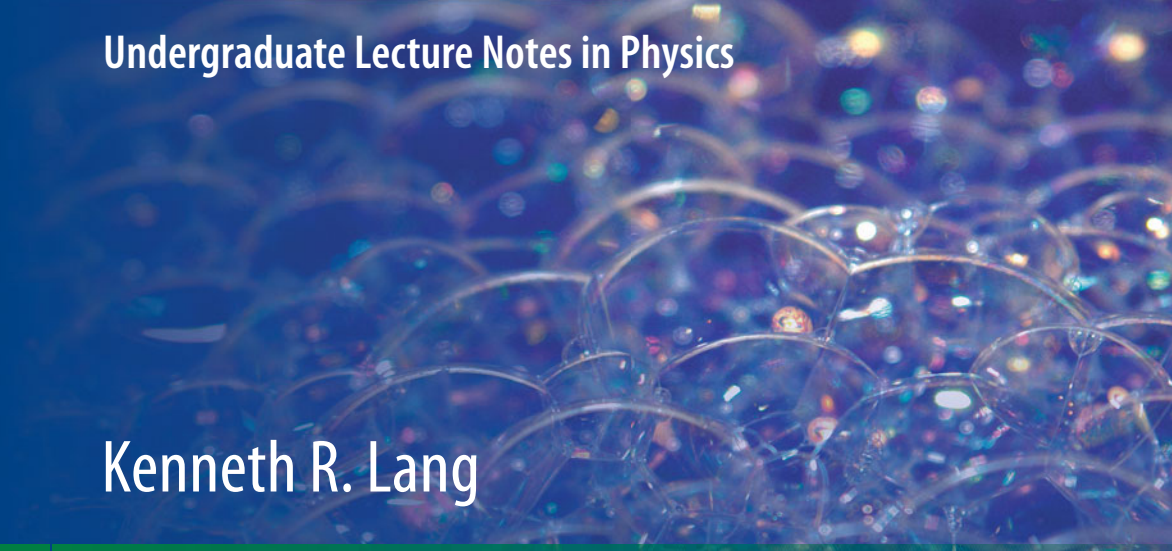
Undergraduate Lecture Notes in Physics



Kenneth R. Lang

Essential Astrophysics

Springer

Undergraduate Lecture Notes in Physics

For further volumes:
http://www.springer.com/series/8917

Undergraduate Lecture Notes in Physics (ULNP) publishes authoritative texts covering topics throughout pure and applied physics. Each title in the series is suitable as a basis for undergraduate instruction, typically containing practice problems, worked examples, chapter summaries, and suggestions for further reading.

ULNP titles must provide at least one of the following:

- An exceptionally clear and concise treatment of a standard undergraduate subject.
- A solid undergraduate-level introduction to a graduate, advanced, or non-standard subject.
- A novel perspective or an unusual approach to teaching a subject.

ULNP especially encourages new, original, and idiosyncratic approaches to physics teaching at the undergraduate level.

The purpose of ULNP is to provide intriguing, absorbing books that will continue to be the reader's preferred reference throughout their academic career.

Kenneth R. Lang

Essential Astrophysics

Springer

Kenneth R. Lang
Department of Physics and Astronomy
Tufts University
Medford, MA
USA

ISSN 2192-4791 ISSN 2192-4805 (electronic)
ISBN 978-3-642-35962-0 ISBN 978-3-642-35963-7 (eBook)
DOI 10.1007/978-3-642-35963-7
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012955653

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Essential Astrophysics is a book to learn or teach from, as well as a fundamental reference for anyone interested in astronomy and astrophysics. This unique volume can be used as a textbook, teaching guide, or reference source for just about anyone interested in astronomy and astrophysics.

It serves as a comprehensive, introductory text, which takes the student through the field of astrophysics in lecture-sized chapters of basic physical principles applied to the cosmos. Undergraduate students with an interest in the physical sciences, such as astronomy, chemistry, engineering, or physics, will enjoy this one-semester overview.

The text is of sufficient breadth and depth to prepare the interested student for more advanced, specialized courses in the future. The clarity and comprehensive nature of *Essential Astrophysics* make it a significant resource for the curious reader that is unfamiliar with astrophysics or for professional astronomers who may have forgotten the basics.

Astronomical examples are provided throughout the text, to reinforce the basic concepts and physics, and to demonstrate the use of the relevant formulae. In this way, the student learns to apply the fundamental equations and principles to cosmic objects and situations. All of the example problems are solved with the rough accuracy needed to portray the basic result. Such order-of-magnitude estimates are commonly used in astronomy and astrophysics, where large numbers are involved, and an understanding of the underlying physics does not require engineering accuracy.

Essential Astrophysics is a serious introduction to astrophysics complete with the necessary formulae. These equations sometimes include the calculus of integration, or adding up, and differentiation, that are found in the author's classic *Astrophysical Formulae* and more advanced textbooks. Nevertheless, the end result in *Essential Astrophysics* is always a simple algebraic relationship that can be applied to cosmic objects. These fundamental equations are given in the text and collected at the end of the book in Appendix III, for future reference and use. Therefore, only elementary algebra is required to solve any of the example problems or other numerical conclusions in this book.

There are two types of intended readers. One type will be interested in broad, general conclusions, without use of calculus. This reader will be content with the existing text with no further elaboration. The more mathematically competent reader will want to use *Essential Astrophysics* as a foundation for more advanced considerations, with the guidance of the references, an instructor, or an advanced textbook, using the formulae found in the text or within set aside Focus Elements of *Essential Astrophysics* as a starting point.

The modern SI (International System) units are used in the equations and example problems, which is another unique aspect of this book when compared to most previous texts of astrophysics. A conversion table between the SI and c.g.s. units is provided in the first chapter, to help the reader follow the details of many papers and textbooks that use the older c.g.s. system. Astronomical and physical constants, units, and fundamental equations are provided in appendices, for quick reference.

