunar surface. The image shows volcanic flows with relatively high titanium content (*blue*), volcanic flows that are low in titanium but rich in iron and magnesium (*green, yellow and light orange*), and heavily cratered highlands that are typically poor in titanium, iron and magnesium (*pink and red*). In this view, taken by *Galileo* on 7 December 1992, bright pink highlands surround the lava-filled Crisium impact basin (*bottom*), and the dark blue Mare Tranquillitatis (*left*) is richer in titanium than the green and orange maria above it. The youngest craters have prominent blue rays extending from them. (Courtesy of NASA/JPL.)

magnesium. The heavily cratered highlands are typically poor in titanium, iron and magnesium.

In early 1994, the United States Department of Defense placed a small spacecraft in orbit about the Moon. In sharp contrast to the eight-year, $25 billion *Apollo* program, the tiny unmanned satellite required only two years and $75 million to build and launch.

Because one of the mission's byproducts was to prospect the surface mineral content of the Moon, the spacecraft was given the name *Clementine*, after the miner's darling daughter in the old Gold Rush ballad. However, this was not the main purpose of the spacecraft. It was built primarily as a military test of lightweight electronic imaging sensors that could detect the launch and track the flight of enemy ballistic missiles, possibly for use in a future star-wars missile shield.

Like its namesake, the spacecraft was "lost and gone forever" after orbiting the Moon for two months, but not without first chalking up an impressive list of accomplishments. Unlike the *Apollo Command Modules* that circled the Moon in low near-equatorial orbits, *Clementine* orbited across the lunar poles, permitting a global perspective as different regions rotated into view (Fig. 5.24).

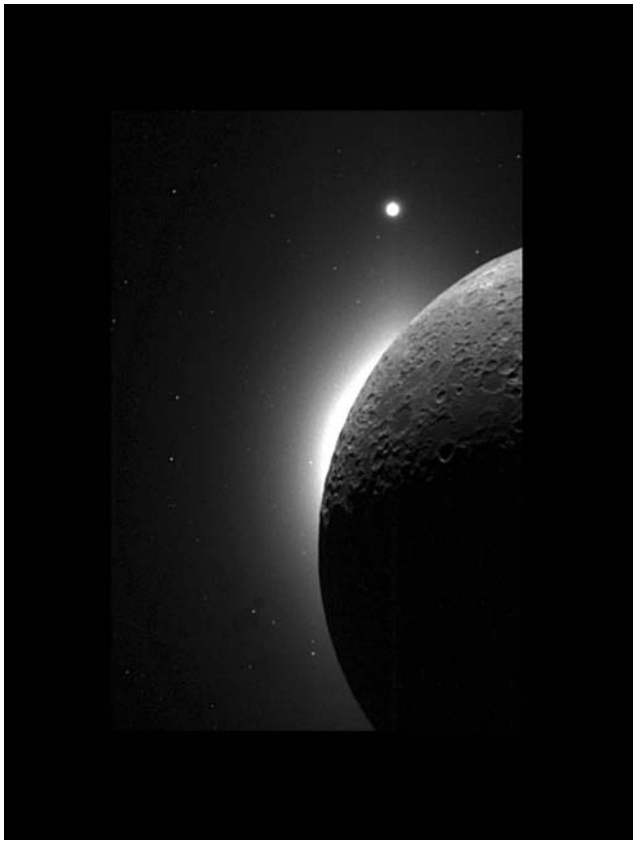


Fig. 5.24 *Clementine* observes the Moon, Sun, and Venus In this image, taken from the *Clementine* spacecraft in 1994, the Sun is just behind the Moon's limb, or edge, so most of the lunar surface is illuminated by light reflected from the Earth, known as Earthshine. Light from the Sun's outer atmosphere, the solar corona, produced the bright glow on the lunar horizon. The planet Venus is seen at the top of the frame. (Courtesy of NASA/JPL/USGS.)

Its ultraviolet, visible and infrared cameras took pictures at eleven different wavelengths used to identify different types of minerals. Since various rock-forming minerals reflect and absorb incident sunlight at different wavelengths, the *Clementine* data could be used to infer the chemical composition of most of the lunar surface, and rock samples could be used to calibrate the global data when it overlapped the *Apollo* landing sites.

The *Clementine* global data was used to map the abundance and distribution of iron on the Moon, showing that the dark, nearside maria consist of iron-rich lava, containing up to 14 percent iron by weight. In contrast, iron is practically absent in the nearside highland crust and across vast tracts of the far side, at about 3 percent iron by weight. These regions of very low iron content are dominated by aluminum-rich anorthosite.

The highland crust on both the near and far sides is just what one would expect if the entire Moon were once covered in liquid rock at least several thousand kilometers deep. The heavy iron sank into this magma "ocean", while low-density feldspar (anorthosite) grains accumulated into floating "rockbergs" that coalesced and cooled to form the eventual constituents of the Moon's highlands. Some of the iron subsequently resurfaced when the Moon heated up inside and magma flowed up to the nearside maria.

Large meteorite impacts dig holes into the lunar crust, exposing deeper material and revealing its composition. For instance, the floor of the South Pole – Aitken basin on the far side has an iron abundance of nearly 10 percent by weight.

Until *Clementine*, we also had no global map of the topography of the Moon. The laser altimeter on the spacecraft fired pulses of light at the Moon once every second and timed how long it took for the light beam to travel down to the lunar surface and back again. This enabled scientists to determine the distance to the surface, over and over again, with an accuracy of 50 meters. When these distances were combined with knowledge of the spacecraft orbit, maps of the elevation, or topography, of the entire lunar surface were obtained (Fig. 5.25).

The new global maps showed an unexpected range of heights, over 16 kilometers and comparable to that seen in the geologically different Earth. The wide range of relief on the Moon is caused by the presence of large impact basins. The huge South Pole – Aitken basin, which is over 12 kilometers deep and about 2600 kilometers across, dominates the farside topography. The near side is relatively

Table 5.4 The *Clementine* and *Lunar Prospector* missions to the Moon

Mission	Launch date	Lunar orbit	Accomplishments
Clementine	25 Jan. 1994	19 Feb. 1994 to 3 May 1994	Global surface composition, global topography, map of South Pole – Aitken basin, possible water ice at poles
Lunar Prospector	6 Jan. 1998	15 Jan. 1998 to 31 July 1999	Global elemental abundance, global magnetic field maps, global gravity maps, detection of lunar core, water ice at poles.

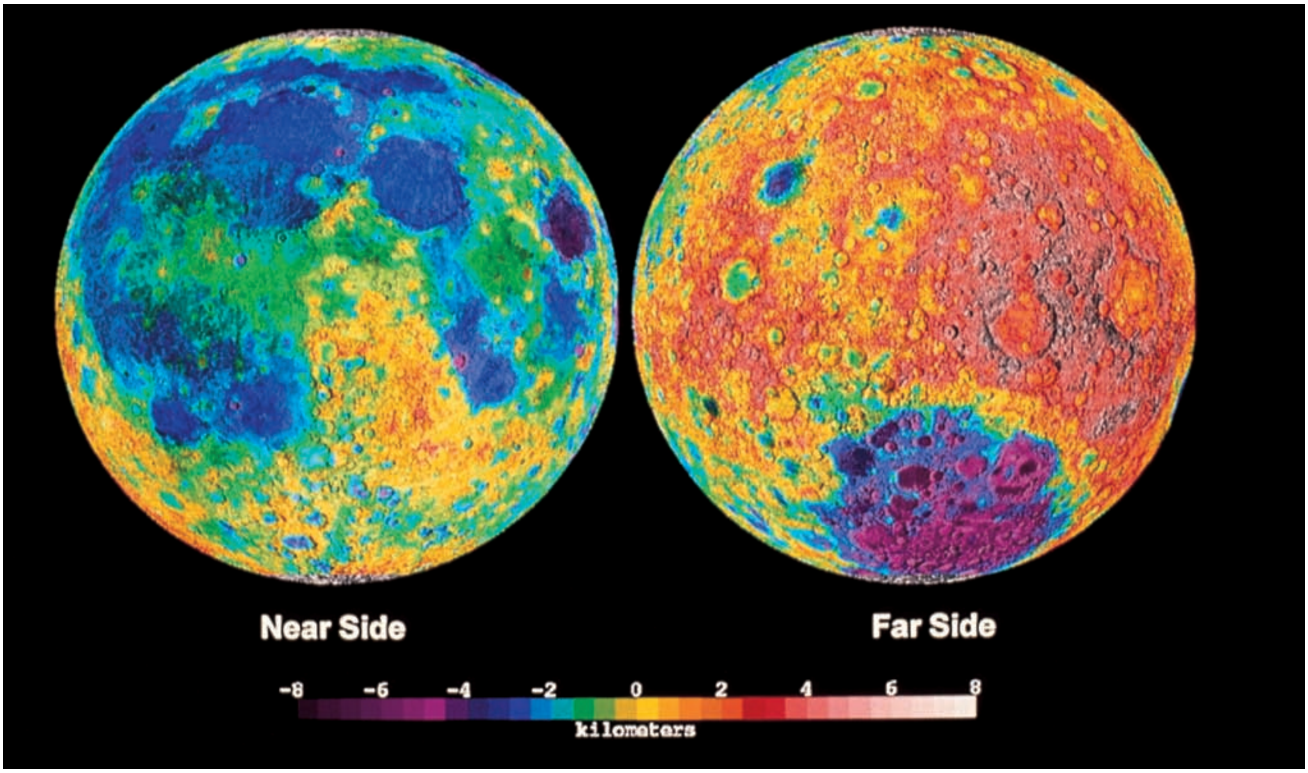


Fig. 5.25 Lunar topography The laser altimeter on *Clementine* provided the first comprehensive topographic map of the Moon. The contour interval is 500 meters, and the altitude in kilometers is coded by color from blue to red (*bottom*). The near side (*left*) is relatively smooth and low (*blue* and *purple*), primarily because of the prominent impact basins, including Imbrium, Crisium and Nectaris, which are at least partly filled with mare basalt. In contrast, the far side (*right*) shows high relief (*red*) and extreme topographic variation comparable to that of the Earth. The Moon's wide altitude range is attributed to ancient impact basins that have been preserved for about 3.9 billion years, while the Earth's wide range stems from ongoing mountain building by colliding tectonic plates. The large circular feature on the southern far side (*right bottom*) is the South Pole – Aitken basin, 2600 kilometers in diameter and 12 kilometers deep. (Courtesy of Paul D. Spudis, Lunar and Planetary Institute.)

smooth, with typical relief of about 5 kilometers, primarily because its impact basins have been filled with mare basalt. As substantiated by *Clementine* gravity data, a thicker crust blocks the outward flow of magma on the far side.

In the face of dwindling budgets and growing public interest in problems here on Earth, scientists and engineers found less-expensive ways of exploring space. NASA adopted a new "smaller, faster, cheaper" mode of operation, in which several cost-effective, high-risk spacecraft rather than a few major, expensive and low-risk ones do science. The *Lunar Prospector* spacecraft is an example. Launched on 6 January 1998, it was designed to obtain global data on elemental abundance, magnetic fields and gravity fields, and it has also achieved an impressive array of accomplishments (Table 5.4).

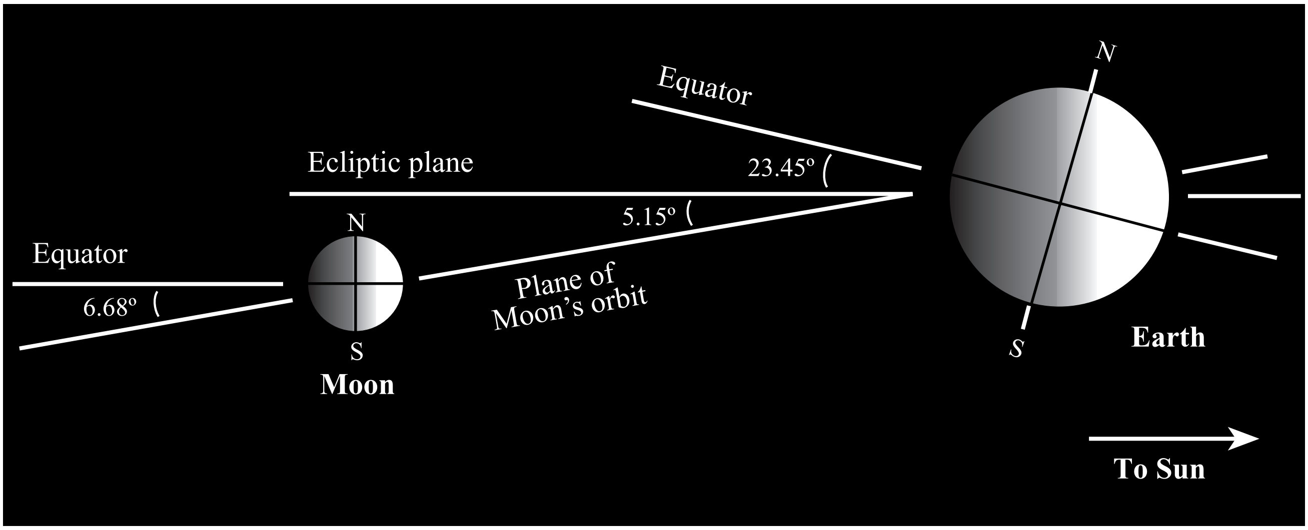


Fig. 5.26 Dark, cold lunar poles The near-vertical orientation of the Moon's north–south rotation axis to the ecliptic plane creates permanent night and deep freeze at the floor of craters located at the lunar poles. These regions might be reservoirs of water-ice, delivered there by comets. The angle between the Earth's equator and the ecliptic, or the plane of the Earth's orbit around the Sun, is 23.5 degrees, and this tilt produces the seasons. The Moon provides a steadying influence for the Earth's tilt, keeping it from varying widely and producing dramatic climate variations. Also note that the plane of the lunar orbit falls neither in the Earth's orbital plane nor in the ecliptic.

The Moon is not completely dry

Bright radar echoes, returned to *Clementine* from the south pole of the Moon, suggested that this region might contain radar-reflective water ice. Instruments aboard the *Lunar Prospector* spacecraft then strengthened the possibility of water ice at the south pole, and also discovered what appears to be additional ice near the north pole. During its passes over the poles, an instrument on *Lunar Prospector* detected substantial quantities of hydrogen, which mission scientists attributed to water ice found in permanently shaded areas near both poles. They estimate that there could be as much as 6 billion tons, or 6×10^{12} kilograms, of water ice located in the polar regions.

The Moon's rotation axis is orientated nearly perpendicular to the ecliptic plane (Fig. 5.26), so the lunar poles are never tilted toward the Sun by more than a very small amount. This means that the bottoms of craters at the poles are in constant shadow and in a perpetual deep freeze with temperatures of 50 to 70 kelvin. Any ice deposited in these frozen reservoirs would be preserved indefinitely in the eternal dark and cold.

Hopes for water on the Earth's Moon were strengthened in 2009 when the Moon Mineralogy Mapper aboard India's *Chandrayaan-1* spacecraft revealed the infrared signatures of water surrounding a small, fresh impact crater on the far side of the Moon (Fig. 5.27). An infrared mapping instrument aboard the *Cassini* spacecraft, on its way to Saturn, confirmed the existence of lunar water, and the *LCROSS* spacecraft bombed a shadowed polar crater, discovering water in the ejected plume. Unfortunately, none of the observations reveal large quantities of water. Even the dampest possibility leaves the lunar surface drier than almost any place on Earth's surface.

But how could there be water ice on the Moon when the rocks returned from the Moon show no signs of ever being exposed to significant amounts of water? The rocks were taken from near-equatorial regions where intense sunlight would boil away any liquid water or water ice, and the Moon's fiery origin seems to have removed volatile elements during the satellite's formation.

Water ice could have been delivered to the Moon by comets, which are essentially big balls of dirty ice. Comet-borne water could have been deposited in the cold traps at the top and bottom of the Moon for more than 4 billion years, ever since the Moon's rocky crust formed, slowly accumulating in amount.

If there were a source of water on the Moon, it would make it more attractive for human outposts. Water could be purified to drink, or it could be chemically split into hydrogen, to burn as a rocket propellant, and oxygen to breathe. This would make it easier to establish a colony on the Moon, or to build a fueling station on it for interplanetary spacecraft. Nevertheless, most of the water on the Moon is not immediately accessible. It is incorporated in the rocky interior of the Moon. Moreover, the concentrations of water on the lunar surface are so low that its extraction is impracticable.

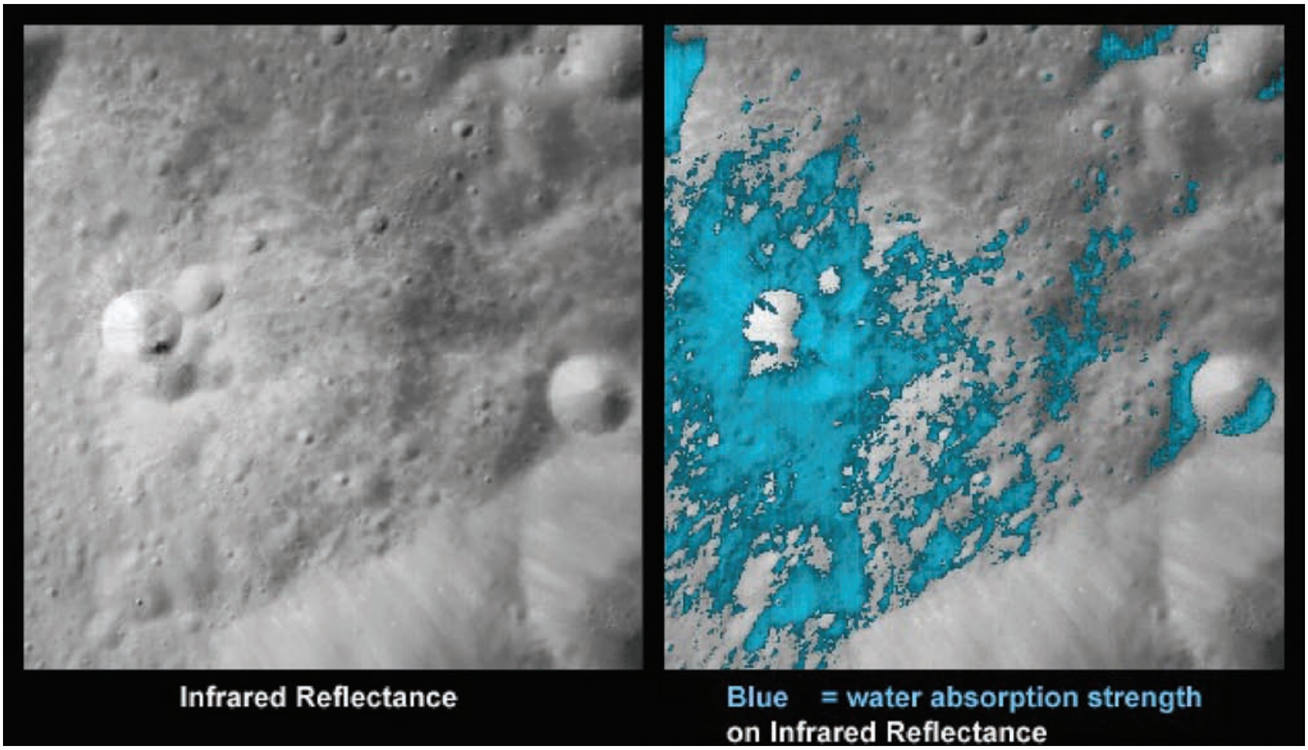


Fig. 5.27 Water around a fresh crater on the Moon These images show a very young crater or the side of the Moon that faces away from the Earth, as viewed from NASA's Moon Mineralogy Mapper instrument aboard the Indian Space Research Organization's *Chandrayaan-1* lunar-orbiter spacecraft. Sunlight reflected at infrared wavelengths (*left*) shows a very young, fresh crater on the far side of the Moon. On the right, the distribution of water-rich minerals, seen in absorption at a specific infrared wavelength (*false-color blue*), is shown around the crater. Small amounts of both water and hydroxyl-rich material were found to be associated with material ejected from the crater, but the amount of water is very small. (Courtesy of ISRO, NASA, JPL-Caltech/USGS/Brown University.)

The magnetized Moon

The Moon has no overall dipolar magnetic field, at least none that is strong enough to be detected. Its magnetic moment is at least 10 million times weaker than the Earth's. Yet some of the lunar rocks returned to Earth are magnetized. They have survived since the time that molten rocks covered the Moon and solidified 3 to 4 billion years ago, preserving fossilized remnants of ancient magnetic fields.

The *Apollo 15* and *16* missions each carried small subsatellites designed to measure the Moon's magnetism from lunar orbit. They found localized regions on the Moon with surface magnetic fields of between one-hundredth and one-thousandth that of Earth; the Earth's equatorial magnetic field strength is 3×10^{-5} tesla. Large blocks of the lunar crust, as broad as 100 kilometers, are magnetized, but they do not combine into an overall global pattern.

Lunar Prospector measurements have shown that the largest concentrations of strong magnetic fields are on the lunar far side, located diametrically opposite to the Imbrium, Serenitatis, Crisium and Orientale impact basins on the near side. The basin rock is itself weakly magnetized, suggesting that the large basin-forming impacts demagnetized the crust at the impact side, while simultaneously magnetizing the crust on the opposite side of the Moon.

According to one explanation, the young Moon may have had its own global magnetic field generated early in its history when molten metal circulated in a small core. Large impacts, like the one creating the Imbrium basin, would create an ionized fireball racing around the Moon, piling the magnetic field up and concentrating it at the point opposite the impact. As the rocks cooled, the strongest, localized magnetic fields survived as fossils after the Moon lost its global magnetic field.