Essential Astrophysics goes beyond the typical textbook by providing comprehensive access to astrophysical discoveries, concepts, and facts that are not available in any other way. It gives us access to that long-forgotten formula, idea, or reference, while also providing the material needed to introduce anyone to a new area of astrophysics. Here, the reader can obtain the background required for a general understanding and find guidance to the relevant literature including seminal discoveries, original research, and comprehensive up-to-date reviews that will enable the curious reader to delve deeper into a particular topic. A more extensive reference compilation of developments in astrophysics, from then to now, can be found in *Astrophysical Formulae*.

We are the benefactors of 300 years of cumulative discovery in astronomy and astrophysics, and *Essential Astrophysics* helps pass on these fundamental insights to the next generation. It also reveals both the exciting moments of the past and relatively recent discoveries. Historical aspects are illuminated through a progressive flow of chapter topics and by guidance to the earliest ideas, with reference to the original sources as well as contemporary reviews. Perhaps because of the rapid pace of modern research, contemporary texts often focus on specialized topics and overlook these broader perspectives that *Essential Astrophysics* provides.

There are 50 set-aside focus elements that enhance and amplify the discussion with fascinating details. They include the intriguing development of particular themes, which is missing in most astrophysics textbooks, or provide further astrophysics or equations for use in examples, problems or further investigations.

In *Essential Astrophysics* we can rediscover basic physical concepts such as space, time, radiation, mass, gravity, motion, heat, atoms, radioactivity, and cosmic rays, which are required to understand the observable universe. These fundamental topics are discussed in the first seven chapters, beginning with the introductory chapter that describes how astronomers observe the contents of the universe and how astrophysicists interpret them. The SI units of distance, mass, time, energy, and luminosity are introduced, together with their astronomical units

such as the Ångström, light-year, parsec, and the Sun's mass, luminosity, and radius. The magnitude unit is also defined, but used sparingly in examples.

Chapter 2 describes radiation, of both the visible and invisible sort, which carries messages from the cosmos and tells us much of what we know about it. Chapter 3 discusses gravity, together with mass that helps determine its strength, and related tidal phenomena and space curvature. Chapter 4 discusses cosmic motion, and its balanced equilibrium with gravitation. Chapter 5 discusses the motion of particles in a gas, together with the related concepts of speed distribution, heat, temperature, and pressure. The inside of the atom is explored in Chap. 6, where the reader learns about atomic spectral lines and their use in determining the composition of stars and the measurement of motions and magnetic fields. The transformation of elements in both radioactivity and by subatomic bombardment is presented in Chap. 7.

The fundamental concepts described in these first seven chapters provide a necessary prelude to the rest of the book. It includes the discoveries that the universe is predominantly hydrogen; that the stars shine by nuclear fusion; that the stars live and die while new ones continue to be formed; that the interstellar spaces are not empty but filled with dust, atoms, and molecules; and that the observable universe is expanding and has a history. The last half of *Essential Astrophysics* also includes relatively recent discoveries, such as pulsars, black holes, the three-degree cosmic microwave background, the formation of stars and galaxies, invisible dark matter, and the dark energy that is now accelerating the expansion of the universe.

Chapter 8 provides an account of the nuclear fusion reactions that make the Sun shine. This is followed in Chap. 9 by modern discoveries of the Sun's expanding atmosphere, the solar winds, explosions on the Sun, the solar flares and coronal mass ejections, and their space–weather threats to spacecraft and humans in space.

Chapter 10 presents an overview of the stars, telling us how far away, bright, luminous, hot, big, and massive they are. It also includes discussions of stellar spectra, as well as the evolution of stars and their role in the origin of the chemical elements.

The space between the stars is discussed in Chap. 11, beginning with bright stars that illuminate nearby space and continuing with the dust, gas, radio emission, and molecules within interstellar space. This is naturally followed in Chap. 12 by the ongoing formation of stars and their planets; recent discoveries of protoplanetary disks and planets around nearby stars can also be found in this chapter.

The final destiny of stars, when they have depleted their nuclear resources, is presented in Chap. 13. It includes planetary nebulae, white dwarf stars, degenerate pressure, novae, supernovae, neutron stars, pulsars, and stellar black holes.

Our last two chapters discuss the observable universe in its entirety, including the Milky Way, the receding galaxies, the big bang with its background radiation, the first atoms, stars, and galaxies, the evolution of galaxies, dark matter and dark energy, and the ultimate destiny of the universe.

A total of 69 tables provide vital facts and physical information for the main types of cosmic objects; students, teachers, and researchers may also consult this information throughout their careers. In alphabetical order, they include the physical properties of atmospheres, clusters of galaxies, the cosmic microwave background radiation, the Earth, emission nebulae, galaxies, our Galaxy, giant molecular clouds, H I regions, H II regions, interstellar molecules, the Milky Way, our Moon, neutron stars, novae, planetary nebulae, planets, pulsars, radioactive isotopes, the Sun, stars, star clusters, supernova explosions, and supernova remnants.

Our tables also include information about cosmic magnetic fields, cosmic rays, cosmological parameters, and nuclear fusion processes, as well as the range of cosmic pressures, cosmic temperatures and stellar luminosity, and the spectral lines of active galaxies, emission nebulae, stars, the Sun's corona, and the Sun's photosphere.

There are also excellent line drawings, prepared by Kacha Bradonjich, and several images of astronomical objects from the ground and space that help cement our newfound knowledge together. They help crystallize a new concept with a visual excitement that adds another dimension to our understanding.

The author also writes another sort of popular book, filled with personal anecdotes, the lives of contributors to the field, and human metaphors, without an equation or reference in sight. For this complementary approach, the reader is referred to the author's two books *The Life and Death of Stars* and *Parting the Cosmic Veil*, which deal with many of the same general topics as *Essential Astrophysics* in a different, lighter perspective.

I am indebted to Gayle Grant for help in assembling this book, and to the Tufts Faculty Research Committee for modest support for typing some equations in it. And last, but not least, the author thanks Ramon Khanna for his skillful editorial suggestions that have made *Essential Astrophysics* a better book.

Medford, November 2012 Kenneth R. Lang

Contents

Focus Elements

Tables

Chapter 1
Observing the Universe

1.1 What Do Astronomers and Astrophysicists Do?

Astronomy is an ongoing, cumulative science in which astronomers either discover previously unseen constituents of the observable universe or determine physical properties of known ones. They measure the mass, luminosity, distance, size, chemical composition, motion, and magnetic fields of planets, stars, galaxies, and their surroundings.

Astrophysicists apply the laws of physics to celestial objects and events, thereby interpreting and explaining the astronomical observations. They assume that the physical laws that apply on Earth are valid throughout the Cosmos, but often under extreme conditions that cannot be achieved on our planet. The diverse aspects of physics used in astrophysics include radiation processes and universal gravitation, cosmic and particle motion, atomic and nuclear physics, and special and general relativity.

Astronomers and astrophysicists together investigate how everything in the universe originates, changes, interacts, moves, and radiates. Theoretical studies, analytical models, and numerical simulations with computers are also employed to help understand these processes.

Astronomy, and therefore astrophysics, is an instrument-driven science. Many of the seminal discoveries in astronomy have been accidental and unanticipated, often made when using unique telescopes, new technology, and novel detection equipment (Lang 2009). These instruments extend our vision to places that are not accessible to direct observation, enabling us to "see" the invisible and permitting us to look at the universe in new ways. Without a telescope, for example, the vast majority of stars cannot be seen, and all but a very few of the billions of galaxies and most of the expanding universe are invisible to the unaided eye.

Observations provide the crucial data for our celestial science. Without them, astrophysicists would have nothing to describe. Fortunately, an astronomical object can be observed over and over again, in different ways, once it has been

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_1, © Springer-Verlag Berlin Heidelberg 2013

discovered. These observations require knowledge of our location on the Earth, the location of the object in the sky, and an understanding of both angular measure and passing time.

1.2 Our Place on Earth

In order to observe cosmic objects with any accuracy, we must first establish our bearings here on Earth. In arguments used by Pythagoras (572−479 BC), 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. Only a spherical body can cast a round shape 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.

So we can, to first approximation, assume the Earth is a sphere, and locate ourselves within a grid of great circles on it. A great circle divides the sphere in half; the name derives 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. Each circle of longitude intersects the equator in two points that are 180° apart. We halve the great circles of longitude into semicircles, called meridians. Long ago, in 1884, it was decided that the half-circle

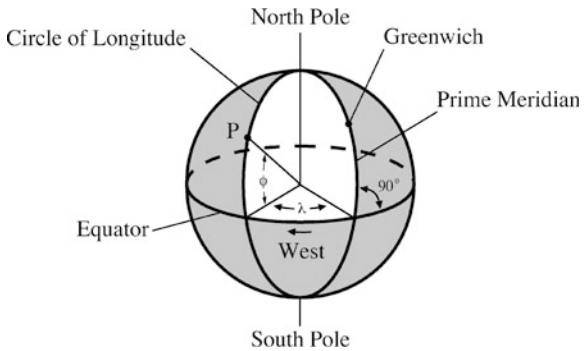


Fig. 1.1 Latitude and longitude Great circles through the North and South Poles of the Earth create circles of longitude. They are perpendicular to the Equator where they intersect it. The circle of longitude that passes through Greenwich England is called the Prime Meridian. The longitude of any point, P, is the angle lambda, λ, measured westward along the Equator from the intersection of the Prime Meridian with the Equator to the equatorial intersection of the circle of longitude that passes through the point. The latitude is the angle phi, ϕ, measured northward (positive) or southward (negative) along the circle of longitude from the Equator to the point. In this figure, the point P corresponds to San Francisco

of longitude passing through the old Royal Observatory in Greenwich, England, would mark 0° longitude. It was designated as the "Prime Meridian", the starting point of counting longitudes.

The longitude, denoted by the Greek letter λ, of any point on the Earth's surface is the angle measured westward from the intersection of the Prime Meridian with the Equator to the equatorial intersection of the circle of longitude that passes through the point (Fig. 1.1). The latitude, designated by the symbol ϕ, is the angle measured northward (positive) or southward (negative) along a circle of longitude from the equator to the point. Sobel (1995) has discussed early determinations of terrestrial longitude, whereas Carter and Carter (2002) have provided a historical account of latitude variations. Alder (2002) has discussed early measurements of the size of the Earth, and the associated beginning of the metric system.

Example: 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. The surface speed of rotation is greatest at the equator and reduces to lower values at higher latitudes. Using an equatorial radius of about 6,378 km, the Earth would have to be rotating at a speed of about 464 m s^{-1} to spin about its equatorial circumference once every 24 h. To calculate this speed, just multiply the equatorial radius by 2π to get the equatorial circumference, and divide by 24 h, where there are 86,400 s per hour. The constant $\pi = 3.1416$. 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.

Every location on the Earth rotates about an axis that pierces the Earth and extends between its North and South Poles. The period of rotation, and the length of the day, is everywhere the same, but the rotation speed is fastest at the equator and systematically lower at higher latitudes.

The geographic description of the location of an observatory includes its height, h, in meters above mean sea level. The geodetic coordinates of longitude, latitude and height, designated λ, ϕ, and h, are specified online in *The Astronomical Almanac* for all observatories engaged in professional programs of astronomical observations.

The Global Positioning System (GPS) is now used to determine reliable location and time information. It is a system of about 30 navigation satellites developed, maintained and operated by the U.S. Air Force for military and civilian purposes. Each satellite is constantly beaming radio signals that contain the exact time. These signals take a few milliseconds to travel from a satellite to the GPS receiver, and it has a built-in computer that calculates its precise position on Earth using signal time delays from four or more satellites. The time differences are

converted into distance by multiplication with the speed of light, and these distances are translated by triangulation into an exact position accurate to about 100 m.

The numbers after the N (north) notation on the GPS receiver indicate its latitude, while the numbers after the W (west) notation indicates its longitude. The GPS devices used in automobiles specify the longitude and latitude of your start and end points, and map the route between them.

The GPS was initially developed by the military and is still used by them. Soldiers can use a GPS device to find an enemy objective, even in the dark or unfamiliar territory; weapons systems can use them to track potential ground and air targets. All GPS receivers capable of functioning above 18 km in altitude and moving faster than 515 m s^{-1} are classified as weapons.