This is not the only explanation for lunar magnetism, for there are hundreds of smaller magnetized regions scattered over the entire Moon, in regions that are not located opposite to a large impact site. Perhaps they swept up, amplified and concentrated the magnetized winds that flow from the Sun. On the other hand, the ancient magnetism may have been incorporated from surrounding

Table 5.5 Voyages to the Moon in the early 21st century

Country/Region	Mission	Launch date	Orbit entry	Accomplishments
Europe	*SMART-1*	27 Sept. 2003	15 Nov. 2004	Tested propulsion technology, mapped lunar surface composition, impacted Moon on 3 Sept. 2006
Japan	*Kaguya*	14 Sept. 2007	3 Oct. 2007	Obtained altitude, geological and gravity data, impacted Moon on 10 June 2009
China	*Chang'e-1*	24 Oct. 2007	5 Nov. 2007	Obtained images of landforms, geological structures, and distribution of chemical elements, impacted Moon on 1 March 2009
India	*Chandrayaan-1*	22 Oct. 2008	8 Nov. 2008	Sent probe into lunar south polar region on 14 Nov. 2008, discovered small amounts of water on lunar surface, mission terminated on 29 August 2009
United States	*Lunar Reconnaissance Orbiter* and *LCROSS*	18 June 2009	23 June 2009	Obtained maps of the lunar surface intended to specify potential landing sites and resources for future human exploration of the Moon. Simultaneous launch of *LCROSS* with impact probe that confirmed trace amounts of water ice and vapor in south polar region of the Moon

material during the early stages of formation, or the Moon could have been magnetized by the Earth, at a time when the two bodies were closer together.

This is one of the remaining mysteries yet to be solved by the future space-age exploration of the Moon, already begun in the early 21st century.

5.7 Return to the Moon

A new space race to the Moon

Space agencies from China, Europe, India, Japan and the United States have all sent spacecraft to the Moon in the early 21st century (Table 5.5), with extended plans to return. Altruistic scientific goals are not the main impetus this time around, any more than they were during the first flights. Each agency is eager to demonstrate its technological skills, and to elevate their international status. In some instances, there might also be close ties between the developing space program and defense or military objectives. But there is a genuine spirit of international collaboration in these new lunar missions, including instrument sharing between countries.

The European Space Agency's first-ever trip to the Moon began in 2003 with the launch of the low-cost *SMART-1* spacecraft, the first of its *Small Missions for Advanced Research in Technology*. As the name suggests, its primary goal was flight-testing of new technologies, such as solar-powered electric propulsion. After a long testing cruise, lasting more than a year, *SMART-1* entered lunar orbit and its instruments mapped and studied the Moon in great detail for more than a year and a half. They surveyed lunar resources, mapping the surface composition with then unrivalled resolution, and investigated potential landing sites and outposts for future missions. The spacecraft was intentionally sent into a controlled impact with the lunar surface on 3 September 2006 in order to study the ejected material.

The next modern lunar mission, *Kaguya*, was launched on 14 September 2007 by Japan. Kaguya is the name of a princess from the Moon found in a bamboo thicket by an old couple, and described in the ancient Japanese folktale *The Tale of the Bamboo Cutter*. This name replaced the more austere one of *SELENE*, an acronym for *SELenological and ENgineering Explorer*. The spacecraft's objectives were to study the origin and evolution of the Moon and to serve as an engineering test for future deep-space missions. Its instruments obtained improved lunar topographical maps and geological data using a laser altimeter and X-ray and gamma-ray spectrometers. After 1 year and 8 months of lunar orbit, *Kaguya* was intentionally impacted on the lunar surface, on 10 June 2009, for an examination of its impact plume by Earth-based telescopes. *Kaguya* also carried and released two smaller satellites to study the Moon's gravity and ionosphere, named *Okina* and *Ouna*, also from Japanese folklore.

The People's Republic of China entered the modern race to the Moon on 24 October 2007 with the launch of *Chang'e-1*, a lunar orbiter designed: to make three-dimensional imagery of the Moon; to study the chemical elements and surface composition of the lunar surface, including the terrestrially-rare isotope helium-3; and to investigate the prospect of mining the Moon. The spacecraft is named for an angel in a Chinese fairy tale who takes a magic potion and flies to the Moon.

Chang'e-1 entered lunar orbit on 5 November 2007 and was operated until 1 March 2009, when it was deliberately taken out of the orbit and made to impact the Moon. During orbit, the spacecraft obtained images of the landforms and geological structures of the entire lunar surface, and analyzed and mapped the distribution of chemical elements. As the number 1 suggests, it is the first of an ambitious series of Chinese spacecraft, including the launch of a sister orbiter, *Chang'e-2*, in 2011, and the landing of a rover on the Moon and a sample-return mission during the subsequent decade. China has also expressed interest in a manned Moon landing.

India's national space agency, the Indian Space Research Organization (ISRO) launched the lunar orbiter *Chandrayaan-1*, or first moon craft, on 22 October 2008, entering lunar orbit on 8 November 2008. A few days after that, it released an impact probe, making India the fourth country to touch down on the lunar surface. The main orbiting spacecraft was originally intended for two years of mapping the chemical and mineralogical characteristics of the lunar surface, as well as its geological and topographical details. Unfortunately, the failure of the spacecraft's star sensors and problems with overheating led to the termination of the mission after 10 months in space. Many of its primary objectives were nevertheless obtained, including the discovery of trace amounts of water on the lunar surface and confirmation that the Moon was once completely molten. ISRO is planning a *Chandrayaan-2*, and hopes to land a motorized rover on the Moon in 2012 as part of this second mission. Manned spaceflights and a manned mission to the Moon are also part of their future agenda.

In 2004 President George W. Bush (1946–) announced plans to send American astronauts back to the Moon and later Mars. Summoning the spirit of discovery, exemplified by the Lewis and Clark exploration of the vast American West two centuries before, Bush called on the human need to venture into space, "for the same reason we were once drawn into unknown lands and across the open sea", urging NASA to return people to the Moon no later than 2020.

NASA's first step back to the Moon was achieved with the launch of its *Lunar Reconnaissance Orbiter* (*LRO*) on 18 June 2009. Instruments aboard the spacecraft are providing scientists with a detailed global topographical map of the Moon, specifying its impactor populations as well as potential landing sites, and with global high-resolution infrared maps of the Moon, enabling detection of previously unseen compositional differences. The *Lunar Crater Observation and Sensing Satellite* (*LCROSS*) was launched with the *LRO* mission. *LCROSS* included a small probe that was sent into a controlled impact within a crater in the south polar region of the Moon. Scrutiny of the ejected material confirmed that small amounts of water were present.

This bold plan to return Americans to the Moon was grounded by President Barack Obama (1961–) in February 2010, after NASA had already spent 9.1 billion dollars on the program. Obama instead proposed 6 billion dollars over five years to encourage private companies to build spacecraft that NASA could rent to ferry astronauts to the *Space Station* – after the *Space Shuttle* program is terminated.

Nevertheless, the idea of a human return to the Moon still has its attractive aspects. Technology can now be used that wasn't even imagined when humans first walked on the Moon, more than 40 years ago. The *Apollo* spacecraft, for example, carried handheld cameras to photograph the lunar surface. Now high-resolution imaging and remote sensing from lunar orbit are common.

NASA's cancelled plans included the future use of robotic lunar rovers, similar to those used on Mars, to pave the way for riskier human missions. Eventually a lunar outpost was envisaged, as a permanent base and way station for trips to Mars. With only one-eightieth of the Earth's mass, the lower gravitational pull of the Moon would make it easier and less costly to launch missions to other parts of the solar system, using spacecraft assembled and provisioned on the Moon with its abundant resources.

Although the prospects look dim right now, the United States might eventually establish a permanent human settlement on the Moon. It could serve as a support station for exploration of the rest of the solar system, including manned trips to Mars. Oxygen, hydrogen and metals might be extracted from the lunar soil to produce water, air, fuel and construction material. The Moon might even be mined for scarce resources on the Earth, such as helium-3: the gas that makes balloons float and might be used in fusion reactors in the future. The global reserves of helium-3 on the entire Earth amount to about 100 kilograms, but there is perhaps 10 million times as much helium-3 on the Moon's surface, implanted there by the Sun's winds. And since the Moon has no appreciable atmosphere, some promising astronomy could be accomplished from it.

Astronomy from the Moon

A manned scientific station on the Moon would provide excellent opportunities for astronomy in the future. With no significant atmosphere, every wavelength of radiation, from the longest radio waves to the shortest gamma rays, streams down without absorption to the lunar surface. Telescopes in artificial Earth satellites are now used to observe cosmic ultraviolet and X-ray emission that is partially or totally absorbed in the Earth's atmosphere, but such instruments are constrained by their weight, size, complexity and cost. Telescopes constructed on the rock-solid surface of the Moon would rest on a more stable platform than an orbiting satellite or space station, permitting high-resolution observations of the Universe at wavelengths that cannot be seen from the ground, including X-rays and gamma rays.

The near absolute-zero temperatures and airless environment on the Moon would also avoid the problem of image blurring caused by atmospheric turbulence that usually limits angular resolution at visual wavelengths to 1.0 seconds of arc. A lunar optical telescope as small as 1 meter in diameter would have a resolution of about 0.1 seconds of arc, while also avoiding interruption caused by cloudy weather and daytime. Relatively small telescopes could be linked together electronically, using the techniques of interferometry to achieve angular resolutions at visible wavelengths that are thousands of times better than those currently available from either space or the ground. The great increase in resolution might, for example, be used to investigate Earth-like planets around nearby stars, or to obtain focused infrared observations for weeks and months at a time, detecting the most distant galaxies as they are forming.

The Moon's far side, permanently shielded from the noisy, interfering Earth, is an ideal site for future radio astronomy. It could also be used to open a new window on the Universe at wavelengths longer than about 20 meters that are reflected by the Earth's ionosphere and do not reach the ground.

5.8 The Moon's history

The age of the oldest rocks – Moon, Earth and meteorites

The time at which different features on the Moon originated can be determined from rocks returned from them. These relics have remained unaffected by the erosion that removed the primordial record from most terrestrial rocks. The ages of the lunar rocks can be determined by examining unstable radioactive elements and their stable decay products.

What matters are not the actual quantities of each element present, but the proportions – the ratio of stable elements, like lead, to unstable ones like uranium and thorium. When this ratio is combined with the known rates of radioactive decay, the time since the rock solidified and "locked in" the radioactive atoms is found.

The method is known as *radioactive dating*, and it works this way. Certain types of nuclei, known as unstable parent isotopes, decay at a constant rate into stable lighter isotopes known as daughters. By measuring the amount of daughter material and knowing the rate of decay, the age of the rock can be estimated. The detailed mathematical treatment is given in Focus 5.4. The method is something like determining how long a log has been burning by measuring the amount of ash and watching a while to determine how rapidly the ash is being produced.

The daughter isotopes must be trapped in the rock and not escape or the estimated age will be too short. In fact, the daughters can escape quite easily when the rock is molten; only when it cools and solidifies do the daughters start to accumulate. For this reason, the ages determined for the rocks are really the times since the rock became solid. And if the rock is re-melted, say by the impact of a meteorite, its radioactive clock is reset, and the age will measure the time since the last solidification.

The radioactive dating method has been used to study rocks returned from the Moon. The oldest lunar samples, returned from the light, rugged highlands, indicate an age of nearly 4.6 billion years, and the lava flows that created the dark, nearside maria are dated at 3.2 to 3.9 billion years ago.

The oldest rocks found on Earth are about 3.9 billion years old, and the oldest known terrestrial minerals are found in crystals of zircon that have been dated to between 4.1 and 4.3 billion years old. Erosion by wind, water and geological processes has wiped out the oldest terrestrial rocks. The deep-ocean sediments, that are least affected by continuing geological activity on Earth, have ages of about 4.55 billion years.

Primitive meteorites known as carbonaceous chondrites have an age of 4.566 billion years, with an uncertainty of about 0.002 billion years. These meteorites are thought to date back to the earliest days of the solar system.

Rounding off the numbers and allowing for possible systematic errors, we can say that the Earth, Moon and primitive meteorites solidified at about the same time some 4.6 billion years ago, with an uncertainty of no more than 0.1 billion years. If the solar system originated as one entity, then this should also be the approximate age of the Sun and the rest of the solar system.

Focus 5.4 Radioactive dating

Radioactive elements can be used to clock the age of rocks on the Earth's surface, meteorites, and lunar rock samples. The number, N, of radioactive atoms in a rock changes with the time, t, since its solidification according to the differential equation:

$$\frac{dN}{dt} = -\lambda N$$

where λ is the decay rate. This equation integrates to give the number of radioactive atoms, N_t, at time t:

$$N_t = N_0 \exp(-\lambda t) = N_0 \exp\left(\frac{-0.693\, t}{\tau_{1/2}}\right)$$

where N_0 is the number of atoms at time $t = 0$, the time of solidification. The radioactive decay constant λ is $0.693/\tau_{1/2} = \ln(2)/\tau_{1/2}$, and $\tau_{1/2}$ is the half-life of the radioactive species. Half-lives for the decay of radioactive isotopes are given in Table 5.6.

The number of radioactive atoms in the rock will be halved in a time equal to the half-life. Radioactive uranium ^{238}U decays, for example, into lead ^{206}Pb, which is stable with a half-life of about 4.47 billion years; so every 4.47 billion years the amount of uranium-238 in a rock will be halved. We can apply the equations to ^{238}U, and express the abundance in terms of another kind of lead, ^{204}Pb, that is not a radioactive decay product. If a terrestrial rock, lunar sample or a non-terrestrial meteorite became a closed system at time $t = 0$, then the present abundance of lead and uranium are related by the equation:

$$\left(\frac{^{206}\mathrm{Pb}}{^{204}\mathrm{Pb}}\right)_t = \left(\frac{^{238}\mathrm{U}}{^{204}\mathrm{Pb}}\right)_t \left[\exp(\lambda_{238}\, t) - 1\right] + \left(\frac{^{206}\mathrm{Pb}}{^{204}\mathrm{Pb}}\right)_0$$

where the subscripts t and 0 denote the present and initial abundance, respectively.

If all of the rock samples have the same initial ^{206}Pb / ^{204}Pb abundance, and if all of them have the same age, t, then a plot of $(^{206}\mathrm{Pb}/^{204}\mathrm{Pb})_t$ against $(^{238}\mathrm{U}/^{204}\mathrm{Pb})_t$ should lie in a straight line of slope $[\exp(\lambda_{238}\ t) - 1]$. Such a plot is called an *isochron*. If a system formed t years ago and initially contained no lead, then a curve of the ratios ^{207}Pb / ^{206}Pb and ^{238}U / ^{206}Pb also provides the age t.

These and similar methods have been used to show that the Earth, Moon and meteorites are 4.6 billion years old, with an uncertainty of no more than 0.1 billion years, where 1 billion years = 10^9 years = one Gyr.

Table 5.6 Radioactive isotopes used for dating

Radioactive parent [name (symbol) mass no.]	Stable daughter [name (symbol) mass no.]	Half-life (millions of years)
Rubidium (Rb) 187	Strontium (Sr) 87	48 800
Rhenium (Re) 187	Osmium (Os) 187	44 000
Lutetium (Lu) 176	Hafnium (Hf) 176	35 700
Thorium (Th) 232	Lead (Pb) 208	14 050
Uranium (U) 238	Lead (Pb) 206	4 470
Potassium (K) 40	Argon (Ar) 40	1 270
Uranium (U) 235	Lead (Pb) 207	704
Samarium (Sm) 146	Neodymium (Nd) 142	100
Plutonium (Pu) 244	Thorium (Th) 232	83
Iodine (I) 129	Xenon (Xe) 129	16
Palladium (Pd) 107	Silver (Ag) 107	6.5
Manganese (Mn) 53	Chromium (Cr) 53	3.7
Aluminum (Al) 26	Magnesium (Mg) 26	0.72

Formation of the highlands and maria

The retrieved lunar rocks have taken us back into time, to the formative stages of the Moon. They record events from the earliest history of the solar system that have been erased on Earth by water, wind and geologic activity. Radioactive dating of Moon rocks and primitive meteorites indicates, for example, that the Moon was assembled a mere 50 million years after the solar system itself was born 4.6 billion years ago.

The cataclysmic bombardment associated with the final stages of the Moon's formation was so energetic that the globe was melted to depths of several hundred kilometers. In this global magma ocean, lighter mineral

Table 5.7 Lunar timescales

Period	Age[a] (billions of years)	Characteristics
Pre-Nectarian	4.6 to 3.92	Moon accumulated 4.5 to 4.6 billion years ago in Earth orbit, newly formed Moon wrapped in molten rock, a magma ocean, until 4.4 billion years ago, solidification of lunar crust and formation of oldest impact basins
Nectarian	3.92 to 3.85	Nectaris basin probably formed 3.9 billion years ago, Serenitatitis, Crisium and other impact basins are most likely this old, as are most lunar rocks
Imbrian	3.85 to 3.15	Period of lunar volcanism when most maria were formed, Imbrium basin excavated 3.85 billion years ago and the last big basin, Orientale, created 3.8 billion years ago; these basins were filled with lava to create the lunar maria up until 3.15 billion years ago
Eratosthenian	3.15 to about 1.0	Craters that are slightly degraded but without rays; most of lunar surface remains unchanged, but some mare volcanism and large impacts
Copernican	About 1.0 to present	Youngest craters formed, most of which have preserved rays; crater Copernicus excavated 0.85 billion years ago and crater Tycho created just 0.10 billion years ago; most of lunar surface remains unchanged

[a] These ages are estimates based on inferences of the geological setting of the lunar samples.

species floated to the top and formed the Moon's crust, and the denser material sank to the interior. The lunar highlands still contain the remains of these early low-density rocks, the anorthosites.

The magma ocean gradually cooled and crystallized between 4.6 and 4.4 billion years ago, forming a thin, low-density lunar crust. Portions of this crust are today's highlands, on both the near and far sides of the Moon, rich in light elements such as aluminum and poor in heavy ones like iron.

When the radioactive dating method is applied to highland rocks retrieved from the Moon, there are two significant results. Highland rocks that crystallized from internally generated magmas, such as the anorthosite rocks, date from the earliest times, 4.6 to 4.1 billion years ago – a span of 500 million years after the origin of the Moon. In contrast, highland rocks that assembled from pre-existing rocks by impact, the breccias, all date from 3.9 to 3.8 billion years ago. The two results attest to prolonged igneous evolution on the early Moon, followed by an apparently short, "cataclysmic" impact bombardment about 3.9 billion years ago. This hail of meteoritic debris cratered the lunar highlands and obliterated most of the direct evidence of the first half billion years of lunar history.

The large impact basins were formed during the final stages of this heavy bombardment. The Nectaris basin was created 3.92 billion years ago, and the Imbrium impact took place an estimated 3.85 billion years ago. These impacts have been used to mark key events in the lunar history (Table 5.7). In pre-Nectarian time, the Moon's crust solidified. The major impact basins were gouged out during the heavy bombardment of Nectarian time. When the Imbrium basin was excavated, it marked the start of the Imbrian period of lunar volcanism that created the maria.

As the external cratering rate was declining rapidly, internal processes set to work. The radioactive decay of long-lived unstable elements, such as uranium and thorium, produced heat that gradually warmed up the lunar interior. There followed an era of volcanism, lasting for 700 million years, from 3.9 to 3.2 billion years ago. The outer zone of solid rock gradually cooled from the outside in, becoming thicker, and lava worked its way from deeper and deeper in the Moon. The magma flow may have stopped 3.2 billion years ago, since the youngest lunar lava samples are this old, but some mare basalts could be as young as 1 billion years old.

As molten basaltic rock welled up from the interior, it penetrated the thin crust beneath the great impact basins on the near side of the Moon, flooding them with lava and producing the dark circular maria that can be seen today (Fig. 5.28). Successive lava flows set their marks in some maria, showing that they were not formed in a single quick pulse of volcanism, but by repeated outpourings that gradually filled the nearside basins.

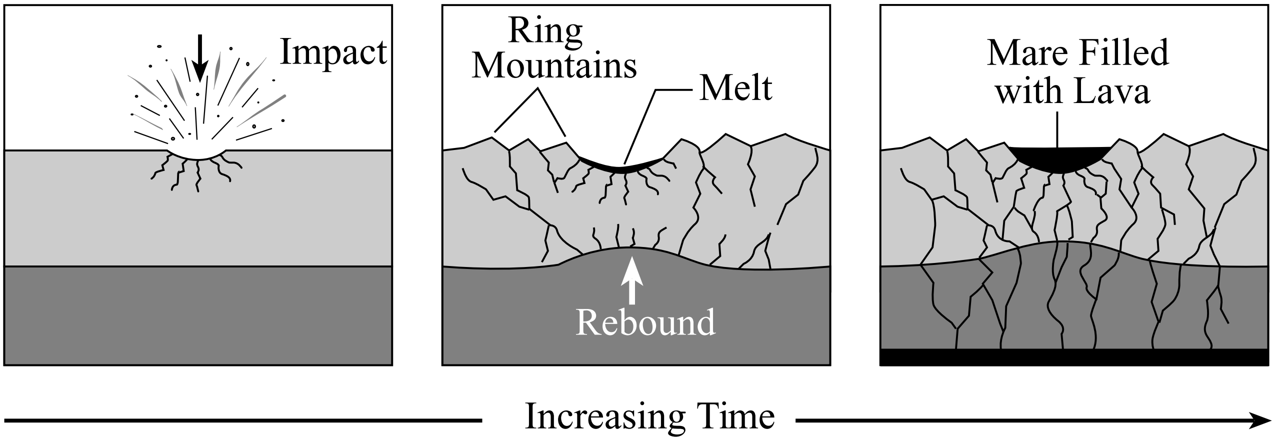


Fig. 5.28 Mare formation on the Moon Disintegrating and vaporizing as it strikes, a meteorite blasts a huge impact basin out of the lunar surface (*left*), while the associated shock waves create fractures in the rock beneath the basin. The blast hurls up mountain ranges around the basin (*middle*), and the underlying rock adjusts to the loss of mass above it by rebounding upward. The uplifted mantle causes additional fractures in the rock, while a pool of shock-melted rock solidifies in the basin. All the major impact basins on the Moon were created in this way between 4.3 and 3.9 billion years ago. Later, interior heat from radioactivity caused partial melting inside the Moon, and magma rose along the fractures, filling the basin with lava to form a dark mare (*right*). The lunar maria were filled by this volcanic outpouring between 3.9 and 3.1 billion years ago.

The liquid lava moved quickly away from its vents, covering the sources rather than piling up, so no volcanic mountains were created. Flowing with about the consistency of motor oil, the lava spread for hundreds of kilometers before hardening into a thin veneer, only a few hundred meters or less in thickness. And since the crust on the far side of the Moon was relatively thick to begin with, the molten rock had greater difficulty in penetrating it, explaining why there are so few maria on that side of the Moon.