1.3 Location in the Sky

You may have watched the stars as they rise at the horizon on one side of the Earth, slowly move overhead, and eventually set on the other side of the planet, only to reappear the next night. This slow coursing of stars was initially attributed to a revolving celestial sphere, which carried its embedded stars about a stationary Earth, but appearances can be deceiving. The Earth is instead spinning under an imaginary celestial sphere concentric with the Earth, on which the stars and other astronomical objects are placed. Such a celestial sphere explains why people located at different places on Earth invariably see just half of all the stellar sky. As the Earth rotates, day turns into night and these stars glide by.

Astronomers define points and circles on the celestial sphere (Fig. 1.2). If you extend the Earth's rotation axis in both directions, it intersects the celestial sphere at the north and south celestial poles. They are the pivotal points of the night sky's apparent daily rotation. When the plane of the Earth's Equator is extended outward in all directions, it cuts the celestial sphere in half, at the celestial equator. The point where the Sun crosses the celestial equator going northward in spring is called the Vernal Equinox. The Vernal Equinox is sometimes called the first point of Aries, and is given the symbol γ.

The projection of the plane of the Earth's orbit onto the celestial sphere is known as the ecliptic, and the angular separation between the ecliptic and the celestial equator is called the obliquity of the ecliptic, designated ε, which is about 23.5°. The obliquity is also the angle between the Earth's rotational axis and a line perpendicular to its orbital plane. On the standard reference date of January 1.5, or at noon in January 1, in the year 2000.0, the slowly changing obliquity had the exact value of:

$$\varepsilon = 23^\circ\, 26'\, 21.406'',$$

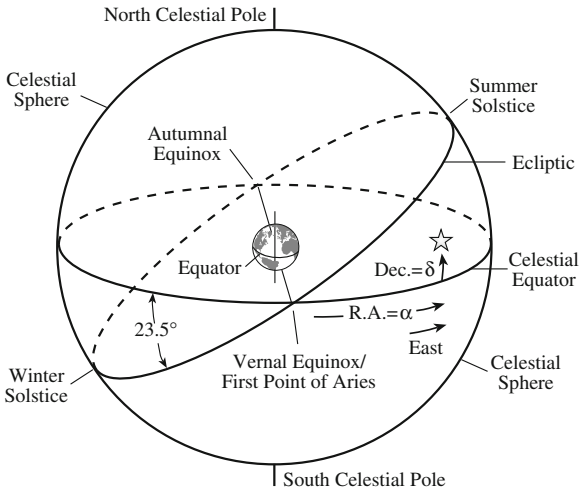


Fig. 1.2 Celestial coordinates Stars, galaxies, and other cosmic objects are placed on an imaginary celestial sphere. The celestial equator divides the sphere into northern and southern halves, and the ecliptic is the annual path of the Sun on the celestial sphere. The celestial equator intersects the ecliptic at the Vernal Equinox and the Autumnal Equinox. Every cosmic object has two celestial coordinates. They are the right ascension, designated by the angle alpha, α, or by R.A., and the declination, denoted by the angle delta, δ, or Dec. Right ascension is measured eastward along the celestial equator from the Vernal Equinox to the foot of the great circle that passes through the object. Declination is the angular distance from the celestial equator to the object along the great circle that passes through the object, positive to the north and negative to the south. Precession results in a slow motion of the Vernal Equinox, producing a steady change in the celestial coordinates

where the symbol ° denotes angular degrees, the $'$ symbol designates minutes of arc or angle, and the symbol $''$ denotes seconds of arc.

Positions on the celestial sphere are defined by angles along great circles. By analogy with terrestrial longitude, right ascension, denoted α, is a celestial object's longitude, but it is measured eastward along the celestial equator from the Vernal Equinox. The right ascension is expressed in hours and minutes of time, with 24 h in the complete circle of 360 degrees, denoted as 360°. For conversion, 1 h of time is equivalent to 15° of angle, or 1 h = 15°; 1 s of time is equal to 15 s of arc, or 1 s = $15''$; and 1 min of arc equals 4 s of time, or $1' = 4$ s.

Just as latitude is a measure of a one's distance from the Equator of the Earth, declination, denoted δ, is a celestial object's angular distance from the celestial equator. The declination is positive for objects located north of the celestial equator, and they are observed by inhabitants of the northern hemisphere of the Earth. Celestial objects in the southern sky have negative declinations, and people living in the southern half of our planet observe them.

Example: What can you see in the night sky from where you are?

An observer can see only half of the celestial sphere – the half above the local horizon. The celestial objects that can be seen depend on only two things, the observer's latitude, denoted by ϕ, and the object's declination, designated δ. At the Earth's geographic North Pole, and latitude $\phi = 90°$, the north celestial pole is directly overhead and located near the star Polaris, and the horizon runs along the celestial equator. This observer can therefore only see the northern half of the celestial sphere, and objects with positive, northern declination; all the objects with negative, southern declinations are forever invisible from this location. For an observer located at different northern latitudes ϕ, stars can be observed with declinations greater than $\phi - 90°$, or $\delta > \phi - 90°$, for $0 \leq \phi \leq 90°$; only these stars rise above the horizon. The southern half of the celestial sphere is fully visible from the geographic South Pole, while the northern sky is unseen. At the Equator where $\phi = 0$, the complete range of positive and negative declinations are visible. Of course, the observer has to wait for the rotating Earth to bring any potentially observable object above the horizon and into the observable half of the celestial sphere.

For centuries, astronomers have used catalogues of right ascension, α, and declination, δ, of celestial objects to locate them in the sky (Focus 1.1). Modern celestial positions of the highest accuracy are referred to the center of mass of all bodies in the solar system, known as the solar system barycenter, and specified within an International Celestial Reference System (Kaplan 2005).

Focus 1.1 Astronomical catalogues

The positions, brightness, spectra, angular size, and other data of different kinds for celestial objects are given in catalogues provided over centuries of meticulous observations by dedicated astronomers. Stars were the first celestial objects to be catalogued. Their accurate positions were compiled and used to discover such things as stellar motions, the planet Uranus, and the first asteroid. The positions and spectral classification of about 235,000 stars were compiled in the famous *Henry Draper Catalogue*, published between 1918 and 1924. The letters "HD" followed by the listed number in this catalogue often designates a star. Gliese (1969) and Gliese and Jahreiss (1979) have catalogued the closest stars other than the Sun. Planets are now being discovered around such nearby stars designated by "GJ" followed by their number in this catalogue – GJ 581, for example.

The French astronomer Charles Messier (1730−1817) compiled one of the most famous catalogues (Messier 1781). His list of just over 100 bright non-stellar objects includes some of the most widely studied objects in the universe, now known as emission nebulae, galaxies, star clusters, and supernova remnants. The letter "M" followed by the number in the Messier

Catalogue indicates them. The Crab Nebula supernova remnant, M 1, is the first on the list, and M 31 is the closest spiral galaxy, also known as Andromeda.

In astronomical parlance, a nebula is a diffuse, non-stellar object. Some extragalactic nebulae, which reside outside our Milky Way, were eventually designated as galaxies. They contain as many as 100 billion stars, as well as diffuse gaseous nebulae. An emission nebula consists of interstellar gas glowing from the ultraviolet light of a nearby luminous star.

William Herschel (1738−1822) dramatically increased the number of known non-stellar objects to 2,500, during 20 years of systematically observing the heavens, from 1783 to 1802. This sweep of the sky's northern hemisphere was extended to the southern hemisphere by William's son, Sir John Herschel (1792−1871), who published data for 5,079 objects in his *General Catalogue* in 1864, the combined result of more than half a century of painstaking observations.

Using the Herschel catalogue as a basis, J. L. E. Dreyer (1852−1926) published his *New General Catalogue* (NGC) of nebulae and star clusters, followed by two *Index Catalogues*, designated IC. Many galaxies, as well as emission nebulae and star clusters, are still known by their NGC and IC numbers.

Later on, the photographic Palomar Sky Survey, using the wide-angle 1.2 m (48 in.) Schmidt telescope on Palomar Mountain, was used to catalogue tens of thousands of galaxies; an observatory telescope is often designated by a name and its diameter in meters or inches. In 1958, George Abell (1927−1983) used it to create a catalogue of 2,712 rich clusters of galaxies; the designation "A" followed by the number in his catalogue is still used today. Millions of galaxies were catalogued in the mid-twentieth century using photographs taken using large telescopes, but galaxy catalogues have now become computerized. An important example is the Sloan Digital Sky Survey, created from a dedicated, computer-driven 2.5 m (98 in.) telescope.

Bright radio sources are designated by "3C" followed by the number in the *Third Cambridge Catalogue of Radio Sources* published in 1959, a famous example is 3C 273, the first quasar to be discovered. The most intense radio and x-ray sources have been named after the constellation they appear in. For example, Cygnus A is a bright radio galaxy, Cygnus X-1 is a bright x-ray source and a candidate black hole, and Centaurus X-3 is an x-ray pulsar.

Specific types of objects, such as pulsars, supernova remnants, and white dwarf stars, have their own catalogues, and are often designated by letters like "PSR", "SNR", or "WD" followed by their celestial position.

The international celestial reference frame is defined by a catalogue of exceedingly accurate positions for extragalactic radio sources observed with Very Long Baseline Interferometry. At optically visible wavelengths, the Tycho-2 catalogue of positions for more than 2.5 million stars, observed from the *HIPPARCOS* satellite, is used. Accurate positions of major solar system bodies are given as a function of time in an *Ephemerides* provided by the Jet Propulsion Laboratory.

As the Sun moves along the ecliptic, it crosses the celestial equator twice, on its way north at the Vernal Equinox, on about March 20, and then at the Autumnal Equinox on about September 23. On either equinox, the Sun lies in the Earth's equatorial plane, so the twilight zone that separates night and day then cuts the Earth in equal parts and the days and nights are equally long. The point at which the Sun is farthest north, is the Summer Solstice (on about June 21), and its most southerly point is the Winter Solstice on about December 22. The days in the northern hemisphere are the longest on the Summer Solstice, and shortest on the Winter Solstice. So the crossing of the Sun at the equinoxes and solstices mark the beginning of the seasons in the Earth's northern hemisphere, and the location of these points on the celestial sphere are given in Table 1.1.

Right ascension and declination provide celestial bearings in the equatorial coordinate system. Another celestial coordinate system is the horizon, or horizontal, coordinate system that employs great-circle angles measured with respect to the observer's zenith and horizon. The zenith is located above your head, directly away from the center of the Earth. It is the point of intersection of the celestial sphere with the upward prolongation of the observer's plumb line, whose bob is drawn to the terrestrial center by gravity. If a plane is extended outward from the observer's feet, perpendicular to the plumb line, it intersects the celestial sphere in a great circle known as the horizon. Celestial objects are only visible if they are above the horizon. The altitude measures the angular distance from the horizon to the object in question, along a great circle that intersects the object and the zenith. The azimuth, an angle measured along the horizon, provides the second dimension to this coordinate system. Angles are used to designate celestial positions in both types of celestial coordinate systems.

Table 1.1 Celestial positions of the equinoxes and solstices

Position	Right ascension α (2000.0)	Declination δ (2000.0)
Vernal (Spring) equinox	0^{h}	$0°$
Summer solstice	6^{h}	$23° 26' 21.4''$
Autumnal (Fall) equinox	12^{h}	$0°$
Winter solstice	24^{h}	$-23° 26' 21.4''$

1.4 Measuring Angle and Size

Astronomers measure angles in degrees, designated by the superscript °, and there are 360° in a circle. They also use the second of arc, or arc second, denoted by the symbol ″, and the minute or arc, or arc minute, abbreviated by ′, as a units of angle. The units mimic a clock with 60 s in a minute, or $60'' = 1'$. A full degree of angle contains 60 min of arc, so $1° = 60' = 3{,}600''$.

Mathematicians use a different unit of angular measurement called a radian. The radian is the ratio between the length of an arc and its radius. The ratio of linear size to distance is expressed in radians, where an angle of one radian, when viewed from the center of a circle, results in an arc on that circle equal to the radius of the circle (Fig. 1.3). That is, the radian unit of angular measure is defined such that an angle of one radian subtended from the center of a unit circle produces an arc length of one.