The lava inundated all craters in its path, wiping the slate clean of previous impacts and preparing a fresh surface to record new impacts, which, by this time, had greatly diminished in intensity. Thus the maria are relatively unscarred and most of their craters are small and relatively young.

Despite its violent beginnings, the Moon then settled down into a long quiet life. The airless Moon's face has remained largely unchanged for 3.2 billion years. By that time, the Moon had cooled so much that magma could no longer break through and erupt, and the pummeling by meteorite impacts continued at a much lower rate. Impacts occasionally blasted out craters like Copernicus, some 800 million years old, and Tycho, dated at 108 million years, but the ongoing rain of lesser impacts mainly churned up the lunar surface, covering it with rock fragments. This rubble forms a dusty covering to a magnificent, ancient museum of craters and basins formed billions of years ago.

5.9 Tides and the once and future Moon

The pattern of the tides

Walking along the ocean beach some morning, we might notice that the waves seem to be reaching farther and farther up the sand. The tide is flooding the beach. A few hours later, it hesitates and then begins to ebb, retreating onto the flats where the clams may often be found. The high tides occur simultaneously and symmetrically on opposite sides of the Earth; they return every 12 hours 25 minutes in each location, although not precisely to the same height. The time between consecutive high tides is slightly more than half a day because the Moon's revolution around Earth is in the same direction as the Earth's rotation on its axis, so Earth needs an extra 25 minutes of rotation to out-race the Moon and get into position.

The Moon creates two high tides because the gravitational force of the Moon draws the ocean out into an ellipsoid, or the shape of an egg. We can understand this by remembering that the gravitational force decreases with distance, so the Moon pulls hardest on the ocean facing it, and least on the opposite ocean; the Earth between is pulled with an intermediate force. As a result, the water directly beneath the Moon is pulled up away from the Earth's center, and the Earth's center is pulled away from the water on the opposite side, causing another high tide. Thus the differences of the gravitational attraction of the Moon on opposite sides of the Earth produce two tidal bulges – one facing the Moon and one facing away (Fig. 5.29).

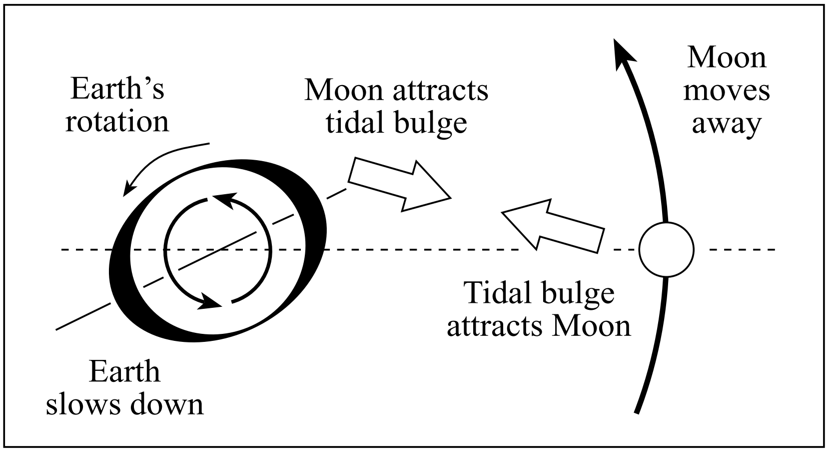


Fig 5.29 Cause of the Earth's ocean tides The Moon's gravitational attraction causes two tidal bulges in the Earth's ocean water, one on the closest side to the Moon and one on the farthest side. The Earth's rotation twists the closest bulge ahead of the Earth–Moon line (*dashed line*), and this produces a lag in time between the time the Moon is directly overhead and the highest tide. The Moon pulls on the nearest tidal bulge, slowing the Earth's rotation down. At the same time, the tidal bulge nearest the Moon produces a force that tends to pull the Moon ahead in its orbit, causing the Moon to spiral slowly outward.

As the Earth's rotation carries the continents past the tidal humps, we experience the rise and fall of water. In mid-ocean the tide is only 0.01 to 0.30 meters in height and usually goes unnoticed. But, when a shore blocks the tide, it often runs 2 or 3 meters high. Tides can also resonate in estuaries of the right shape, amplifying and building up the tides to 10 or 20 meters in height, as in the Bay of Fundy in Nova Scotia. You can create a similar effect by sloshing the water in your bathtub back and forth at just the right rate.

On a slowly rotating planet without continents, the tide would be highest along the line joining the centers of the planet and its moon – that is, when the moon is overhead. This is not the case for the Earth. The friction of the continents and the rapid rotation of the Earth carry the ocean's tidal bulge forward so it precedes the Earth–Moon line by about 3 degrees (Fig. 5.29). This means that in the open ocean the high tide actually occurs about 12 minutes after the Moon is overhead.

There are further delays for tides at the continental shores. When the flood tide moves in from the ocean, it may have to work its way among islands and peninsulas and along channels; this twisted path will delay the arrival by amounts that vary with location and with the time of the month. This time delay is called the *establishment of the port*, and the result is that high tide usually occurs an hour or two after the Moon is overhead, and occasionally more. Similar delays can be noted in tidal pools. They are lowest long after low tide.

Most people think that the Moon alone causes the tides, but that is not the case. The Sun and the Moon both contribute to the formation of the tides, but the major portion of this rhythmic ebb and flood is driven by the Moon, whose tide is 2.2 times as high as the Sun's. In the course of a month, the changing alignment of the Sun and Moon causes the tides produced by these two bodies to alternately reinforce and interfere, leading to the cycle of spring tides and neap tides. Once a month the Sun and Moon are aligned, both on the same side of the Earth, producing the spring tide that swells exceptionally high.

The spring tide occurs near new and full moon, when the Sun and Moon reinforce each other's tides, and the neap tide occurs near first and third quarter, when they interfere with each other (Fig. 5.30). The spring tides can be two or three times as high as the neap tides. On new moon nights, the spring tide can pull a boat much farther out than the average quarter-moon neap tide. The Sun's tides also vary by a small amount over the year as the Earth travels around its eccentric orbit and the Sun–Earth distance changes, with the greatest solar tides when the Earth is nearest the Sun.

The days are getting longer

As the Earth rotates, the bulge raised on its surface by the Moon's gravity is always a little ahead of the Moon rather than directly under it. The Moon pulls back on the bulge, and in the process it slows the whole planet down. In other words, our planet meets resistance in its daily rotation caused by the tidal interaction of the Moon with the Earth.

As the ocean tides flood and ebb, they create eddies in the water, producing friction and dissipating energy at the expense of the Earth's rotation. The motion of the

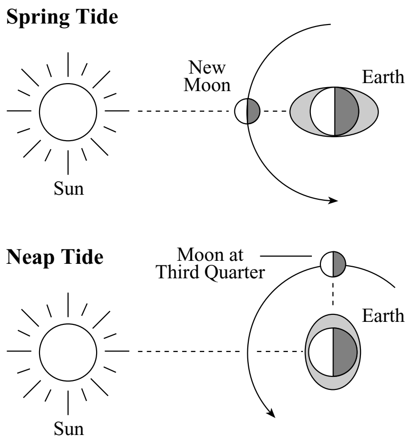


Fig. 5.30 Earth's spring and neap ocean tides The height of the tides and the phase of the Moon depend on the relative positions of the Earth, Moon and Sun. When the tide-raising forces of the Sun and Moon are in the same direction, they reinforce each other, making the highest high tides and the lowest low tides. These spring tides (*top*) occur at new or full moon. The range of tides is least when the Moon is at first or third quarter, and the tide-raising forces of the Sun and Moon are at right angles to each other. The tidal forces are then in opposition, producing the lowest high tides and the highest low tides, or the neap tides (*bottom*). The height of the tides has been greatly exaggerated in comparison to the size of the Earth.

tides heats the ocean water ever so slightly and the Earth's rotation is slowed. The tides therefore act as brakes on the spinning Earth, slowing it by friction in much the way that the brakes of a car slow its wheels and become warm. As the result of this tidal friction, the rotation of the Earth is slowing down and the day is becoming longer at a rate of 2 milliseconds, or 0.002 seconds, per century (Focus 5.5). In other words, the days are getting longer at the rate of one second every 50 000 years, and tomorrow will be 60 billionths of a second longer than today.

This tidal effect on the Earth's rotation can help us understand the ancient astronomical records, and conversely those records also help us understand the effect of the tides. If the current, slow rotation rate is used to wind the Earth backward 2500 years to the time of an eclipse, the Earth would rotate about a quarter-turn too little, putting those predicted eclipse paths several thousand kilometers west of their actual locations. These paths would then conflict with the reported occurrences. As we shall see, the length of the month is also increasing; this reduces the discrepancy considerably, so the full story is a bit complicated.

Indirect historical measures of the Earth's rotation have been made by paleontologists through studies of fossil corals. The growth patterns of these corals consist of annual bands and fine daily ridges, produced by the effect of seasonal and daily changes of water temperature on the growth rate. The days were shorter in the past, but the year was the same, so the number of days per year increases as we go back in time. Ancient corals confirm this, and they show a greater number of daily ridges per annual band than modern corals. Careful counting reveals that the day was only 22 hours long when we look back 400 million years. Studies of daily grown increments have been extended to fossilized algae called stromatolites, which indicate that the day may have been only 10 hours long 2 billion years ago.

Aside from such historical and paleontological determinations, this change of the Earth's rotation is imperceptible to humans, and it has not yet been measured directly. It is also mixed in with an erratic rate produced by the vagaries of the weather and the seasons, so all in all the Earth's rotation is no longer the best choice of a clock. Astronomers now prefer to rely on atomic clocks due to their stability and continued accuracy.

The planets Mercury and Venus have exceptionally slow rotation periods, of 58.646 and 243 Earth-days, respectively. The youthful energy of their fast initial rotation has probably been tempered by tidal interaction with the massive nearby Sun. These would be tides in the solid body of the planets, for there are no oceans on Mercury or Venus. Such a tidal interaction is suggested by the fact that Mercury spins on its axis exactly three times during two full revolutions about the Sun, so its rotation period is exactly two-thirds of Mercury's orbital period of 87.969 Earth days.

Earth's tidal influence on the Moon

The Moon pulls the Earth's oceans, and the oceans pull back, in accord with Newton's third law that every action has an equal and opposite reaction. The net effect is to swing the Moon outward into a more distant orbit. This is because the tidal bulge on the side facing the Moon is displaced ahead of the Moon and this bulge pulls the Moon forward.

As the Earth slows down, the angular momentum it loses is transferred to the Moon, which speeds up in its orbit around us. It is not hard to see that this will swing the Moon away from the Earth if we look at the key equations (Focus 5.6). When we do the arithmetic, we find that the change of 0.002 seconds per century in the length of the day implies an outward motion of the Moon amounting to about 0.04 meters per year. Small as it is, this value is just measurable with the laser reflectors planted on the Moon by the *Apollo* astronauts. The lunar laser ranging data indicate that the Moon is moving away from the Earth at a rate of 0.0382 ± 0.0007 meters per year.

Focus 5.5 Tidal friction slows the rotation of the Earth

In most of the ocean, the tidal currents are confined to the top of the deep sea, never reaching its bottom. Most of the tidal energy is therefore dissipated in shallow seas near land, where the turbulent tidal water reaches the ocean bottom, at depths of 100 meters or less.

When the tide moves toward a beach at velocity V, the frictional energy ΔE dissipated by tidal currents on the sea bottom per unit time Δt and unit area ΔA is

$$\frac{\Delta E}{\Delta t \Delta A} = \gamma \rho V^3$$
$$\approx 2 \text{ joule per second per square meter}$$

where the density of seawater is $\rho \approx 1000$ kilograms per cubic meter, a typical velocity is $V \approx 1$ meter per second, and the stress on the sea bottom is $\gamma \rho V^2$ with an empirical drag coefficient $\gamma \approx 0.002$ for wind stress on the ground and a river's stress on its bed as well as tidal currents in the bottom of the sea.

In 1919, Sir G. I. Taylor (1886–1975), a British expert on turbulence in air and water, used this equation to obtain $\Delta E/\Delta t \approx 5 \times 10^{10}$ joule per second for the Irish Sea alone, and in the following year Sir Harold Jeffreys (1891–1989) estimated that the total energy dissipated by tidal friction in the shallow seas surrounding Europe, Asia, and North and South America is

$$\text{Energy lost by tidal friction} = \frac{\Delta E}{\Delta t}$$
$$\approx 10^{12} \text{ joules per second}$$

This is comparable to the estimate obtained by considering the flux of energy convected into the shallow seas by tidal currents.

The lost energy comes from the Earth's rotational energy, which is equal to

$$\text{Rotational energy} = \frac{1}{2} M_{\mathrm{E}} V_{\mathrm{rot}}^2 = \frac{2\pi^2 M_{\mathrm{E}} R_{\mathrm{E}}^2}{P_{\mathrm{E}}^2}$$

where the mass of the Earth M_{E} is 5.972×10^{24} kilograms, the mean radius of the Earth R_{E} is 6.371×10^6 meters, and the rotation period of the Earth P_{E} is 24 hours = 8.616×10^4 seconds. For a period change ΔP in time interval Δt, the loss in rotational energy is

$$\text{Energy lost by rotation} = \frac{4\pi^2 M_{\mathrm{E}} R_{\mathrm{E}}^2 \Delta P}{P^3}$$

Setting this equal to $\Delta E/\Delta t$ and collecting terms, we obtain:

$$\frac{\Delta P}{\Delta t} = \frac{P^3}{4\pi^2 M_{\mathrm{E}} R_{\mathrm{E}}^2} \frac{\Delta E}{\Delta t}$$
$$\approx 10^{-12} \text{ seconds per second}$$
$$\approx 0.003 \text{ seconds per century}$$

where one century equals 3.156×10^9 seconds.

Will the Moon's outward motion carry it away from the Earth altogether? Probably not, because there is not enough energy in the Earth–Moon system for these bodies to overcome their binding energy and go their separate ways. Only the intrusion of a massive third body could achieve that, or some fantastic project to attach enormous rockets to the Moon and launch it into space.

What will ultimately happen is the following. The combination of the slowing Earth and the receding Moon means that the Earth's day will eventually catch up with the length of the month. When the day and the month are equal, the Moon-induced tides will cease moving; from then on the oceans will rise and fall much more gently under the influence of the Sun. The Moon will hang motionless in the sky, and will be visible from only one hemisphere. At that stage the recession of the Moon will stop.

Then, billions of years from now, the Sun's tidal action will take over; slowing the Earth's rotation even further, until the day becomes longer than the month. At this point, angular momentum will be drawn from the Moon, and it will begin approaching the Earth, heading on a course of self-destruction until it is finally torn apart by the tidal action of the Earth. Perhaps it will form a ring around our planet. In any case, it will probably end its years where it apparently began – close to the Earth. By this time, however, the brighter Sun will have boiled the oceans away, and the Earth will have become a dry and barren place.

Stabilizing the Earth

The orientation of the Earth's rotation axis causes the annual seasonal variations of our climate, and small variations in its orientation contribute to the advance and retreat of the ice ages. The obliquity of the Earth, the angle that its rotation axis makes with the perpendicular to its orbital plane, is now a modest 23.5 degrees. This is sufficient to bring summer and winter as the northern or southern

Focus 5.6 Conservation of angular momentum in the Earth–Moon system

One of the fundamental, unbreakable laws of physics is the law of conservation of angular momentum, and this means that the angular momentum that the Earth loses in slowing down will be transferred to the Moon. That is, the product of mass M, velocity V and radius R is unchanged in a closed system, which is not subject to an outside force. Thus:

$$\begin{aligned}\text{Conservation of angular momentum} &= M \times V \times R \\ &= \text{constant}\end{aligned}$$

For the Earth, the angular momentum is rotational, with $V = 2\pi\ R_E/P_E$, where P is the Earth's rotation period of one day and the subscript E denotes the Earth. So, we have

$$\text{Earth's rotational angular momentum} = \frac{2\pi\ M_E\ R_E^2}{P_E}$$

Since the length of the Earth's day is increasing as time goes on, the Earth's rotational angular momentum is decreasing by the amount

$$\text{Decrease in rotational angular momentum} = \frac{2\pi\ M_E\ R_E^2\ \Delta\ P_E}{P_E^2}$$

The loss has to be made up by an equivalent gain somewhere else in order to conserve angular momentum. This is done by an increase in the Moon's orbital angular momentum, which is given by:

$$\text{Moon's orbital angular momentum} = M_M \times V_M \times D_M$$

where the subscript M denotes the Moon, M_M is the mass of the Moon, D_M is the distance between the Earth and the Moon, and the orbital velocity of the Moon can be estimated by the escape velocity of the Earth at the Moon's distance, or by

$$V_M = \left(\frac{2G\,M_E}{D_M}\right)^{1/2}$$

where G is the gravitational constant. Substituting this velocity expression into the angular momentum relation, we obtain

$$\text{Moon's orbital angular momentum} = M_M\,D_M(2G\,M_E/\,D_M)^{1/2}$$

Since the mass of the Moon and the Earth do not change, the Moon's distance has to increase by an amount $\Delta\,D_M$ to provide an increase in the angular momentum:

$$\text{Increase in orbital angular momentum} = M_M\ \Delta\ D_M \left(\frac{2G\,M_E}{D_M}\right)^{1/2}.$$

Setting the loss in rotational angular momentum equal to the gain in orbital angular momentum and collecting terms, we obtain

$$\begin{aligned}\Delta\ D_M &= \frac{2\pi\ M_E\ R_E^2\ \Delta\ P_E}{M_M\ P_E^2\left(\frac{2G\,M_E}{D_M}\right)^{1/2}} \\ &\approx 1.3 \times 10^{-9}\text{meters per second} \\ &\approx 0.04\ \text{meters per year}\end{aligned}$$

where $M_E = 5.972 \times 10^{24}$ kilograms, $R_E = 6.371 \times 10^6$ meters, $\Delta P_E = 0.002$ seconds per century $\approx 0.66 \times 10^{-12}$ seconds per second, $M_M = 7.348 \times 10^{22}$ kilograms, $P_E =$ 24 hours $= 8.616 \times 10^4$ seconds, $G = 6.673 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, $D_M = 3.844 \times 10^8$ meters and 1 year $= 3.156 \times 10^7$ seconds. This is the amount measured by sending laser pulses from the Earth to the corner reflectors left on the Moon by astronauts.

hemisphere is tilted toward or away from the Sun. Variation in the Earth's obliquity as small as ±1.3 degrees, around a mean value of 23.3 degrees, may contribute to, or trigger, the ice ages.

The climate forecast for a Moon-less Earth would be a lot bleaker. The gravitational pull of our large Moon acts as an anchor, limiting excursions in the Earth's rotation axis and keeping the climate relatively stable (Fig. 5.31). Without the Moon, the tilt of Earth's rotation axis would vary chaotically between 0 and 85 degrees. Such large variations in the planet's obliquity would result in dramatic changes in climate. With an obliquity of 0 degrees, there would be no seasonal variation in the distribution of sunlight on Earth. At 85 degrees, the Earth's axis would be tipped completely over. The equatorial tropics could then be permanently in cold winter snows, and the poles would be alternately pointed almost directly at or away from the Sun over the course of a single year. Such wide climate changes might be hostile to many forms of life on Earth.

The nearby massive Sun holds the tilt of Mercury and Venus in place, but there is no help for Mars. Located far from the Sun and with two puny satellites, the obliquity of the red planet exhibits wild variations from 0 to

Fig. 5.31 Steadying influence of the Moon on the Earth The brightly colored, sunlit half of the Earth contrasts strongly with the darker, subdued colors of its Moon, which reflects only about one-third as much sunlight as our world. The Moon holds the Earth upright in space, stabilizing its orientation and keeping the planet from tilting over. Without the Moon's influence, chaotic forces could tip the Earth's rotation axis down so far that its poles are pointing at or away from the Sun, producing wild swings in the Earth's climate. This image of the Moon and Earth was taken from a distance of 6.2 billion meters, by the *Galileo* spacecraft on 16 December 1992, soon after swinging around the Earth on its way to Jupiter. (Courtesy of NASA/JPL.)

60 degrees on timescales of about 5 million years, with profound changes in the Martian climate.

5.10 Origin of the Moon

What models of the Moon's origin must explain

There are several facts that must be explained by a successful account of how the Moon originated. Some of them have been known for more than a century, while others result from laboratory investigations of the rocks returned from the Moon.

Any origin theory must, for example, explain why the Earth has a relatively massive Moon when Mercury and Venus have no known moons, and Mars only has two miniscule ones that may be captured asteroids. In other words, our Moon is an unusual event in the formation process of the rocky terrestrial planets.

A satisfactory theory for the origin of the Moon must also explain the Moon's peculiar orbit, which lies neither in the plane of the Earth's orbit around the Sun, the ecliptic plane, nor in the Earth's equatorial plane. Our satellite revolves around the Earth inclined about 5 degrees to the ecliptic plane, which is itself tilted 23.5 degrees with respect to the Earth's equatorial plane.

Perhaps even more important is the mean mass density of the Moon, just 3344 kilograms per cubic meter, much lower than the Earth's mean mass density of 5513 in the same units. The Moon's overall mass density is much closer to the terrestrial mantle than that of the Earth as a whole, which includes its dense iron core.

If we extrapolate the outward motion of the Moon back into the past, assuming a constant rate of 0.04 meters per year, we see that the Moon was 1.8×10^8 meters closer to the Earth 4.6 billion years ago, when the Earth and Moon were formed. That's more than half the current distance to the Moon, of 3.844×10^8 meters, suggesting that the Moon might well have formed near or even out of the Earth in the distant past, for stronger tidal interaction probably propelled the Moon outward at a quicker rate in the past.

Comparison of the lunar rocks to terrestrial rocks provides further constraints on the Moon's parentage (Table 5.8). The oldest rocks on the Moon solidified 4.5 billion years ago, which means that the Moon is about as old as the Earth. One important distinction comes from the similar quantities of oxygen isotopes, or light and heavy oxygen atoms, in Moon rocks and Earth rocks, indicating a close kinship and suggesting a common ancestry, instead of the Moon forming elsewhere and then being

Table 5.8 Constraints on models for the origin of the Moon

Constraint	Implication
No massive moons on other rocky planets	Moon formation is an unusual process
Moon orbit tilted to Earth's equator	Origin process must explain peculiar orbit
Low mean mass density of Moon	Moon has no large iron core
Moon moving away from Earth	Moon once closer to Earth
Some Moon rocks 4.5 billion years old	Moon about as old as Earth
Oxygen isotope ratios	Earth and Moon formed nearby
Depletion of volatiles	Moon formed at high temperature
Enrichment of refractories	Condensation at high temperature
Depletion of metals	Removal of iron prior to formation

captured by the Earth's gravity. Objects formed in other parts of the solar system exhibit different oxygen isotope ratios. This indicates that the Earth and Moon formed in roughly the same part of the primeval solar nebula, unlike all of the meteorites and planetary samples found to date.

A second key constraint is in the compositional differences between the Earth and the Moon. The Moon rocks lack any detectable water-bearing minerals. They are also missing other kinds of volatile elements, with low melting points, that could have been boiled out into space at high temperatures. Relative to Earth rocks, the Moon rocks are also highly depleted in siderophile, or "metal-loving", elements such as cobalt or nickel, which tend to occur in rocks containing iron.

Yet when compared to the Earth our satellite is enriched in non-volatile substances. Called refractories, these elements are the opposite of volatiles; they have high melting points and remain solid at high temperatures and require extraordinary heat to vaporize.

Early origin hypotheses

There are three classical hypotheses for the origin of the Moon that have been advocated for more than a century. They are the fission, capture and accretion models, nicknamed the daughter, pickup and sister theories (Fig. 5.32). But as Sherlock Holmes said in *The Adventure of Silver Blaze*, "I am afraid that whatever theory we state has very grave objections to it." So we will briefly discuss the advantages and flaws of each of these hypotheses, and then move on to the more successful giant impact theory.

The fission hypothesis supposes that the Earth had no satellite in its earliest youth, but that it was once spinning so fast that a large fraction of its mass tore away to create the Moon. If this occurred after the Earth's iron had settled to the center, the Moon would naturally be depleted in metals and would have a low mean mass density characteristic of the outer layers of the Earth. Once the Moon had separated, tidal friction caused it to move slowly away toward its present orbit.

The fission hypothesis does not easily account for the compositional differences between the Moon and the Earth, such as the depletion of volatiles on the Moon. There are also two dynamical difficulties. First, the primordial

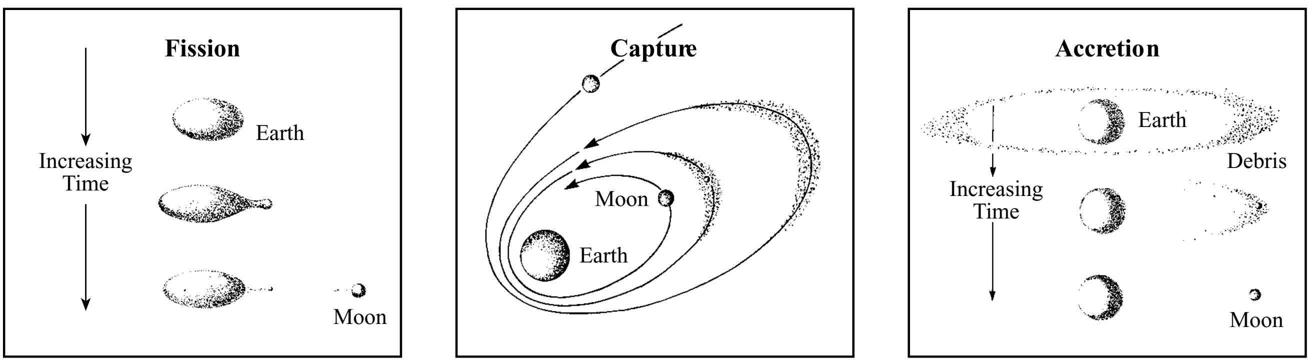


Fig. 5.32 Classical Moon origin hypotheses According to the fission hypothesis (*left*), the rotational speed of the young Earth was great enough for its equatorial bulge to separate from the Earth and become the Moon. In the capture hypothesis (*middle*), a vagabond Moon-sized object once passed close enough to be captured by the Earth's gravitational embrace. We have pictured disruptive capture, with subsequent accretion, but the Moon might have been captured intact. The accretion hypothesis (*right*) asserts that the Moon formed from a disk near the young Earth.

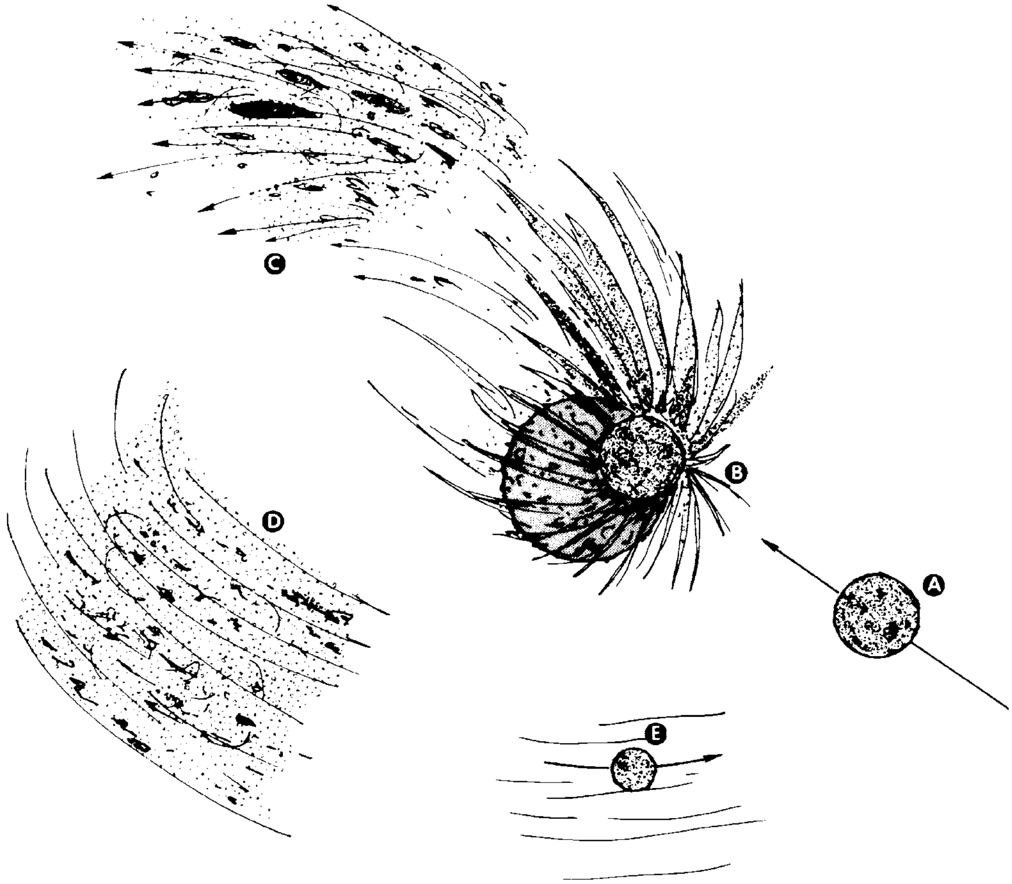


Fig. 5.33 Giant impact hypothesis for the origin of the Earth's Moon According to the giant impact hypothesis, a massive projectile (A), about the size of Mars, struck the young, still-forming Earth (B) in a catastrophic, glancing blow nearly 4.6 billion years ago, resulting in a tremendous explosion and the jetting outward of both projectile and Earth mass. Some fraction of this mass remained in Earth orbit (C), while the rest escaped Earth or impacted again on Earth's surface. A proto-Moon began to form from the orbiting material (D), accreting neighboring matter, and finally became the Moon (E). It may be mostly derived from the crust and mantle of the Earth and/or the impacting object, accounting for the Moon's relatively low mean mass density and lack of iron when compared with the Earth. The Moon accumulated so rapidly that the outer crust was molten, helping to account for the relative lack of water and other volatile elements. Then, as the crust cooled, the newborn Moon swept up the remaining objects nearby, blasting out impact basins and pockmarking the surface with numerous craters. (Courtesy of Alan P. Boss, Carnegie Institution of Washington.)

Earth would have had to rotate exceptionally fast, once every 2.5 hours, if it were to throw off the material that became the Moon. Second, the Moon's orbital plane is tilted from the equatorial plane of the Earth; if the Moon spun off the planet's equator, the two planes ought to be coincident. There are ways around both of these problems, but the fission theory loses its attractive simplicity when it is doctored in these contrived ways.

If the Moon was not plucked out of the Earth, perhaps our satellite is a maverick that formed elsewhere, strayed too near the Earth, and was captured in orbit – either intact or as fragments torn apart by our planet's strong gravity. The principal advantage of this capture hypothesis is that it easily permits compositional differences between the Earth and Moon since they were formed in different locations within the solar system. The main obstacle is understanding how the capture could have taken place. A passing body would either collide with the Earth or receive a gravitational boost that would hurl it away from the Earth in slingshot fashion. In order to go into orbit about the Earth, the approaching Moon would have to slow down, and the chances of this happening are exceptionally low.

The third classical hypothesis suggests that the Moon and Earth formed concurrently from a cloud of gas and dust through a process not unlike the probable formation of the planets around the Sun. The raw materials for the Moon came from a disk of material in orbit around the Earth, and the planets originated in a similar disk orbiting the Sun.

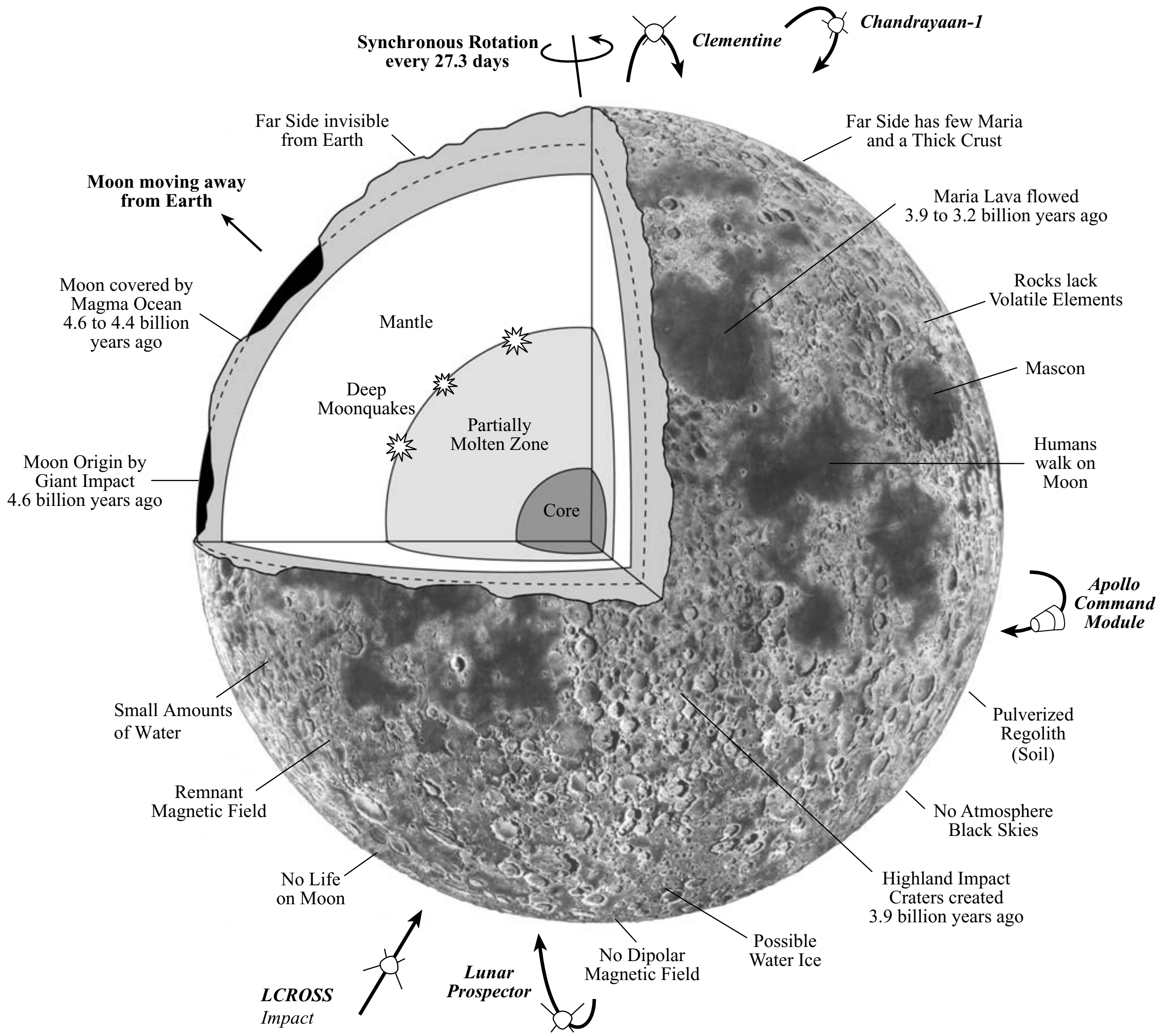


Fig. 5.34 Summary diagram

Such a model seems to apply nicely to giant gaseous planets, such as Jupiter, that have families of satellites resembling the solar system. But where does that leave the rocky terrestrial planets that have no known moons, Mercury and Venus? And what about Mars, with only two tiny satellites? If the process that formed our massive Moon is the natural way of things, we have difficulty understanding these other planets. And why should the chemistry of the Earth and Moon be so different, and how were the volatile elements driven out of the Moon if it always orbited the Earth? Special assumptions can help extricate the accretion theory of its difficulties, but it also loses its appeal when these special assumptions are introduced.

The giant impact hypothesis

Today many astronomers favor a hybrid of the fission and capture theories, with a violent and catastrophic lunar birth. According to this newer, giant impact hypothesis, a massive rogue projectile, perhaps 2.5 to 3.0 times the mass of Mars, sideswiped the Earth and dislodged the material that would become the Moon. The collision may have knocked the Earth away from its original upright position, giving the planet its current axial tilt of 23.5 degrees.

This glancing, planet-shattering blow occurred almost 4.6 billion years ago, during a heavy bombardment that marked the last stages of the solar system's formative period. At this time, iron had already sunk to the core of the Earth, and a rocky crust was beginning to congeal around the partially molten planet.

The giant impact mechanism permits the Moon to form initially in the same part of the solar system as the Earth and to undergo a process that explains both the dearth of metals and volatile elements, as well as an enrichment of refractory elements, before solidification. It might have shattered the colliding object to smithereens and vaporized parts of the iron-poor upper layers of the Earth, blasting off a mix of terrestrial and impactor material into orbit where it coalesced within about a century to form the Moon (Fig. 5.33). The intense heat of the collision vaporized water and most volatile materials from the cast-off material, which clumped together and reassembled into the orbiting Moon.

The glancing blow would have knocked the collision debris into the Moon's current tilted orbit, and if the satellite included material in the mantle of the impactor, this could also help explain the compositional differences between the Moon and Earth. The searing heat of such a collision would explain why the Moon holds no appreciable amounts of water and few volatile elements. All were boiled away. The new hypothesis therefore seems to explain the facts with a minimum of assumptions. As one astronomer stated, "it requires no magic, no special pleading, no extra twiddling and no *deus ex machina*".

Another reason that this model became at least imaginable was the discovery of extremely large impact basins on the Moon, such as the South Pole – Aitken basin on the far side. It was a short mental leap from very large impact basins to a planetary collision.

Thus, exploration of the Moon has resulted in the probable solution of the ancient mystery of the Moon's origin. It was most likely born in a fiery cataclysm out of the infant Earth, the result of an enormous, off-center collision during the early days of the solar system when such events were common.

The giant impact that gave rise to the Moon is a natural consequence of planet formation, making astronomers more aware of impact catastrophes in solar system history. Similar collisions with larger or smaller projectiles could explain major planetary anomalies, such as the removal of Mercury's low-density mantle, the off-kilter, backward spin of Venus, the global crustal dichotomy of Mars, the planet Uranus' bizarre, sideways orientation, and even the demise of the dinosaurs that redirected the course of life on Earth 65 million years ago.

The voyage to the Moon became the stepping-stone to outer space, resembling the first, tentative steps of a child testing the water before learning to swim. It opened a path to the rest of the solar system, to an ongoing, close-up exploration of the planets and their satellites, which have been mapped and surveyed with a detail surpassing that of most countries on our home planet Earth.

Our satellite was the first port of call in our captivating voyage to the planets, to which we now turn, beginning with Mercury whose composition probably resulted from a giant impact of its own.

6 Mercury: a dense battered world

- Because of its close proximity to the Sun, the innermost planet Mercury cannot be studied from Earth against the dark night sky; many astronomers and most people have never seen the elusive planet.
- During the daytime, Mercury's ground temperature reaches 740 kelvin, hot enough to vaporize water and melt lead; at night it plunges to a freezing 90 kelvin.
- Although Mercury is one of the Earth's nearest planetary neighbors, only two spacecraft have ventured near Mercury. They are the *Mariner 10* spacecraft in 1974–75 and the *MErcury Surface, Space ENvironment, GEochemistry, and Ranging* (*MESSENGER*) spacecraft in 2008–11.
- There is a simple three-to-two resonance between Mercury's rotation period of 58.646 Earth days and its orbital year of 87.969 Earth days. This spin–orbit coupling is produced by solar tides in the solid planet.
- The interval from sunrise to sunset at a given location on Mercury is 87.969 Earth days, and the night lasts 87.969 Earth days more, so the day on Mercury lasts 175.938 Earth days and is twice Mercury's year.
- Mercury's rotation axis is aligned perpendicular to its orbital plane, so there are no seasons on the planet, and its polar regions never receive the direct rays of sunlight. Radar echoes suggest that water ice may reside in permanently shaded regions within deep craters near Mercury's poles.
- Mercury has highland craters and impact basins that resemble those found on the Moon. The craters and basins on both objects were most likely formed during a late heavy bombardment by meteorites 3.9 billion years ago.
- An ancient period of volcanic flow, during the late heavy bombardment of Mercury 3.9 billion years ago, obliterated small craters, partially filled larger craters, and created intercrater plains that are not found on the Moon.
- Smooth volcanic plains have filled old craters and impact basins after they formed, and covered approximately 40 percent of the surface of Mercury.

- Irregularly shaped depressions surrounded by bright material have been attributed to volcanic vents on Mercury.
- Long, winding cliffs, or rupes, are found on Mercury, and not on the Moon. They are attributed to the contraction of the young planet as it cooled.
- Relative to its size, Mercury has the biggest iron core of all the terrestrial planets. Mercury's core is much larger than the core of the Moon.
- Mercury may have been blown apart by an ancient collision with a planet-sized object, removing its low-density rocky mantle.
- Mercury has a dipolar magnetic field with a magnetic axis closely aligned with the planet's rotation axis.
- The magnetosphere of Mercury can be opened on its dayside by magnetic reconnection during interaction with the magnetic fields emanating from the Sun.
- Rotational twists discovered by radar observations of Mercury suggest that it has a liquid core in which the planet's magnetic field might be generated.
- More than a century ago, astronomers found that Mercury did not appear in its expected place, leading Einstein to develop a new theory of gravity in which the Sun curves nearby space.

6.1 Fundamentals

Mercury is the smallest of the four rocky, terrestrial planets. It is the planet closest to the Sun, moving around it with the fastest speed and shortest year of any planet. Mercury spins with a slow rotation period of 58.646 Earth days, just two-thirds of its year. It has practically no atmosphere at all, but retains an unexpectedly strong magnetic field.

6.2 A tiny world in the glare of sunlight

Mercury revolves closer to the Sun than any other known planet, with a mean distance from the Sun of just 0.3871 AU. It also has the most eccentric orbit of any major planet, with a distance varying from 0.3072 AU at perihelion, its closest approach to the Sun, to 0.4667 AU at aphelion, when it is furthest from the Sun.

Table 6.1 Physical properties of Mercury[a]

Mass	3.301×10^{23} kilograms $= 0.0553\ M_{\mathrm{E}}$
Mean radius	2439.7 kilometers $= 0.382\ R_{\mathrm{E}}$
Bulk density	5427 kilograms per cubic meter
Sidereal rotation period	58.6462 Earth days
Sidereal orbital period	87.97 Earth days = 0.240 846 7 Earth years
Mean distance from Sun	5.79×10^{10} meters = 0.387 AU
Age	4.6×10^{9} years
Exosphere	Hydrogen, helium, sodium, potassium atoms
Surface pressure (exosphere)	10^{-12} bars
Surface temperature	90 to 740 kelvin
Magnetic field strength	0.0033×10^{-4} tesla at the equator $= 0.01\ B_{\mathrm{E}}$
Magnetic dipole moment	5.54×10^{13} tesla meters cubed

[a] The symbols M_{E}, R_{E} and B_{E} respectively denote the mass, radius and magnetic field strength of the Earth.

Mercury also has the shortest year – about 88 Earth days – and the highest orbital speed of any planet. Like a moth about a flame, Mercury races around the Sun at an average speed of 48 kilometers per second. Its rapid motion explains why Mercury is named after the wing-footed messenger of the gods in Roman mythology.

Any planet so close to the Sun is subject to intense sunlight, and therefore has to become very hot. Mercury's maximum daytime surface temperature, when closest to the Sun at perihelion and on the equator, is 740 kelvin. This is hot enough to melt tin, lead and even zinc. Because there is no atmosphere to hold in the heat, the surface temperature plummets to about 90 kelvin when the sunlit side rotates into the Sun's shadow during the planet's long night.

As planets go, Mercury is a tiny world, with the smallest size of any terrestrial planet and slightly smaller than Jupiter's moon Ganymede and Saturn's moon Titan. Mercury's linear radius is easy to measure from its angular radius and distance. Its radius is 2439.7 kilometers or about 1.4 times the radius of the Earth's Moon. Mercury's mass has been a more elusive quantity to determine, because the planet has no satellites. The mass was first estimated from Mercury's gravitational influence on the orbital motion of Venus and passing comets, and improved by the planet's gravitational deflection of the space probes *Mariner 10* and *MESSENGER*. The planet's mass is 3.301×10^{23} kilograms.