A full circle subtends 2π rad and 360°, where $\pi = 3.141592654$, so $1 \text{ rad} = 360/(2\pi) = 57.2958°$, and the conversion factors between seconds of arc and radians are:

$$1'' = 4.848 \times 10^{-6} \text{ rad},$$

and

$$1 \text{ rad} = 2.06265 \times 10^{5\,''} = 57.2958°.$$

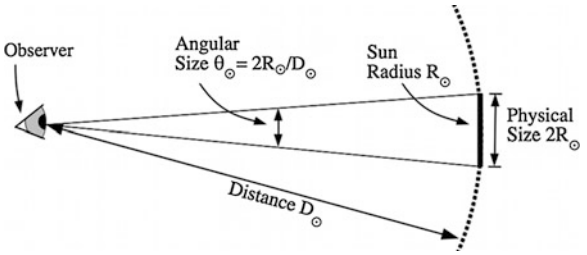


Fig. 1.3 The Sun's angular size and radius The solar radius can be determined from the Sun's angular size and distance. As long as this angle is small, the physical size is only a small arc of a large circle, denoted by the dashed line, and the angular size is the ratio of the physical size to the distance. Astronomers specify this angle as a partial arc of a full circle of 360°; for the Sun it is about 32 min of arc, in which there are 60 min of arc in 1°. This angle has been enlarged to display it in this illustration. In mathematics, the *radian* is the standard unit of angular measure. It describes the angle subtended by a circular arc as the length of the arc divided by the radius of the arc. When the arc length is equal to the arc radius, the angle is 1 rad. We can convert between the two methods of describing angles by noting that the circumference of a circle is 2π times its radius; therefore 1 rad is equal to $360°/(2\pi)$, or 57.2958°. For the Sun, the angular size $\theta_\odot = 2R_\odot/D_\odot$ radians, where $R_\odot$ denotes the Sun's radius and the mean distance of the Sun, $D_\odot$, is 1 AU. The observed angular size of the Sun corresponds to a radius of 695.5 million meters

A telescope can be used to measure the angular size, θ_{size}, of a celestial source if the angular resolution of the telescope is smaller than the angular size of the source. If the source distance, D, is known, then one can infer its linear extent, L, perpendicular to the line of sight, or if spherical its radius, R, using the angular size and the relation:

$$\theta_{\text{size}} = \frac{L}{D} = \frac{2R}{D} \text{ rad.}$$

Example: Measuring the size of the Sun

The linear radius of the Sun, denoted by $R_\odot$, can be determined from the Sun's angular diameter, denoted by $\theta_\odot$, using

$$\theta_\odot = \frac{2R_\odot}{D_\odot} \text{ rad}$$

where the mean distance between the Earth and the Sun has a value of $D_\odot = 1$ AU $= 1.496 \times 10^{11}$ m. As illustrated in Fig. 1.3, this expression uses the mathematician's radian unit of angular measurement, and it has to be converted to astronomical measurements of angle. The radian is the ratio between the length of an arc and its radius.

The mean equatorial angular diameter of the Sun is $\theta_\odot = 31.97$ min of arc $= 31.97' = 1{,}918.2''$, and this is equal to 0.009299 rad, since 1 rad = 206265 s of arc. So the radius of the Sun is given by

$$R_\odot = \frac{\theta_\odot D_\odot}{2} = 6.956 \times 10^8 \text{ m,}$$

which is 109 times the radius of the Earth.

By the way, the angular diameter of the Sun is about the same angle as that subtended by the thumb when viewed at arm's length. In one of those fascinating coincidences, the angular diameter of the Sun is also about the same as the angular diameter of the Moon, which is much closer to us and smaller in radius than the Sun. Because of this similarity in angular size, the Moon can pass in front of the Sun during a total solar eclipse, blocking out the sunlight.

1.5 The Locations of the Stars are Slowly Changing

While most stars sweep by as the Earth rotates, a star that is aligned with our planet's rotation axis, at the north celestial pole, seems to remain placed in an unchanging location at 90° north declination. The Earth's northern rotation axis, for example, now points close to Polaris, also known as the North Star or the Pole

Star, which would lie approximately overhead when viewed from the Earth's geographic North Pole. The latitude of any location in the Earth's Northern Hemisphere is equal, within about 1°, to the angular altitude of Polaris. The uncertainty is due to the fact that Polaris is not exactly at the north celestial pole, where the north end of the Earth's rotation axis pierces the night sky.

We can locate Polaris by following the line joining the two stars farthest from the handle of the Big Dipper, which accounts for the phrase "follow the drinking gourd" used by southern slaves escaping to the northern parts of the United States. Mariners have also used the North Star for navigation, to find the direction of north and the latitude of their ship.

Nevertheless, everything in the universe is in a state of perpetual change, and the locations of the so-called fixed stars on the celestial sphere are no exception. Their change in position is related to the Earth's elongated shape (Focus 1.2), which has sent the Earth into a wobbling rotation that resembles a spinning top. This causes a very slow change of the celestial positions of the north celestial pole, the Pole Star and all the other stars, called precession. The changing positions of bright stars on the celestial sphere were first observed by Hipparchus, a Greek astronomer who lived in the second century BC (Hipparchus, 125 BC); the telescope was not invented until 17 centuries after Hipparchus established the stellar positions using his eyes.

Focus 1.2 The elongated shape of the Earth

The Earth isn't precisely spherical in shape. It has a slight bulge around its equatorial middle and is flattened at its poles, with a shape more like an egg than a marble or billiard ball. This elongated, oblate shape is caused by the Earth's rapid rotation. The outward force of rotation opposes the inward gravitational force, and this reduces the pull of gravity in the direction of spin. Since this effect is most pronounced at the equator, and least at the poles, the solid Earth adjusts into an oblate shape that is elongated along the equator.