Mercury is surprisingly massive for its size. Its volume is only slightly larger than the Moon's and yet it has four times the Moon's mass. This implies a mean mass density of 5427 kilograms per cubic meter, which is nearly as high as that of the Earth, 5513 in the same units, and a little more than Venus at 5243.

The planet's small apparent size and its proximity to the Sun make it difficult to see from Earth. The innermost planet never wanders more than 27.7 degrees in angular separation from the Sun (Fig. 6.1). This angle is less than that made by the hands of a watch at one o'clock. From Earth's perspective, Mercury's tight orbit never reaches into the dark night sky, and it can thus be observed only during the day.

Mercury is visible to the unaided eye only in twilight when it is low in the sky and must be seen through a thick layer of air. Astronomers have therefore taken to observing Mercury near mid-day, when it is far from the horizon and can be seen through a relatively thin layer of air. At such times, Mercury can only be seen with telescopes.

But the terrestrial atmosphere limits the resolution of even the best Earth-based telescopes to features on Mercury that are a few hundred kilometers across or wider – a resolution far worse than that for the Moon with the unaided eye. Moreover, space-borne telescopes with better resolution, such as the *Hubble Space Telescope*, cannot point at Mercury, because even stray light from the nearby Sun could damage their sensitive instruments.

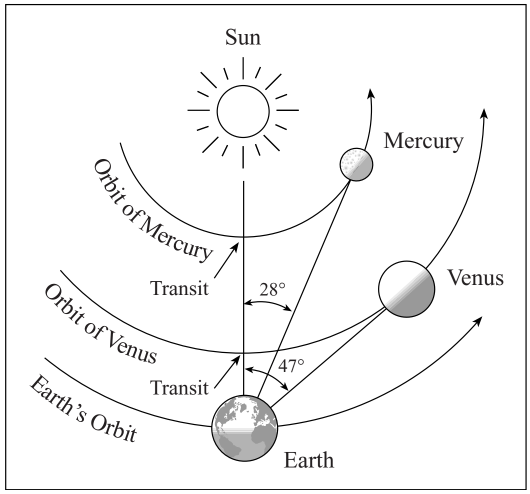


Fig. 6.1 Maximum elongation Unlike the planets that orbit the Sun beyond the Earth, Mercury and Venus can never be seen high in the sky in the dead of night. Because they are close to the Sun and inside Earth's orbit, these planets are always seen soon after sunset or shortly before sunrise, and they show phases much as the Moon does. The elongation of Mercury and Venus, or their angular distances from the Sun as viewed from the Earth, never exceeds 28 degrees and 47 degrees, respectively.

Most people have never seen Mercury. Even Copernicus complained that it eluded him. So it is not surprising that little was known about this enigmatic planet until it was probed by terrestrial radar (short for "radio detection and ranging"), and scrutinized during close encounters with two spacecraft.

6.3 Space-age investigations of Mercury

Close-up scrutiny of Mercury was first carried out from the *Mariner 10* spacecraft, which was launched on 3 November 1973, and obtained data during its three flybys of the planet on 29 March and 21 September 1974 and 16 March 1975. Instruments aboard *Mariner 10* imaged about 45 percent of the surface at an average resolution of 1 kilometer, and indicated that the planet retains an unexpectedly strong magnetic field. As far as we know, *Mariner 10* is still moving by Mercury every six months, but without obtaining any observational data.

Ground-based radar observations had already demonstrated in 1965 that Mercury spins with a slow rotation period of 58.646 Earth days, just two-thirds of Mercury's year of 87.97 Earth days. Detailed radar investigations of the rotation in 2007 provided new information about the planet's liquid core, and radar echoes from Mercury's polar regions indicated that water ice might be trapped in its permanently shadowed craters.

On 3 August 2004, the *MESSENGER* spacecraft was launched to explore Mercury close-up for the first time in more than 30 years. The name *MESSENGER* is an acronym of *MErcury, Surface, Space ENvironment, GEochemistry, and Ranging*. The name also reminds us that to the ancients Mercury was the messenger of the gods.

After an Earth flyby and two Venus flybys, *MESSENGER* passed nearby Mercury on three occasions, on 14 January and 6 October 2008 and 29 September 2009. During the first flyby, its instruments imaged 20 percent of Mercury's surface not previously seen by spacecraft, and it imaged an additional unseen 30 percent during the second flyby. Measurements of the planet's magnetic field, exosphere, surface composition, and gravitational field were also made.

After seven years and six planetary encounters, *MESSENGER* will have slowed enough to insert it into orbit around Mercury, on 17–18 March 2011, using a conventional retro-rocket. The spacecraft would have been traveling at such a high speed during a direct flight, such as the four-month *Mariner 10* trip, that it could not have been directed into orbit around the planet.

6.4 Radar probes of Mercury

The halting spin of old age

Astronomers once supposed that solar tides in the body of Mercury would cause the planet to rotate on its axis once every 87.97 Earth days, in step with its sidereal orbital period. (The sidereal rotation or orbital period is the time taken to complete one rotation or orbit relative to the "fixed" stars.) Just as the Moon always presents the same face to the Earth, it was thought that one side of Mercury was always turned toward the Sun. To test this idea, the Italian astronomer Giovanni Schiaparelli (1835–1910) monitored Mercury's surface markings seen though his 0.46-meter telescope, and he concluded that the same side of the planet did, indeed, always face the Sun. For three-quarters of a century, telescopic observers agreed with his conclusion. All of these astronomers were wrong!

In 1965, Mercury's true rotational period was determined with radio signals that rebounded from the planet. The world's largest radio telescope, located in Arecibo, Puerto Rico, was used to transmit megawatts of pulsed radio power at Mercury, and to receive the faint echo (Fig. 6.2). This technique is known as radar, and it is also used to locate and guide airplanes near airports.

Each pulse was finely tuned, with a narrow range of wavelengths. Upon hitting the planet, its rotation de-tuned the pulse, slightly spreading the range of wavelengths (Fig. 6.3). One side of the globe was rotating away from the Earth, while the other side was rotating toward our planet. These motions produced slight changes in the wavelength of the echo; from these changes, the speed of the surface and the rotational period were calculated, using the well-known expression for the Doppler effect (Focus 6.1; Fig. 6.4).

The result came as an unexpected surprise. The sidereal rotation period was 58.646 Earth days, or exactly two-thirds of the 87.969-day period that had been accepted so long. Thus, with respect to the stellar background, Mercury spins on its axis three times during two full revolutions about the Sun, so we say that there is a 3-to-2 spin–orbit resonance in Mercury's rotation. This relationship follows from $3 \times 58.646 = 2 \times 87.969$, and it is technically known as spin–orbit coupling. In comparison, the Moon has a one-to-one spin–orbit resonance in which its rotation period is equal to its orbital period.

Cause of Mercury's spin–orbit resonance

But why are Mercury's day and year related to each other in such a simple 3-to-2 ratio? The answer lies in the Sun's varying tidal forces as Mercury revolves about its elongated orbit. The solar gravity pulls hardest on Mercury when the planet is closest to the Sun, at perihelion, and least at the opposite side of its eccentric orbit, at aphelion. This extra gravitational pull of the Sun at perihelion gives an abrupt twist to Mercury's non-spherical body, speeding up the rotation rate and forcing it into synchronism at perihelion with the 3-to-2 resonance. If Mercury's orbit around the Sun were much closer to a circular shape, like the nearly round orbit of the Moon around the Earth, then the Sun's tidal forces would have slowed Mercury's rotation into synchronism with its orbital motion, in a one-to-one resonance with a rotation period equal to its 87.969-day orbital period.

Long, hot afternoons on Mercury

The days are certainly long on Mercury, longer than the planet's year. Mercury's solar day, the time from sunrise to sunset and back to sunrise, is two Mercury years long (Fig. 6.5). At any given point in its orbit, the same hemisphere does not face the Sun during each orbit, but during every other orbit.

Any markings at a given location on Mercury's surface will return to the sunlit side after two orbital revolutions, and this may have misled early astronomers. After two of

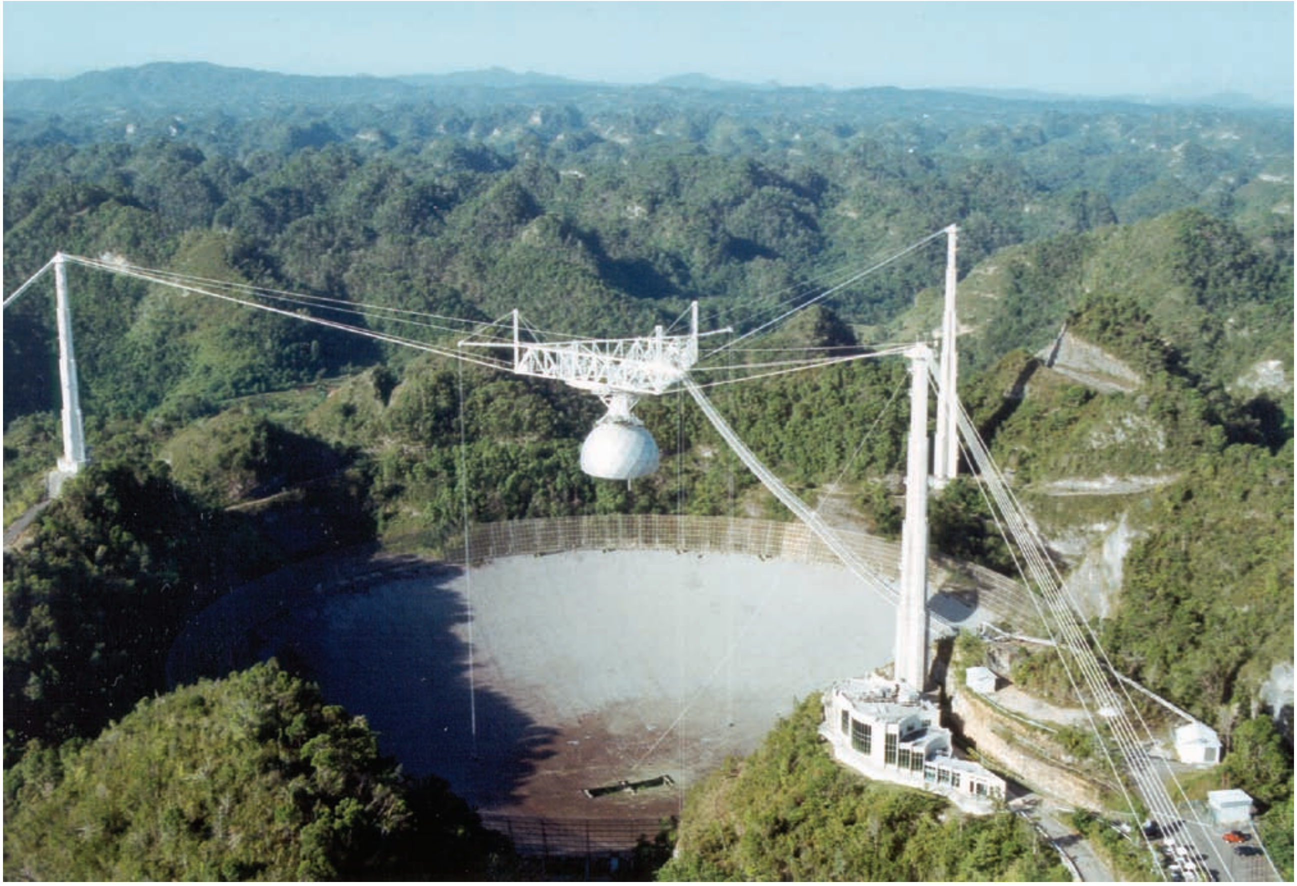


Fig. 6.2 Arecibo Observatory The world's largest radio telescope is nestled into the hills near Arecibo, Puerto Rico. Its metal reflecting surface has a spherical shape with a diameter of 305 meters. The reflected radio signals are focused to detectors suspended on the triangular structure hanging from the three towers. The facility can also transmit powerful radio pulses, sending them off the metal surface into space. Such pulses, sent and received from this giant telescope, first measured Mercury's rotation period in 1965. Until then it had been wrongly thought that Mercury kept one side permanently facing the Sun, with a rotation period equal to its orbital period.

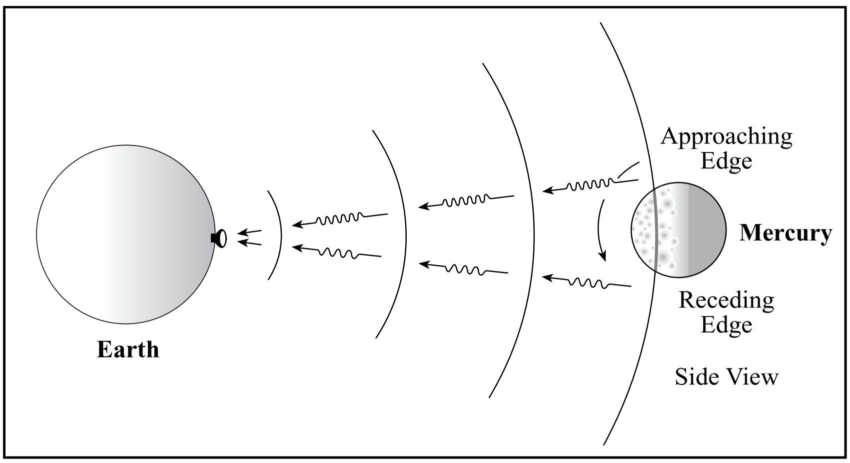


Fig. 6.3 Radar probes of Mercury A radio signal spreads out as a spherical wave, and Mercury intercepts only a small fraction of them. As the wave sweeps by the planet, it is reflected in spherical wavelets whose wavelengths are Doppler-shifted by the rotational motion of Mercury's surface. The waves from the receding side are red-shifted towards longer wavelengths and those from the approaching side are blue-shifted to shorter wavelengths. The total amount of wavelength change, from red to blue, reveals the speed of rotation, and the rotation period can be obtained by dividing the planet's circumference by this speed.

Focus 6.1 The Doppler effect

Just as a source of sound can vary in pitch or wavelength, depending on its motion, the wavelength of electromagnetic radiation shifts when the emitting source moves with respect to the observer. This Doppler shift is named after the Austrian scientist, mathematician and schoolteacher Christian Doppler (1803–1853), who discovered it in 1842. If the motion is toward the observer, the shift is to shorter wavelengths, and when the motion is away the wavelength becomes longer (Fig. 6.4). You notice the effect when listening to the changing pitch of a passing ambulance siren. The tone of the siren is higher while the ambulance approaches you and lower when it moves away from you.

If the radiation is emitted at a specific wavelength, λ_{emitted}, by a source at rest, the wavelength, $\lambda_{\text{observed}}$, observed from a moving source is given by the relation

$$\text{Redshift} = z = \frac{\lambda_{\text{observed}} - \lambda_{\text{emitted}}}{\lambda_{\text{emitted}}} = \frac{V_r}{c}$$

where V_r is the radial velocity of the source along the line of sight away from the observer, and c is the velocity of light, $c = 2.9979 \times 10^8$ meters per second. The parameter z is called the redshift since the Doppler shift is toward the longer, redder wavelengths in the visible part of the electromagnetic spectrum. When the motion is toward the observer, V_r is negative and there is a blue shift to shorter, bluer wavelengths.

A rotating object will produce a blueshift on the side spinning toward an observer, and a redshift on the opposite side. Their combined effect will broaden a finely tuned radio pulse at wavelength λ by an amount $\Delta\lambda$ given by the expression

$$\frac{\Delta\lambda}{\lambda} = \frac{V_{\text{rot}}}{c}$$

where V_{rot} denotes the rotation velocity.

its orbital periods they would see the same markings on the sunlit side and would find no disagreement with the 87.969-day period that they expected, probably ignoring or missing conflicting observations during the intervening orbit.

Possible water ice at the poles of Mercury

Radar astronomers have found evidence for water ice in both the north and south polar regions of Mercury. Both the intensity and orientation, or polarization, of the bright radar echoes suggest the presence of highly reflective water

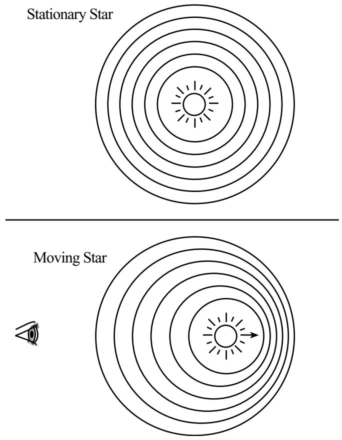


Fig. 6.4 Doppler effect A stationary source of radiation (*top*) emits regularly spaced light waves that get stretched out or scrunched up if the sources moves (*bottom*). Here we show a star moving away (*bottom right*) from the observer (*bottom left*). The stretching of light waves that occurs when the source moves away from an observer along the line of sight is called a redshift, because red light waves are relatively long visible light waves; the compression of light waves that occurs when the source moves along the line of sight toward an observer is called a blueshift, because blue light waves are relatively short. The wavelength change, from the stationary to moving condition, is called the Doppler shift, and its size provides a measurement of radial velocity, or the velocity of the component of the source's motion along the line of sight. The Doppler effect is named after the Austrian physicist Christian Doppler (1803–1853), who first considered it in 1842.

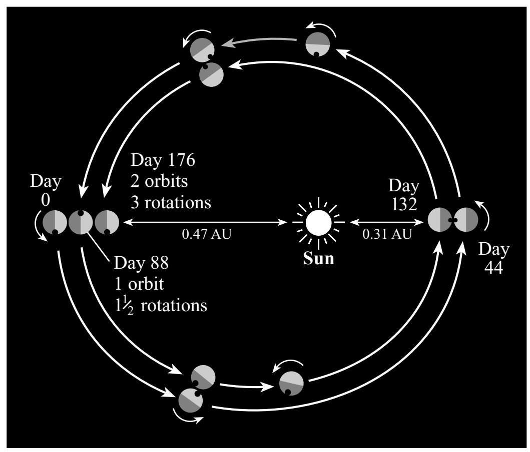
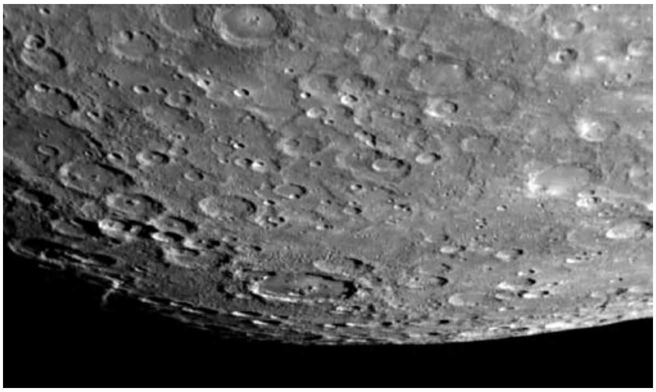


Fig. 6.5 The days are long on Mercury Because of its spin–orbit coupling, Mercury rotates once every 58.6 Earth days, and orbits the Sun in 88.0 Earth days. After two orbits the planet has rotated three times, but from Mercury's surface the Sun appears to have moved only once around the planet. So sunrise is repeated at a given point on the planet's surface (*black dot*) every two orbits, and a solar day on Mercury lasts two of the planet's years and 176 Earth days. The labels on this figure refer to Earth days.



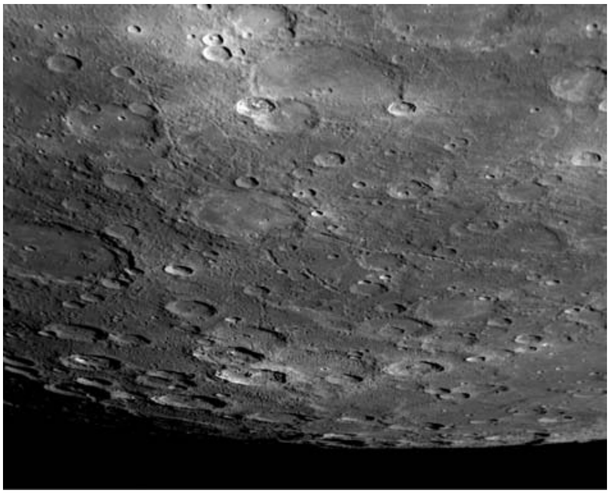


Fig. 6.6 Mercury's south pole During its first flyby, the *MESSENGER* spacecraft obtained high-resolution images of a side of Mercury's south polar region not previously seen by spacecraft (*top*), while the opposite side, which was previously seen from *Mariner 10*, was imaged during *MESSENGER*'s second flyby (*bottom*). Radar-reflection data suggest that water ice might be present in the cold, dark, permanently shadowed floors of craters in the polar regions of Mercury. (Courtesy of NASA/JHUAPL/CIW.)

ice. The polar radar characteristics are similar to those returned from the large ice-covered moons of Jupiter and the residual water-ice cap of Mars.

The prospect of a planet so close to the Sun having any water ice seems preposterous. Yet the ice could reside at the bottoms of polar craters (Fig. 6.6). Mercury's rotation axis is very nearly perpendicular to the plane of its orbit, with an obliquity of zero degrees, so the planet does not experience seasons and its polar regions never receive the direct rays of sunlight. In other words, the planet's rotation axis does not tilt, so the Sun never gets very high in the sky at Mercury's poles. Moreover, the atmosphere of Mercury is so thin that heat from the Sun-baked areas does not spread to colder ones. That is why the temperatures during Mercury's night are so cold, down to 90 kelvin, in spite of long dayside highs of about 700 kelvin.

The floors of deep craters near the poles are never exposed to sunlight and ought to always be colder than about 112 kelvin, well below the freezing point of water at 273 kelvin. Moreover, the radar-bright features are concentrated in specific, fresh-looking polar craters with high rims and permanently shadowed floors; the bottoms of degraded, shallow craters are not in perpetual shadow and do not exhibit bright radar signals.

The water could have been delivered to the crater floors by the impacts of comets, subsequently remaining cold-trapped within the craters for as long as billions of years, depending on when the collision occurred.

The shaded polar craters on Mercury might nevertheless contain other volatile substances, such as sulfur, which could produce strong radar echoes but have a higher melting temperature than water ice. So, we may not definitely know if there is water ice at the top and bottom of Mercury until inquisitive robot spacecraft land there and make the appropriate tests. In the meantime, we turn to the startling results of the *Mariner 10* and *MESSENGER* spacecraft.

6.5 A modified Moon-like surface

At first glance, Mercury closely resembles the Moon, for both worlds are small, heavily cratered, and without a significant atmosphere to cause erosion. They both contain large impact basins, ubiquitous craters upon craters, and vast smoother places attributed to volcanic flows. Yet there are differences, for volcanism on Mercury differs from the lunar variety, and the planet exhibits long cliffs that traverse the surface for hundreds of kilometers and are not found on the Moon.

Craters on Mercury

In 1974–75 the *Mariner 10* spacecraft penetrated the glare surrounding Mercury, providing images of about half of Mercury's surface, and showing features thousands of times smaller than those seen from the best telescopes on Earth. These close-ups revealed a landscape that had never been seen before (Fig. 6.7), and they gave us a glimpse of the planet's past. Then in 2008–09 the *MESSENGER* spacecraft returned fascinating images taken at close range (Fig. 6.8) for most of the other half of Mercury that had not been seen before.

Like the Moon, the planet Mercury has highlands that are pockmarked with impact craters, ranging in diameter from impact basins 1000 kilometers across to craters only 100 meters in diameter. As on the Moon, there are small bowl-shaped craters, larger craters with terraces and central peaks, relatively young craters with bright rays, and huge impact basins on Mercury. Once identified, the

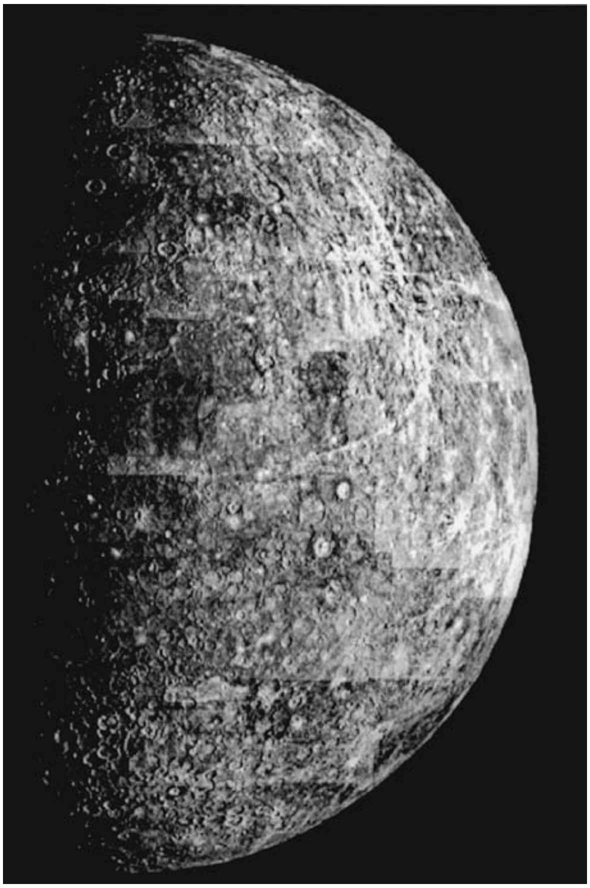


Fig. 6.7 Moon-like surface of Mercury After passing the dark side of Mercury, the *Mariner 10* spacecraft looked back at the sunlit hemisphere of the planet and photographed these images that have been assembled into a mosaic. It shows large tracts of smooth plains, which may be due to extensive volcanism. Partially visible along the day–night terminator (*left*) is half of the Caloris basin (*above center toward the top*), a gigantic multi-ringed impact scar. (Courtesy of NASA/JPL.)

Focus 6.2 Names of surface features on Mercury

The impact craters on Mercury have been mostly named after famous deceased authors, artists, and musicians, such as Dickens, Matisse or Michelangelo, and Mozart. The biggest ones, in order of decreasing size, are named Rembrandt, Beethoven, Dostoevskij, Tolstoj, Goethe, Shakespeare, Raphael and Homer; they range from 720 to 314 kilometers in diameter.

New craters found in *MESSENGER* images and approved by the International Astronomical Union (IAU) include, for example, Calvino, Gibran, Hemingway and Poe, after the authors Italo Calvino, Kahil Gibran, Ernest Hemingway, and Edgar Allan Poe.

Cliffs (*rupes* – Latin for "cliff") take the names of famous ships, such as the British naval vessel *Beagle*, which the naturalist Charles Darwin (1809–1882) traveled on, or *Discovery*, Captain James Cook's (1728–1779) ship on his last voyage to the Pacific, and *Santa Maria*, Christopher Columbus' (1451–1506) flagship on his first voyage to America in 1492. Plains, or plaitiae, take the names of Mercury, as a planet or god, in various cultures.

The most prominent impact basin viewed by *Mariner 10* has a unique designation, the Caloris Basin or Basin of Heat, so-named because it nearly coincides with one of the hottest locations on the planet.

A complete description of all the naming details can be found in the "Gazetteer of Planetary Nomenclature" of the IAU Working Group for Planetary System Nomenclature, available on the Internet at http://planetarynames.wr.usgs.gov/.

craters are mostly named after famous writers, painters, and composers (Focus 6.2).

The ubiquitous craters on Mercury strongly resemble their lunar counterparts, indicating that they were formed by meteoritic impact, and most of the Mercurian craters probably record the period of late heavy bombardment by meteorites of all size, which formed the lunar highlands about 3.9 billion years ago.

Because the force of gravity on the surface of Mercury is about twice that on the Moon, material ejected from a crater on Mercury is thrown about half as far. Thus, secondary craters are closer to the primary crater rim on Mercury than they are on the Moon. The comparatively high gravity on Mercury also helps flatten crater walls, perhaps explaining why Mercury's craters appear to be shallower than similarly sized craters on the Moon (Fig. 6.9).

The freshest craters on Mercury have extensive ray systems, some of which extend over 1000 kilometers (Fig. 6.10). The bright rays are thrown out during a crater-forming explosion when the meteorite impact occurs on the surface of an airless body like Mercury or the Moon. But the rays fade with time as tiny meteorites and particles from the solar wind strike the surface and darken the rays. So the rayed craters were formed relatively recently, on the order of 1 billion years ago or less.

Some craters on Mercury have dark rims or nearby dark "halos" surrounding them (Fig. 6.11). There are two possible explanations for these features. A subsurface layer of dark material might have been excavated when the crater-producing impact penetrated to the right depth. Alternatively, the heat of impact might have melted some of the surface, and the molten rock splashed to the edge of the crater where it re-solidified as a dark, glassy substance.

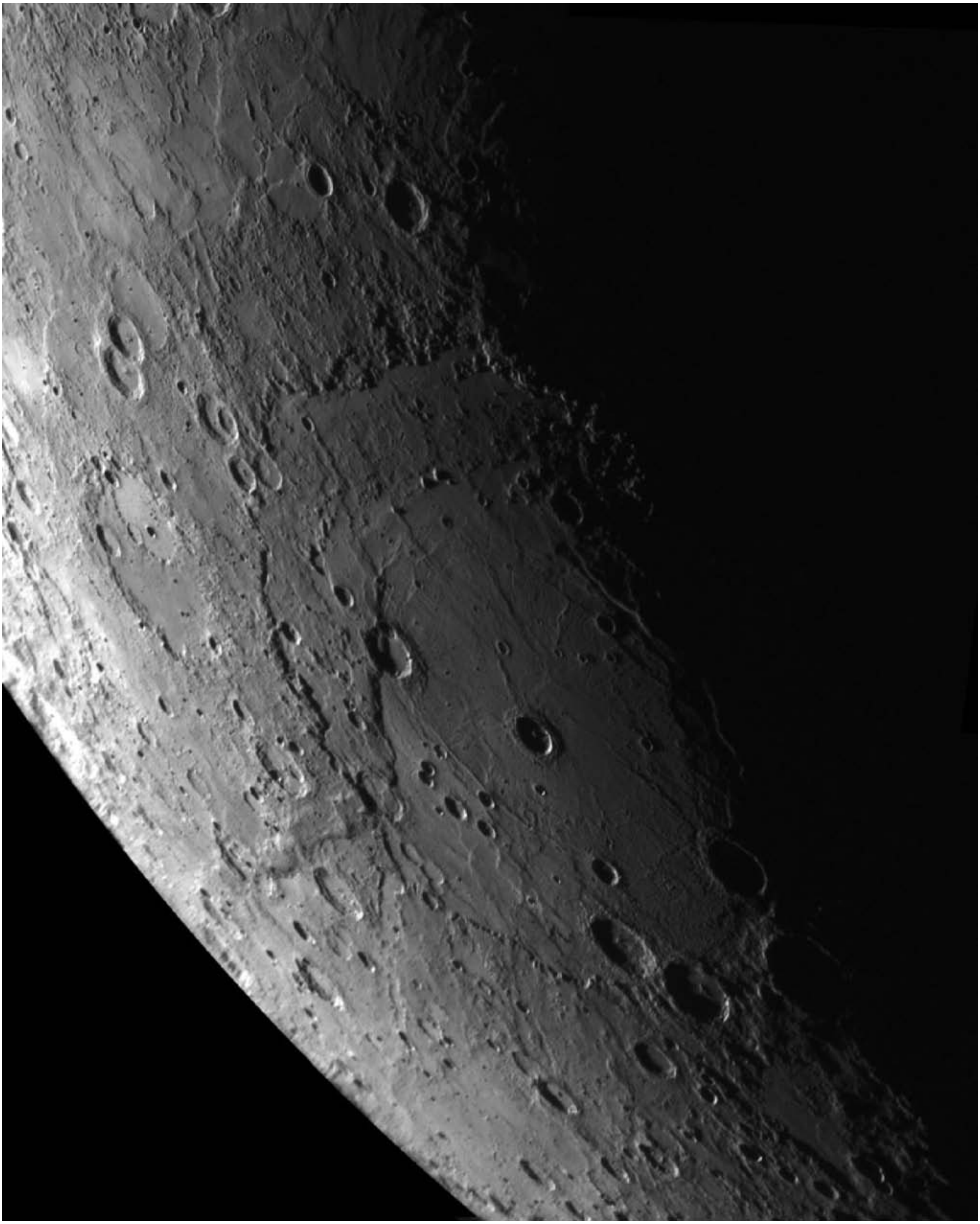


Fig. 6.8 Previously unseen side of Mercury A mosaic of images collected as the *MESSENGER* spacecraft viewed Mercury obliquely on its second approach on 6 October 2008. The Rembrandt impact basin, 715 kilometers in diameter and also shown in Fig. 6.14, is seen at the center as night was falling across the planet's eastern (right) edge. (Courtesy of NASA/JHUAPL/CIW.)

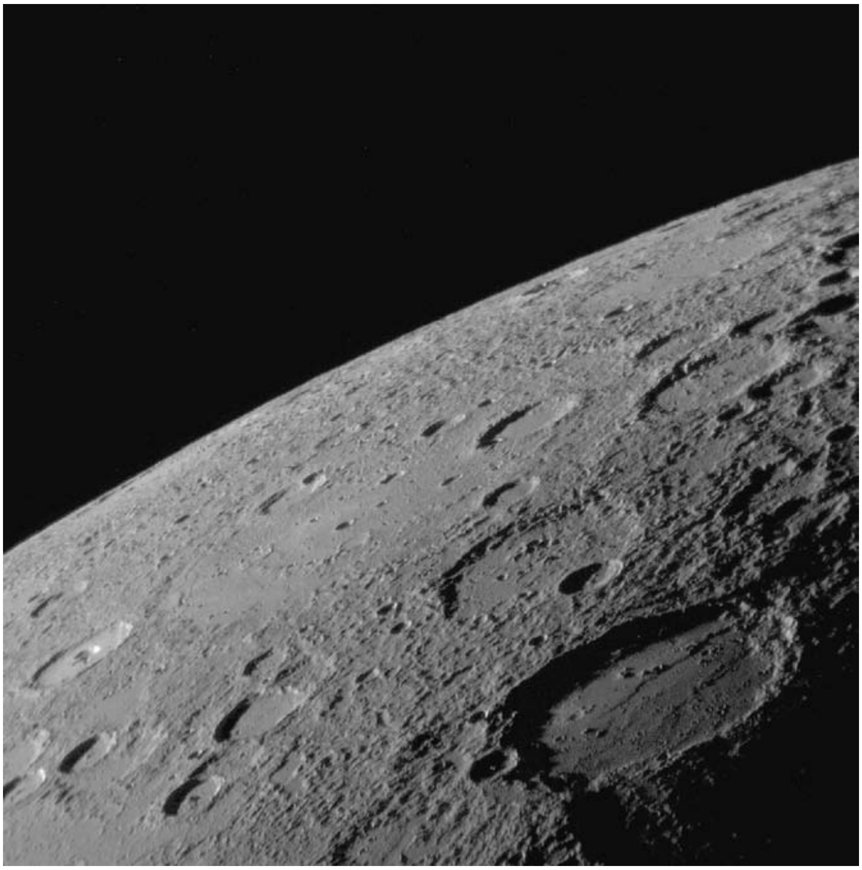


Fig. 6.9 Shallow craters on Mercury
This mosaic image was acquired from the *MESSENGER* spacecraft during its first flyby of Mercury. Many of the impact craters appear to be shallower than similarly sized craters on the Moon. The comparatively high gravity of Mercury helps flatten tall structures like crater walls. Smooth volcanic plains are seen near the center of the image. The large crater to the lower right is 200 kilometers wide. The shadowed area to the right of this crater marks the boundary between the sunlit day side and dark night side of the planet. (Courtesy of NASA/JHUAPL/CIW.)

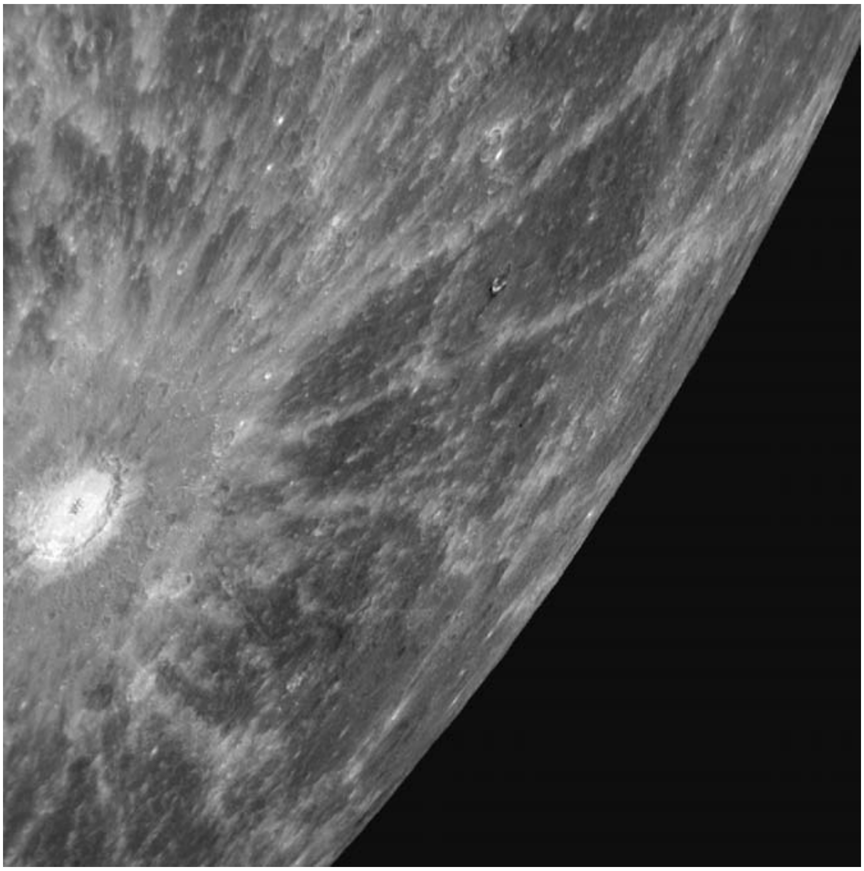


Fig. 6.10 Bright rayed crater on Mercury A relatively young crater, about 80 kilometers in diameter, has splashed bright rays far around Mercury. The bright rays are produced by impacts that excavate and eject relatively unweathered subsurface material. Space-weathering effects, such as the continued bombardment by small meteorites or solar-wind particles, will eventually darken the ejected material and erase the rays from view. Bright-rayed craters are also found on the Moon, and *MESSENGER* images, such as this one, indicate that Mercury has more of them, perhaps due to more violent impacts caused by Mercury's greater mass and proximity to the Sun. (Courtesy of NASA/JHUAPL/CIW.)

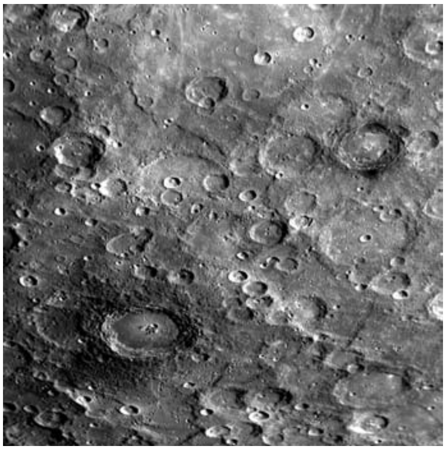


Fig 6.11 Dark halo craters on Mercury Two of the larger craters in this *MESSENGER* image appear to have darkened crater rims and partial "halos" of dark material immediately surrounding them. The explosive impact that produced the dark-haloed craters could have excavated darker subsurface material or melted part of the surface, splashing it across the surface where it re-solidified as dark material. (Courtesy of NASA/JHUAPL/CIW.)

The Moon also has dark-haloed craters; Tycho is a well-known example. In the lower gravity of the Moon, dark material will be ejected a greater distance from the crater than on Mercury.

The heavily cratered terrain on Mercury also exhibits a paucity of smaller craters when compared to the lunar highlands, most likely due to volcanic flows during the late heavy bombardment that obliterated the small craters on Mercury, but not in the lunar highlands. Mercury's craters are also less densely packed than their counterparts on the Moon, possibly due to volcanism that obscured the older craters on Mercury and created extensive intercrater plains not found in the lunar highlands.

Old intercrater plains on Mercury

In many important respects, Mercury's resemblance to the Moon is superficial. The planet's most densely cratered surfaces are not as heavily cratered as the lunar highlands, and Mercury does not contain regions of overlapping large craters and basins. Also unlike the Moon, the heavily cratered terrain on Mercury includes large regions of gently rolling, intercrater plains that were discovered in *Mariner 10* images (Fig. 6.12).

These plains obliterate or partially fill older craters, but the plains have other craters superposed upon them, suggesting that lava flows created the intercrater highland plains during the late heavy bombardment roughly 3.9 billion years ago. This widespread volcanic resurfacing covered the older craters, erasing them from view, and the planet's heavily cratered terrain was then excavated out of the cooled and solidified lava near the end of the heavy bombardment recorded on the Moon.

Active volcanic past on Mercury

Widespread areas of Mercury are covered by relatively flat, sparsely cratered terrain called the smooth lowland plains (Fig. 6.13). They are younger than the intercrater highland plains, and they are about 2 kilometers lower. Unlike the dark maria on the Moon, the smooth lowland plains on Mercury have about the same brightness or color as the heavily cratered terrain and intercrater plains in the highlands of Mercury.

Detailed investigations from the *MESSENGER* spacecraft indicate an active volcanic past for Mercury. The Rembrandt impact basin (Fig. 6.14) has, for example, apparently been covered by effusive volcanism after the basin's formation, in multiple stages of smooth volcanic flow and surface contraction and deformation.

Smooth plains are found within and near craters and impact basins all over Mercury. They are attributed to gentle, effusive volcanic flows that occurred after the late heavy bombardment that excavated the older craters, but before the craters superimposed on them. The smooth plains are distributed across the planet, covering about 40 percent of its surface.

High-resolution imaging from *MESSENGER* has also revealed several non-circular, irregularly shaped depressions on Mercury, which are surrounded by bright, highly reflective halos with a distinctive color (Fig. 6.15). The depressions are interpreted as explosive volcanic vents, and the surrounding bright material as deposits ejected during volcanic eruptions at the vents.

Cliffs, or rupes, on Mercury

Mercury has another type of surface terrain not found on the Moon. They are the remarkable winding cliffs, or *rupes* – Latin for "cliffs" (Fig. 6.16), which are widely distributed over the planet and occur on a global scale. Individual cliffs range from 20 to more than 500 kilometers in length, and have heights from about 300 meters to about 3 kilometers. The cliffs cut across craters, and few craters are superimposed on them. This indicates that the cliffs were formed relatively late in early Mercury history, after the formation of the heavily cratered regions and the intercrater plains.