An ellipse of eccentricity, e, and major axis, a_e, which is rotated about the polar axis, defines the Earth's reference ellipsoid at sea level. The planet's equatorial radius is a_e, and its polar radius, a_p. They are given by:

$$a_p = a_e(1-f) = a_e(1-e^2)^{1/2},$$

where the flattening factor $f = (a_e - a_p)/a_e$ is related to the eccentricity, e, by $e^2 = 2f - f^2$.

The mean surface radius of the Earth, $\langle a \rangle$, is given by:

$$\langle a \rangle = \left(a_e^2 a_p\right)^{\frac{1}{3}} \approx 6.371 \times 10^6 \,\mathrm{m},$$

which is the radius of a sphere of volume equal to the Earth ellipsoid. Geophysicists use another definition of mean radius given by $(2a_e + a_p)/3$.

The radius, r, of the surface of the Earth geoid at any latitude ϕ is given by

$$r = a_e(1 - f \sin^2 \phi).$$

Two of the primary constants of the International Astronomical Union are (Kaplan 2005):

$$\text{Equatorial radius of the Earth} = a_e = 6.3781366 \times 10^6 \text{ m},$$

and

$$\text{Flattening factor for Earth} = f = 0.0033528197 = 1/298.25642.$$

These values of a_e and f give a polar radius for the Earth of $a_p = 6.356752 \times 10^6$ m, and the difference between the equatorial and polar radius is 21,385 m or about 21 km.

The world geodetic system, which is the basis of terrestrial locations obtained from the Global Positioning System, or GPS for short, uses an Earth ellipsoid with $a_e = 6.378137 \times 10^6$ m and $f = 1/298.257223563$.

The changing locations of celestial objects are caused by the gravitational action of the Moon, Sun and planets on the spinning, oblate Earth. As a result of this gravitational torque, the Earth's rotation axis is constantly changing with respect to a space-fixed reference system.

The precessional motion of the Earth's rotation axis is caused by the tidal action of the Moon and Sun on the spinning Earth. That is, because the Moon and the Sun lie in the ecliptic plane, which is inclined by 23.5° to the plane of the Earth's Equator, they exert a gravitational force on the Earth's equatorial bulge. This causes the rotation axis to sweep out a cone in space, centered at the axis of the Earth's orbital motion and completing one circuit in about 26,000 years (Fig. 1.4).

So the Earth is not placed firmly in space; instead it wobbles about causing the identity of the Pole Star to gradually change over time scales of thousands of years. The northern projection of the Earth's rotation axis is currently within about 0.75° of Polaris and will move slowly toward it in the next century. After that, the north celestial pole will move away from Polaris and, in about 12,000 years, the Earth's rotation axis will point to within 5° of the bright star Vega.

The slow conical motion of precession carries the Earth's Equator with it; as that Equator moves, the two intersections between the celestial equator and the Sun's path, or ecliptic, move westward against the background stars. One of these intersections is the Vernal, or Spring, Equinox, from which right ascension is measured. This equinox point moves forward (westward) along the ecliptic at the rate of about 50 s of arc, denoted $50''$, per year, which is equivalent to 3.33 s per year.

As the Earth's rotational axis precesses, declinations also change, through a range of 47°, or twice 23.5°, over 26,000 years.

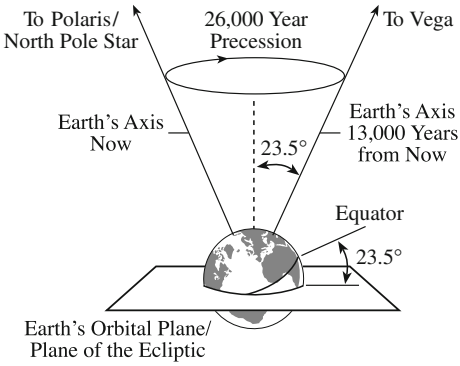


Fig. 1.4 Precession The Earth's rotation axis traces out a circle on the sky once every 26,000 years, sweeping out a cone with an angular radius of about 23.5°. The Greek astronomer Hipparchus (c. 146 BC) discovered this precession in the second century BC. The north celestial pole, which marks the intersection of this rotation axis with the northern half of the celestial sphere, now lies near the bright star Polaris. However, as the result of precession, the rotational axis will point toward another bright north star, Vega, in roughly 13,000 years. This motion of the Earth's rotational axis also causes a slow change in the celestial coordinates of any cosmic object

Because the long period, 26,000 year, conical motion of the Earth's rotation axis is caused by the gravitational action of the Moon and Sun, it is called lunisolar precession. Simon Newcomb (1835−1909) derived the detailed theory for computing the corrections to astronomical coordinates for this precession (Newcomb 1895).

In addition to this steady, progressive motion, there are small, periodic variations in both precessional speed and axial tilt caused by the gravitational action of the planets on the Earth's equatorial bulge. The most important term in this nutation, first observed by James Bradley (1693−1762), induces an 18.6 year periodic wobble in the precessional motion with a size of $17''$ in the direction of precession and $9''$ perpendicular to it (Bradley 1748).

Because of positional changes caused by precession and nutation, the equinox, or reference date, must be given when specifying the right ascension or declination of any cosmic object. The standard epoch that is now in used for celestial positions is:

$$\text{J2000.0} = \text{2000 January 1.5} = \text{JD2451545.0},$$

where JD denotes Julian Date and the prefix J denotes the current system of measuring time in Julian centuries of exactly 36,525 days in 100 years, with each day having a duration of 86,400 s.

The combination of lunisolar and planetary precession is called general precession, and the astronomical constants at standard epoch J2000.0 include (Kaplan 2005):

$$\text{General precession in longitude} = \rho = 5{,}028.796195'' \text{ per Julian century.}$$

The nutation term, N, for that epoch is

$$\text{Constant of nutation} = N = 9.2052331''.$$

Kaplan (2005) provides modern formulas for precession and nutation. They describe the transformation of celestial coordinates from one date to another, as a function of time since a reference epoch.