The long cliffs on Mercury look like the wrinkled skin of a shriveled apple, and the shrinking of the planet as the interior cooled probably caused them. The total shrinkage in the radius of the planet implied by the cliff heights is

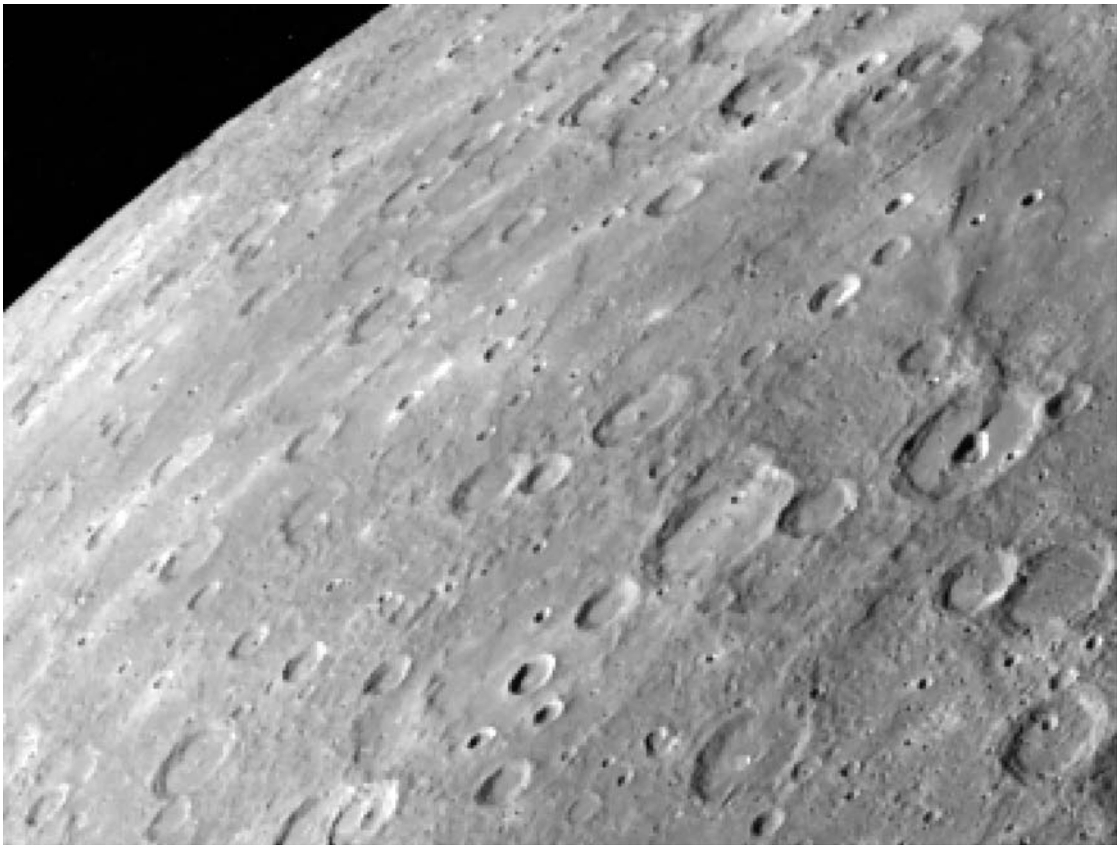


Fig. 6.12 Intercrater plains on Mercury The heavily cratered regions of Mercury exhibit abundant craters, just as the lunar highlands do, but unlike the Moon intercrater plains are found surrounding the craters in Mercury's highlands. This image, taken from the *Mariner 10* spacecraft, also exhibits long, dark cliffs, known as scarps or rupes, and a long ridge that runs along the right side of the image, cutting across a large crater about 80 kilometers in diameter. (Courtesy of NASA/JPL/Northwestern University.)

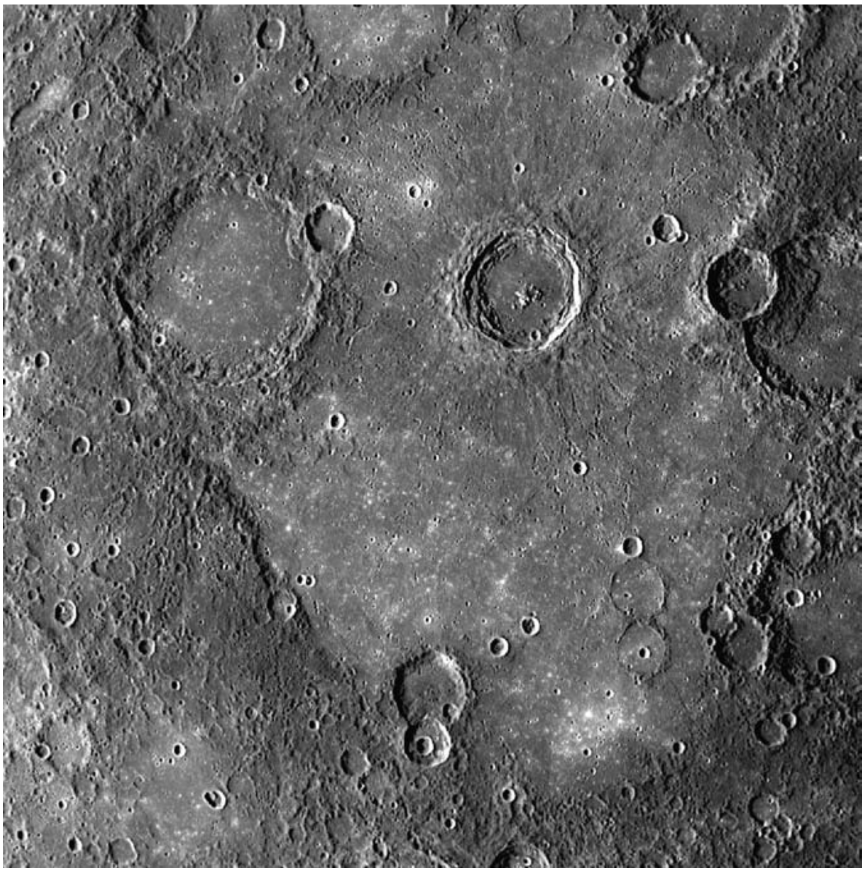


Fig. 6.13 Smooth plains on Mercury A large expanse of smooth plains on Mercury imaged from the *MESSENGER* spacecraft. Craters in these volcanic plains appear to have been significantly flooded with lava, leaving only their circular rims preserved. (Courtesy of NASA/JHUAPL/CIW.)

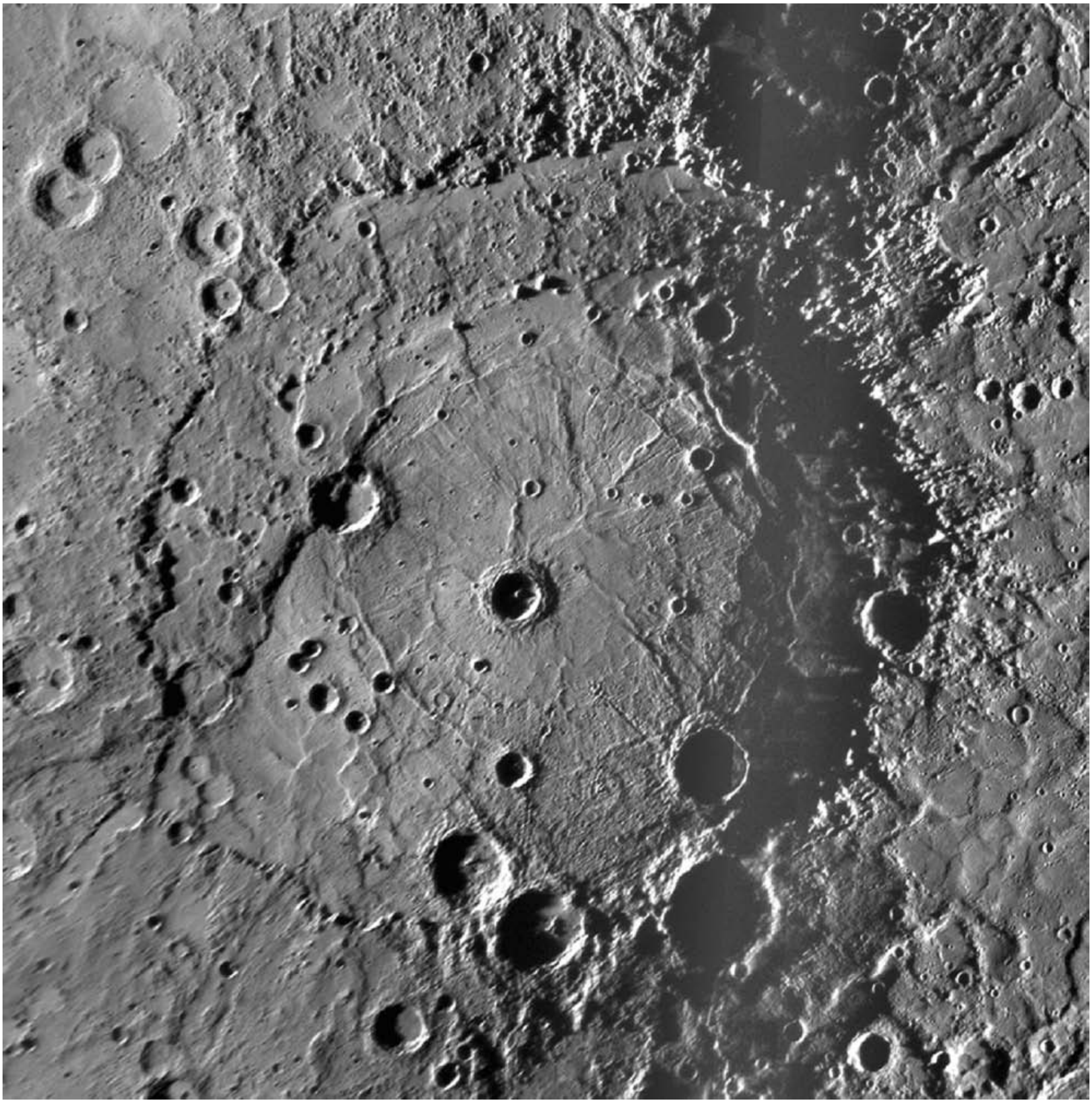


Fig. 6.14 Rembrandt impact basin This well-preserved impact basin was discovered in images taken from the *MESSENGER* spacecraft. Named Rembrandt, after the Dutch painter, it is about 715 kilometers in diameter. Scientists estimate that it was formed about 3.9 billion years ago, near the end of the period of heavy bombardment of the inner solar system. Although ancient, the distributions of smaller craters on its surface indicate that Rembrandt is younger than many other impact basins on Mercury. The basin and some of the small craters have also been flooded by effusive volcanic flows with a smooth appearance. Multi-colored images of the crater floor indicate areas with unusually high amounts of iron and titanium. (Courtesy of NASA/JHUAPL/SSI/CIW.)

estimated to be about 1 kilometer. Since the cliffs formed after most of the craters, but before some of the smaller craters found on them, Mercury probably shrank many hundreds of millions of years after the solidification of the crust, during the cooling of the planet's underlying mantle and partial solidification of its internal core. As the young planet solidified from the outside in, great blocks of its crust shifted up on one side and down on the other, thrusting the cliffs up into the sky.

There is no evidence of global shrinking on the Moon. This is probably because the Moon has a small metallic core, while Mercury has a large one. Mercury's entire globe probably shrank when at least some of its large iron core cooled and solidified.

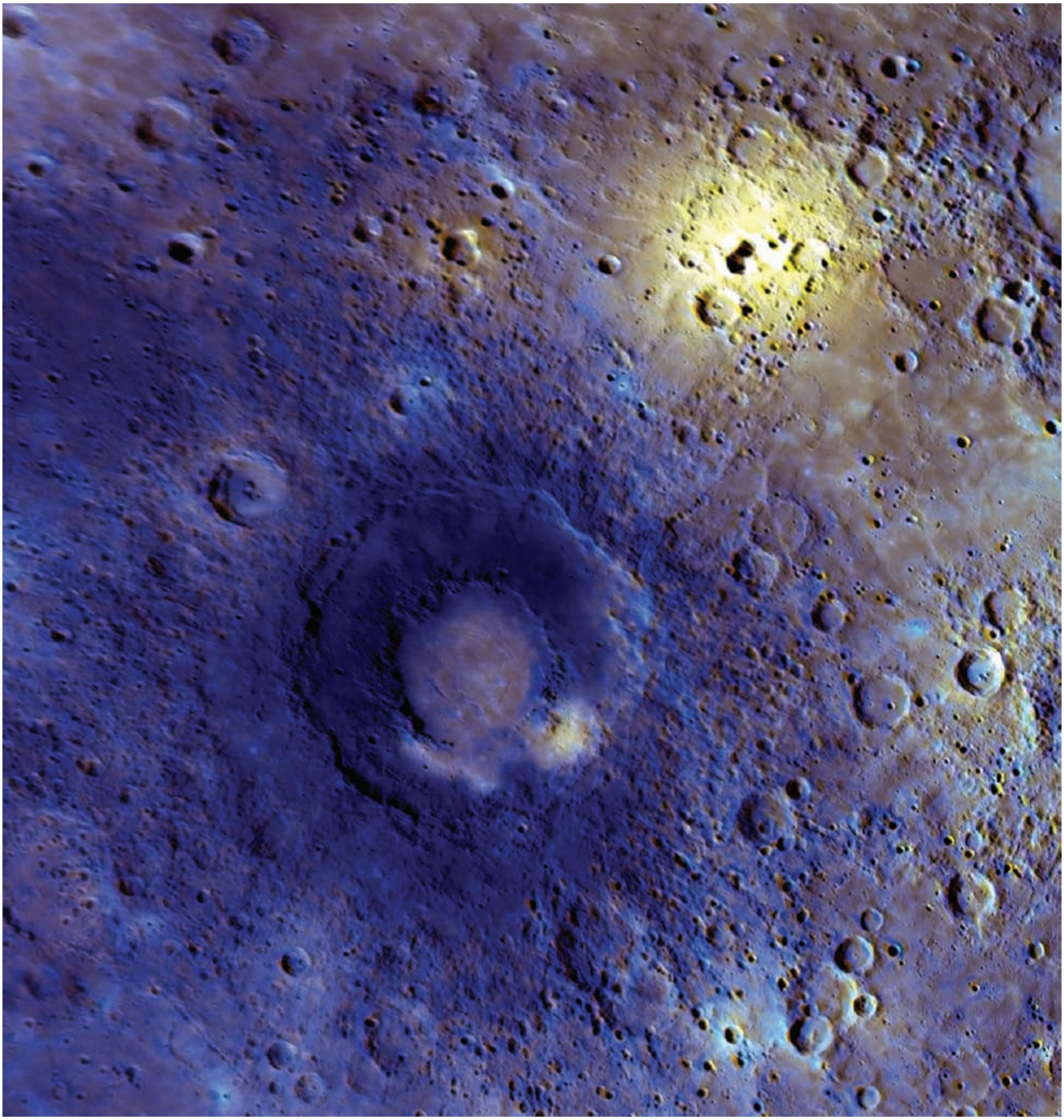


Fig. 6.15 Volcanic activity on Mercury This enhanced-color view, taken during the third flyby of Mercury by the *MESSENGER* spacecraft, enhances compositional differences in the features observed. The bright yellow area near the top right is centered on an irregular depression that is possibly an explosive volcanic vent. The double-ring basin in the center of this image may be filled by the flow of effusive volcanic lava; the basin is 290 kilometers in diameter. Smooth plains, thought to be the result of volcanic activity at earlier episodes than the others shown here, are located in much of the surrounding area. (Courtesy of NASA/JHUAPL/CIW.)

6.6 An iron world

The terrestrial bodies – the Moon, Mars, Venus and Earth – exhibit a fairly linear relationship between size and mean mass density, in which bigger objects have greater density, but Mercury does not conform to this relationship (Fig. 6.17). Although it is less than half the size of the Earth and not much bigger than our Moon, the bulk mass density of Mercury is 5427 kilograms per cubic meter, only slightly less than the Earth's bulk mass density of 5513 in the same units.

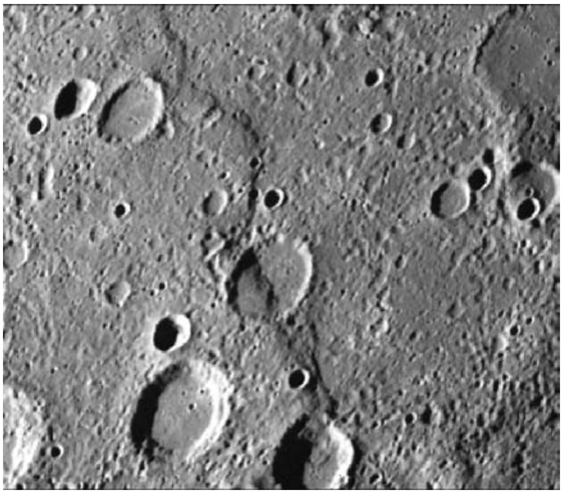


Fig. 6.16 Santa Maria rupes Mercury's surface is distinguished from those of the other terrestrial planets and the Earth's Moon by having enormous cliffs, or rupes, that cut across its surface. The dark lobate cliff that diagonally crosses this *Mariner 10* image has been named Santa Maria rupes, after Christopher Columbus' (1451–1506) flagship on his first voyage to America in 1492; "rupes" is the Latin word for "cliff". This long cliff was probably created when the planet cooled and contracted. The 90-km diameter crater in the upper right hand corner has been filled, or embayed, by intercrater plains. (Courtesy of NASA/JPL/Northeastern U.)

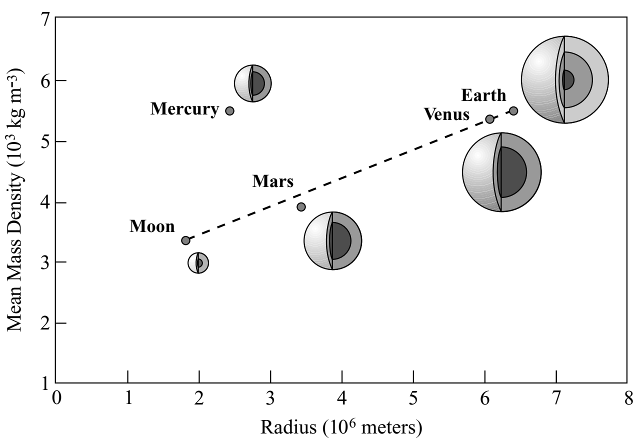


Fig. 6.17 Radius, mass density and interior structure of terrestrial bodies Mercury is an anomaly in this comparison of size and mean mass density. Mercury is just slightly bigger than the Moon and a little smaller than Mars, but its mean mass density is comparable to that of the larger terrestrial planets, Earth and Venus. As shown in the internal cross-sections, Mercury's unexpectedly large mass density is due to a large dense core extending up to 75 percent of its radius. The Moon's core occupies just 20 percent of its radius. The Earth's inner and outer core take up 55 percent of its radius, while that of Mars is smaller.

The most natural explanation of Mercury's high mass density is that it contains an unusual amount of iron, which is cosmically the most abundant heavy element. The dense iron core may take up to 75 percent of the planet's radius, or some 42 percent of its volume, so Mercury is mostly iron core surrounded by a relatively thin silicate mantle. In comparison, the Earth's iron core has a radius of 54 percent of the planet's radius and occupies just 16 percent of the planet's volume.

Soon after its formation, Mercury probably remained molten long enough for the heavy substance to settle at its center, just as iron drops below slag in a smelter. The high weight of the iron atoms would have slowly carried them down into the interior, leaving the lighter silicates to form a rocky mantle on top.

Why does Mercury have so much iron and so little rock? Some astronomers think the silicate rock was blasted off long ago, when Mercury collided with another planet-sized object. The collision might have removed much of Mercury's silicate mantle, leaving the large iron core largely intact. Or both colliding objects could have had thick rocky mantles that were completely vaporized in the collision, while their iron cores clumped together under their mutual gravitational pull. The light vaporized rock particles might have spiraled into the massive Sun, while the iron planet remained in orbit.

Such cataclysmic collisions were apparently common about 4.5 billion years ago, when a giant impact of a Mars-sized object with the nascent Earth created its Moon, with a lot of rock and little metal. A similar collision might have knocked Venus into its slow retrograde spin, in the opposite direction to both the rotation and orbital motion of all the other major planets. Another large impact in the final growth stages of Mars could additionally account for the large-scale dichotomy in its crustal topography.

In a second possibility, the intense radiation and particle emission from the active young Sun could have removed the silicate mantle from once-larger Mercury. In a third possibility, called selective accretion, the accumulation of heavy iron, as opposed to light silicate substances, was favored during the planet's formation; such an enrichment would not occur at greater distances from the Sun. There is not enough evidence to decide between the three possibilities.

6.7 A mysterious magnetic field

An instrument aboard *Mariner 10* showed that Mercury is a magnetized planet, with a magnetic field emanating from it. The increasing strength of this field as the spacecraft approached the planet indicated that magnetism at Mercury's surface would be 0.0033×10^{-4} tesla, about

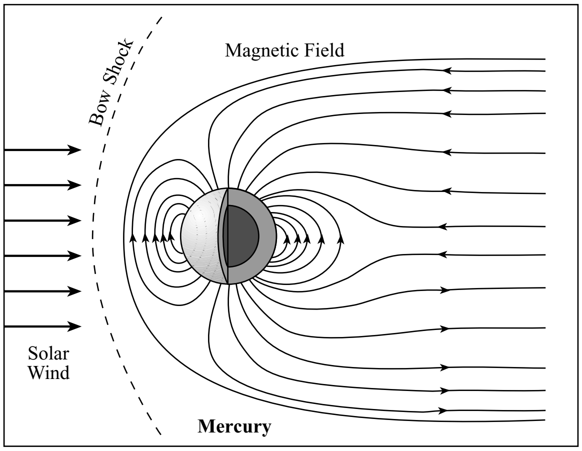


Fig. 6.18 Mercury's magnetic field and large core The magnetic field of Mercury is a miniature version of Earth's magnetic field, complete with bow shock, magnetosphere and magnetotail. Mercury's magnetic axis is closely aligned with its rotation axis, and its polarity is the same as the Earth's, with magnetic south corresponding to geographic north. This magnetic field is probably generated within the planet's large iron core, but the exact mechanism for creating the field remains a mystery. The electrified solar wind compresses the magnetic field into a bow shock on the sunlit side and draws it out into a tail on the opposite side.

1 percent or 0.01 times the strength of Earth's equatorial magnetic field.

Mariner 10's encounter provided magnetic data only for Mercury's eastern hemisphere, but during its second flyby *MESSENGER*'s magnetometer measured the western hemisphere, showing that the planet's magnetism is highly symmetric, with a dipolar magnetic field whose axis is closely aligned, within 2 degrees, with the planet's rotation axis.

The relatively weak magnetic field on Mercury is still strong enough to hold off the solar wind, forming a bow shock on the sunlit side of the planet and carving out an elongated magnetosphere with a tail pointing away from the Sun (Fig. 6.18). It is a scaled-down, miniaturized version of the Earth's magnetosphere, with the planet Mercury occupying a larger fraction of its magnetosphere. That is, Mercury's magnetosphere is relatively small, with a bow shock located at about 1.5 Mercury radii above the planet's surface; the bow-shock distance for Earth's magnetosphere is about 10 times larger, in planetary radii.

Because of Mercury's weak magnetic field strength and the small size of its magnetosphere, the outer boundary of its magnetosphere, known as the magnetopause, is frequently penetrated during interaction with the interplanetary magnetic fields emanating from the nearby Sun. When the interplanetary and planetary magnetic fields are pointing in opposite directions on contact, they merge together and reconnect into a new magnetic configuration, converting magnetic energy into kinetic energy. This opens Mercury's magnetosphere, exposing the planet to energetic charged particles in the Sun's winds.

The magnetometer aboard *MESSENGER* has demonstrated that such perforations of the dayside magnetosphere are large and frequent, with openings of up to 900 kilometers across and with a reconnection rate about 10 times that typical of Earth. The magnetic flux transfer has been likened to magnetic twisters dancing on the magnetopause, rooted deep in the planetary interior and carried into interplanetary space by the solar wind.

The discovery of Mercury's magnetic field was completely unexpected. Its presence implied the existence of an electrically conducting, molten core in which the magnetism is sustained by dynamo action. But back in 1974–75, at the time of the *Mariner 10* encounters with Mercury, most scientists thought that Mercury's core would have solidified long ago.

The argument for a solid core goes something like this. Small objects like Mercury have a high proportion of surface area to volume; that proportion decreases with increasing size of the object. Since internal heat is generated throughout the volume of a satellite or planet, and radiated to space only from the surface, smaller bodies radiate their energy into space faster and cool more rapidly. Since the Earth and Mercury are both the same age, and the much larger Earth has a solid inner core, retaining only a liquid outer core, the interior of Mercury should have completely solidified over its 4.6-billion-year lifetime.

Nonetheless, Mercury does have a global magnetic field; about 1 percent as strong as that found on Earth, suggesting that at least some molten material might be circulating inside the planet, and detailed radar measurements provided strong evidence that the planet does have a molten core. By 2007, radar astronomers had measured the planet's rotation with an accuracy of one-thousandth of a percent, discovering rotational twists back and forth that were double what would be expected for a completely solid body. This meant that the core, or at the very least a thin outer core, is at least partially molten, and decoupled from the overlying solid mantle. The liquid core is not forced to rotate with the mantle, and as the liquid sloshes around it alters the planet's overall rotation with a twisting first in one direction and then in another. Maintaining a molten iron core over up to 4.6 billion years may require that the core also contains an element lighter than iron, such as sulfur, to lower the melting temperature of the core material. When *MESSENGER* begins to orbit Mercury in 2011, it will provide additional evidence about the planet's core and its effect on the spacecraft's orbit.

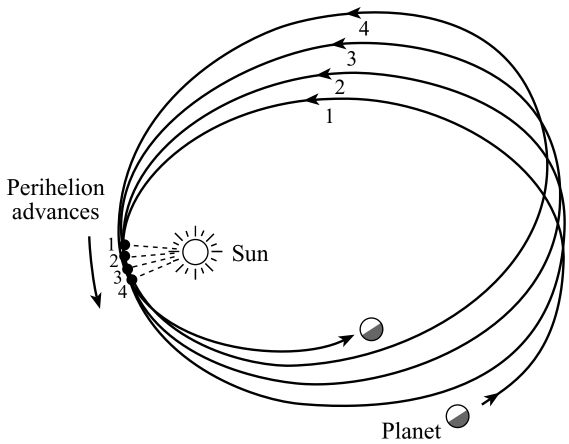


Fig. 6.19 Precession of Mercury's perihelion Instead of always tracing out the same ellipse, the orbit of Mercury pivots around the focus occupied by the Sun. The point of closest approach to the Sun, the perihelion, is slowly rotating ahead of the point predicted by Newton's theory of gravitation. This was at first explained by the gravitational tug of an unknown planet called Vulcan thought to be revolving about the Sun inside Mercury's orbit, but we now know that Vulcan does not exist. Mercury's anomalous motion was eventually explained by Einstein's new theory of gravity in which the Sun's curvature of space makes the planet move in a slowly revolving ellipse.

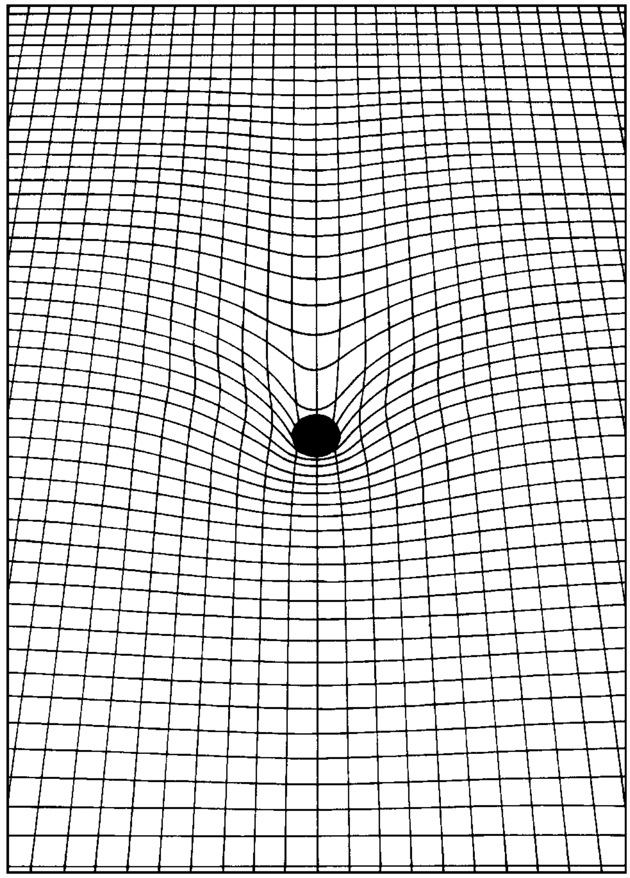


Fig. 6.20 Space curvature A massive object creates a curved indentation upon the flat Euclidean space that describes a world which is without matter. Notice that the amount of space curvature is greatest in the regions near the object, while further away the effect is lessened.

6.8 Einstein and Mercury's anomalous orbital motion

For nearly two and a half centuries, the solar system appeared to behave according to Newton's law of gravitation, which was used to predict the paths of the planets with great precision, but there was something wrong with Mercury's motion. Newton's theory failed to provide the expected connection between old and new measures of the planet's position, and its trajectory could not be precisely specified.

Instead of returning to its starting point to form a closed ellipse in one orbital period, Mercury moves slightly ahead in a winding path that can be described as a rotating ellipse. As a result, the point of Mercury's closest approach to the Sun, the perihelion, advances by a small amount, 43 seconds of arc per century, beyond that which can be accounted for by planetary perturbations using Newton's law (Fig. 6.19). This unexpected effect is known as the anomalous precession of Mercury's perihelion.

The unexplained twist in Mercury's motion was discovered more than one and a half centuries ago, as the result of watching the planet pass in front of the Sun every decade or so. This resulted in accurate determinations of the planet's orbital position. An analysis of these transits led the French mathematician Urbain Jean Joseph Le Verrier (1811–1877) to report the anomalous precession to the French Academy of Sciences in 1849. He subsequently attributed the anomalous motion to the gravitational pull of an unknown planet orbiting the Sun inside Mercury's orbit and moving ahead of it. The hypothetical planet was named Vulcan, and extensive searches were conducted for it. No such planet was ever reliably detected.

The cause of Mercury's anomalous motion remained a mystery until 1915 when Albert Einstein (1879–1955) explained it using a new theory of gravity in a paper entitled "Explanation of the Perihelion Motion of Mercury by Means of the General Theory of Relativity". According to Einstein's theory, space is distorted and curved in the neighborhood of matter (Fig. 6.20). In effect, space has both content and shape. The curved shape is molded into space by its content, the massive objects. That bending, twisting and distortion of space is gravity. But such curvature effects are only noticeable in extreme conditions very close to an exceptionally massive object, and

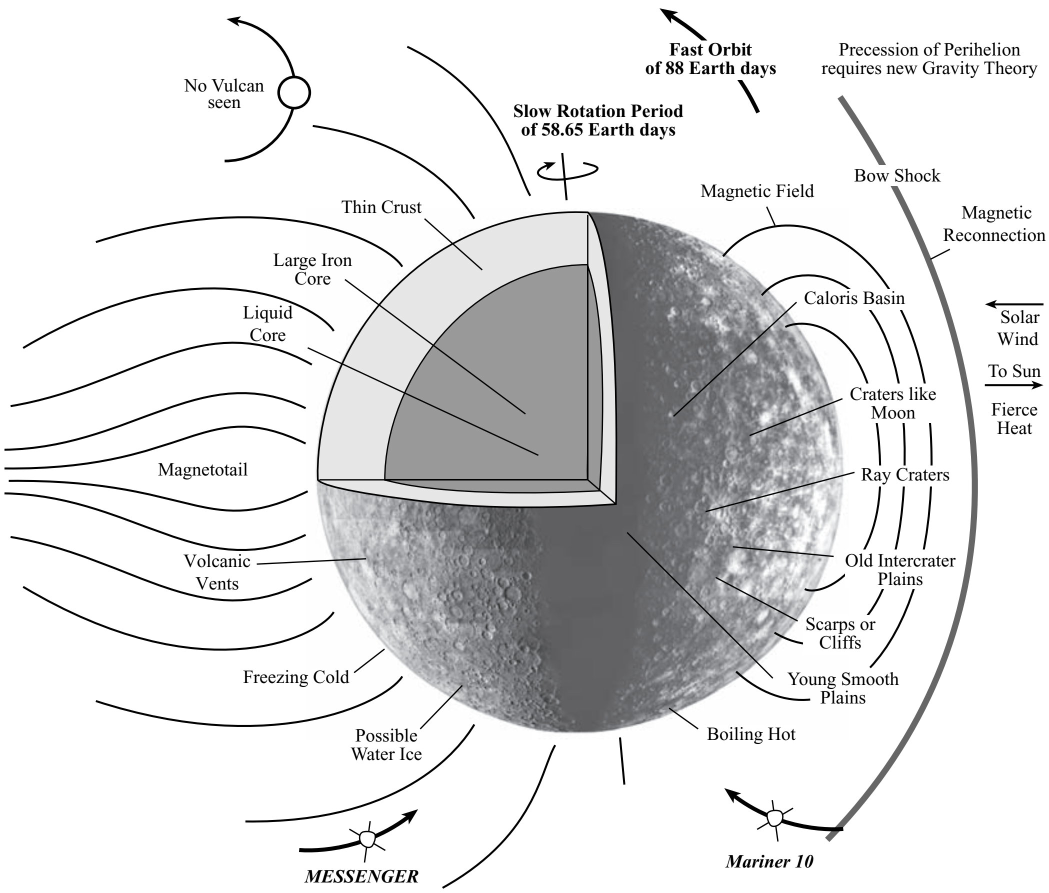


Fig. 6.21 Summary diagram.

the differences between the Newton and Einstein theories are indistinguishable in everyday life.

The curvature of space in the neighborhood of the massive Sun guides the motion of Mercury. The planet is traveling as straight as it can along an invisible curved track in space, like a ball spinning in a roulette wheel, making Mercury overshoot its expected perihelion position by precisely the observed amount of just 43 seconds of arc per century. Because the amount of space curvature produced by the Sun falls off with increasing distance, the perihelion advances for the other planets are much smaller than Mercury's.

The accord between Einstein's calculations and the observed motion of Mercury depends on the assumption that the Sun is a nearly perfect sphere. If the interior of the Sun is rotating very fast, it will push the equator out further than the poles, so its shape ought to be somewhat oblate rather than perfectly spherical. The gravitational influence of the outward bulge will provide an added twist to Mercury's orbital motion, shifting its orbit around the Sun by an additional amount and lessening the agreement with Einstein's theory of gravity. Fortunately, scientists have used sound waves, detected as five-minute pulsations of the Sun's visible disk, to see inside the Sun and show that the differences in internal rotation rates is not enough to produce a substantial asymmetry in the Sun's shape. So we may safely conclude that measurements of Mercury's orbit confirm Einstein's explanation of the planet's motion.

Unlike some of today's theoretical physicists, Einstein realized that his new theory needed to be verified by definitive predictions of other consequences. He noticed that light has energy and therefore mass, and that the Sun's gravity will pull the light toward it. The path of light passing near the Sun will therefore be bent by the curvature of space, and Einstein predicted that the apparent positions of stars would therefore be displaced when passing near the Sun. When Einstein's predicted value, of 1.75 seconds of arc for a Sun-grazing light ray, was confirmed by observations of stellar positions during a solar eclipse on 29 May 1919, it brought him international recognition.

The solar curvature of nearby space has been measured with increasingly greater precision for nearly a century. In one test, radio telescopes on opposite sides of the Earth combine their observations of the changing positions of remote galaxies or quasars when they pass behind the Sun. Another test measures the time for a radio signal to travel from a spacecraft home. When the line of sight passes near the Sun, the radio waves travel along a curved path and take slightly longer to return to Earth. The measurements require extremely precise clocks, for the extra time delay caused by the Sun's curvature of nearby space amounts to only one-ten-thousandth of a second.

In 2003, measurements using radio links with the *Cassini* spacecraft confirmed the predicted curvature to one part in ten thousand or to the fourth decimal place. Einstein's theory has now been verified by so many experiments and to such precision that it has become widely accepted as a brilliant contribution to our understanding of nature — begun by his attempts to account for Mercury's unexplained motion.

7 Venus: the veiled planet

- When visible, Venus is the brightest planet in the sky. It orbits the Sun inside Earth's orbit, appearing in the evening or morning hours and never in the middle of the night.
- No human eye has ever gazed on the surface of Venus, which is forever hidden by a thick overcast of impenetrable clouds.
- *Venera* spacecraft have parachuted through the clouds of Venus, surviving long enough to measure the properties of its torrid surface and even photographing it.
- The deadly efficient greenhouse effect of a thick, carbon-dioxide atmosphere has scorched Venus's surface, raising its temperature to 735 kelvin, even hotter than Mercury's average dayside temperature.
- In size, density and composition, Venus is almost identical to the Earth, but radar signals and space probes have penetrated its clouds to reveal an unearthly surface without a trace of liquid water or life.
- The pale yellow clouds of Venus are composed of concentrated sulfuric-acid droplets.
- The surface of Venus lies under a crushing atmosphere whose surface pressure is 92 times that on Earth.
- It takes only 4 Earth days for the high-flying clouds to move once about Venus, from east to west, blown by fierce, rapid winds, but the slow winds near the surface rotate with the planet, once every 243 Earth days in the same backwards, retrograde direction.
- The high, rapid winds on Venus spiral toward its poles, producing a huge, whirling polar vortex at both poles of the planet.
- There is no detectable magnetic field on Venus, but its dense atmosphere deflects the solar wind.
- Venus has a day longer than its year. The planet rotates once every 243 Earth days, in the opposite, retrograde direction from other planets except Uranus, and it takes 224.7 Earth days for Venus to orbit once about the Sun.
- The radar instrument aboard the *Magellan* spacecraft spent more than four years mapping out the surface of Venus in unprecedented detail, revealing rugged highlands, smooth plains, volcanoes, and sparse, pristine impact craters.

- About 85 percent of the surface of Venus is covered by smooth, low-lying volcanic flows of lava, and much of the remaining 15 percent is high-standing with towering volcanoes.
- The entire surface of Venus was probably covered by rivers of outpouring lava roughly 750 million years ago, wiping out all previous craters and about 90 percent of the planet's history; volcanic activity has continued at a reduced level up to the present.
- Tens of thousands of volcanoes now pepper the surface of Venus; some of the volcanoes could now be active.
- High volcanic rises on Venus are kept up by active motions below.
- Vertical motions associated with upwelling hot spots have buckled, crumpled, deformed, fractured and stretched the surface of Venus.
- Venus exhibits every type of volcanic edifice known on Earth, and some, called arachnoids and coronae, which have never been seen before.
- The surface of Venus moves mostly up and down, rather than sideways.
- Liquid water is non-existent on Venus, and the lack of water could be why Venus does not have moving plates similar to those found on Earth.

7.1 Fundamentals

Venus is the planet most like the Earth in size and mass. Its radius of 6051.8 kilometers is 94.9 percent of the Earth's radius, and its mass is 81.5 percent of the Earth's mass. Like the Earth, the planet Venus is a dense, rocky world, one of the four terrestrial planets. Venus spins in the backward direction, and so slowly that its day is longer than its year. The planet's surface lies under a hot and heavy atmosphere, with a high temperature and pressure, but no magnetic field has been observed on the planet.

7.2 Bright, beautiful Venus

Brilliant torch of the heavens

When visible, Venus is the most brilliant of the planets; it is the brightest object in the night sky, after the Moon. The stunning beauty of Venus must have been known since the dawn of human history. Our name Friday is, for example, derived from the Anglo-Saxon Frigadaeg, combining Friga, or Venus, and daeg, or day.

A female association has been common since the beginning of civilization. As Euripides (484–407 BC) put it:

Table 7.1 Physical properties of Venus[a]

Mass	$4.867\,32 \times 10^{24}$ kilograms = $0.815\ M_E$
Mean radius	6051.8 kilometers = $0.949\ R_E$
Bulk density	5243 kilograms per cubic meter
Sidereal rotation period	−243.018 Earth days, retrograde
Sidereal orbital period	224.7 Earth days = 0.615 197 3 Earth years
Mean distance from Sun	$1.081\,57 \times 10^{11}$ meters = 0.723 AU
Age	4.6×10^{9} years
Atmosphere	96.5 percent carbon dioxide, 3.5 percent nitrogen
Surface pressure	92 bars
Surface temperature	735 kelvin
Magnetic field strength	Less than 3×10^{-9} tesla or $10^{-5}\ B_E$

[a] The symbols M_E, R_E, and B_E denote respectively the mass, radius and magnetic field strength of the Earth.

"Venus, the eternal sway, all race of men obey". In another example, the Chinese named Venus *T'ai-pe* – "the Beautiful White One".

The name Venus is that of the ancient Roman goddess of love and beauty; the Greek equivalent was Aphrodite. The Greeks worshipped Aphrodite on the island of Cythera, and therefore the adjective "Cytherean" has often been applied to the planet.

The oldest recorded observations of Venus are those of the Babylonians, who called the planet Ishtar, "the bright torch of heaven" – the embodiment of all things womanly, the Mother of the Gods, and the goddess who evoked the power of dawn. The Maya built their calendar around the appearances and disappearances of Venus, though for them it was more fearsome than alluring.

There are other non-female names for the planet. The Judeo-Christian Devil is also known as Lucifer, which was originally a Latin name for Venus as a morning star. For the Mayan civilization, which flourished between 300 and 900 AD, Venus was the Sun's brother, the male god named Kukulkan, who preceded the Sun in rising from the underworld of night. The Mayan astronomer-priests could accurately predict Venus' appearance for over a hundred years, but they also got a bit carried away and made human sacrifices to the planet.

The view of Venus from Earth

Venus is visible at the edges of night, lingering near either dawn or dusk. It hangs low and bright in the morning or evening sky, sometimes near the crescent Moon. Because Venus's greatest angular distance from the Sun, known as its maximum elongation, is 47 degrees, it appears as the "evening star" just after sunset or as the "morning star" just before sunrise, but never as both an evening and morning star on the same day.

The brightest planet is never seen in the middle of the night or at midday. In the dark black of midnight, we are on the opposite side of the Earth from the Sun and Venus. At noon the planet is hidden in the full, bright glare of the Sun.

Venus is the second planet from the Sun, the world next door and the nearest planet to us in space. Every 19 months the planet swings to within 100 times the distance of the Moon. At closest approach, Venus is only 0.28 AU, or 38 million kilometers, from us, and its angular diameter is 64 seconds of arc.

Venus moves around the Sun in a nearly circular path once every 224.7 Earth days, like a runner on the inside track, at a mean distance of 0.723 AU. Since the Earth orbits the Sun at 1.00 AU in the same direction as Venus and with a slightly slower rate, it takes 584 Earth days or about 19 months for Venus to catch up with us. That is, every 19 months Venus passes between the Sun and us.

During each 19-month circuit, Venus is visible from the Earth for approximately 260 days as an evening star on one side of the Sun, and for about 260 days as a morning star on the other side of the Sun. Between its evening and morning appearances, Venus disappears from view; it then passes between us and the Sun or moves behind the star.

The approximate 260-day length of a Venus appearance in the morning or evening coincides closely with the average length of a human pregnancy. So the cosmos certainly seems to be in tune with female cycles. Most of us are familiar with a woman's monthly lunar cycle, and there is another Sun-related one that is not so well known. The Sun pulsates, moving in and out, every 5 minutes. This is the average length of a woman's contractions during childbirth; at least it was when my kids were born.

When viewed through a telescope, Venus brightens and fades, and also changes in apparent size, during its dance around the Sun (Fig. 7.1). As noticed by Galileo Galilei (1564–1642) in 1610, the planet exhibits a complete sequence of Moon-like phases, which means that Venus should orbit the Sun rather than the Earth. Its apparent illumination goes from a full round disk to a narrow crescent and back to rotundity again every 19 months. Venus also appears to grow when it approaches us in its orbit and shrinks as it recedes. When Venus is farthest from the Earth, on the opposite side of the Sun, it is fully illuminated and smallest. As the planet comes closer to Earth, it looks partly illuminated and larger.

Venus's exceptional brightness is partly due to its highly reflective clouds, which reflect about 65 percent of the incident sunlight back into space. By way of comparison, Mercury and the Earth's Moon reflect roughly 10 percent of the sunlight reaching them.

The clouds that help make Venus the brightest of planetary worlds also perpetually hide its surface from view (Fig. 7.2). No features can be seen beneath the unbroken layer of clouds by the human eye, or even with a telescope on the ground or in space. The high-flying clouds whip around the planet from east to west every four Earth days, in the same direction as the planet rotates but in the opposite, retrograde direction to its orbital motion.

Spectroscopic observations have been more rewarding than casual visual inspection, showing in the 1930s that the planet's upper atmosphere is mainly composed of carbon dioxide (CO_2). We now know that its thick, massive atmosphere is 96 percent CO_2, and that it contains about 300 000 times as much CO_2 as is present in our air.