The astronomical constants also include an aberration constant, κ, which accounts for the observed position shift of an astronomical object in the direction of the Earth's motion (Focus 1.3). The aberration constant at the standard epoch J 2000.0 is

$$\text{Constant of aberration} = \kappa = 20.49552''.$$

Focus 1.3 Stellar aberration

As the Earth orbits the Sun, the stars all appear to be shifted in the direction of motion, a phenomenon called stellar aberration, described by James Bradley (1693−1767) in 1728. Because the speed of light is finite, the apparent direction of a celestial object detected by a moving observer is not the same as its geometric direction at the time. For stars, the normal practice is to ignore the correction for the motion of the celestial object, and to compute the stellar aberration due to the motion of the observer. This also gave Bradley a means to improve on the accuracy of previous estimates for the speed of light (Bradley 1728).

The magnitude $\Delta\theta$ of stellar aberration depends on the ratio of the velocity of the observer, V, to the speed of light, c, and the angle, θ, between the direction of observation and the direction of motion. The displacement, $\Delta\theta$, in the sense of apparent minus mean place, is given by

$$\Delta\theta \approx \frac{V}{c}\sin\theta - \frac{1}{2}\left(\frac{V}{c}\right)^2 \sin 2\theta + \left(\frac{V}{c}\right)^3 \left(\sin\theta\cos^2\theta - 0.33\sin^3\theta\right) + \ldots$$

As the Earth orbits the Sun, it is moving at a velocity of approximately 30 km s^{-1}, and the speed of light c $\approx$ 300,000 km s^{-1}, so the term of order V/c is 10^{-4} rad or 20 s of arc, denoted $20''$, and the term $(V/c)^2$ has a maximum value of 0.001 s of arc. Bradley (1728) used aberration observations to determine the speed of light as approximately 183,000 miles per second or 294,500 km s^{-1}.

The constant of aberration, κ, at standard epoch J2000.0 is given by

$$\text{Constant of aberration} = \kappa = \frac{2\pi a}{Pc}(1-e^2)^{-\frac{1}{2}} = \frac{V}{c} = 0.9365 \times 10^{-4} \text{ rad},$$

which is equivalent to:

$$\kappa = 20.49552''$$

Here the constant $\pi \approx 3.14592654$, the mean distance between the Earth and the Sun is $a = 1$ AU $= 1.49598 \times 10^{11}$ m, known as the astronomical unit, $e = 0.01671$ is the mean eccentricity of the Earth's orbit, $P = 3.1558 \times 10^7$ s is the length of the sidereal year, $c = 2.997925 \times 10^8$ m s^{-1} is the speed of light, and one radian $= 2.062648 \times 10^5$ ″.

When the observer is moving directly at the star, θ is zero and there is no aberration shift at all. The shift achieves its greatest value of about 20.5″ when the observer's motion is perpendicular to the direction of the star, with $\theta = 90°$.

1.6 What Time is It?

There are two ways of keeping time in common use today. One is atomic time, the basis of the Système International, abbreviated SI, second, and the other is based on the rotation of the Earth (Kaplan 2005; Seidelmann 2005). The SI second is the fundamental unit of atomic time, which is specified by atomic clocks that use cesium-beam and other atomic frequency standards to an accuracy of 1.5×10^{-14}, or to the fourteenth decimal place (Essen 1969). The frequency standards form a standard timescale known as International Atomic Time, abbreviated TAI for Temps Atomique International. The time distributed by the Global Positioning System, or GPS, remains at a constant offset with International Atomic Time. On 21 November 2010, TAI-GPS = 19 s.

The clocks we use in daily life are set to the Earth's rotation with respect to the Sun. It establishes our daily rhythm, from sunrise to sunset and back to sunrise again. The 24 h solar day is the time it takes for the Sun to make one circuit around the local sky.

By definition:

$$1 \text{ solar day} = 24 \text{ h} = 1{,}440 \text{ min} = 86{,}400 \text{ s}.$$

This is known as the unit of Sun time, or solar time. Also by definition, the Julian century has 36,525 solar days and one day is defined as 86,400 s of International Atomic Time, or TAI. So:

$$1 \text{ Julian year} = 365.25 \text{ solar days} = 8{,}766 \text{ h} = 525{,}960 \text{ min} = 31{,}557{,}600 \text{ s}$$

Solar time is the basis of Universal Time, abbreviated UT, which has been defined to be as uniform as possible despite variation in the Earth's rotation. The worldwide system of Civil Time, and the clock on your wall or the watch on your

arm, are synchronized to Coordinated Universal Time, denoted UTC, which has been corrected for small variations in the Earth's rotation, but the atomic clocks in satellites and the time in the Global Positioning System, abbreviated GPS, are not corrected in this way.

Because of irregularities in the Earth's rotation and the lengthening of its rotation period due to Moon-induced tidal friction, this Sun time does not advance at a uniform rate, and it increasingly lags behind the SI-second time scale. UTC is therefore a hybrid time scale using the SI second on the spinning Earth as a fundamental unit, but subject to occasional 1 s adjustments. The difference between atomic time and universal time, or TAL − UTC is an integral number of seconds, which increases by 1 whenever a leap second is introduced into UTC, so the two kinds of clock share the same seconds tick. When necessary, the adjustments to UTC are introduced at the end of June or December, by international agreement.

Universal Time is equivalent to the standard Civil Time for 0° longitude, which is defined to be the Prime Meridian at Greenwich, England. By specifying the longitude of any other location of the terrestrial globe, we can exactly infer the local Sun time. For example, the Sun will be overhead at noon in Boston about 4 h 44 min later than noon in Greenwich; because the longitude of Boston is 71.06° = 4.733 h, where 360° equals 24 h. And in the same way, noon in New York City will occur about 10 min later than in Boston, because New York City is slightly west of Boston.

The world has been divided into standard time zones based on about 1 h, or 15°, increments in longitude, so our watches differ from others in hourly increments and are slightly out of synchronism with the Sun. Standard time is the result of synchronizing clocks in different geological locations within a time zone. Blaise (2000) has described the creation of standard time.

Terrestrial Time (TT) is the modern time standard used for time measurements of astronomical observations from the surface of the Earth. The unit of TT is the SI second. The *Astronomical Almanac* uses TT in the tables of positions, or *Ephemerides*, of the Sun, Moon and planets as seen from the Earth. TT is slightly ahead of atomic time, and can be approximated by

$$\mathrm{TT} \approx \mathrm{TAI} + 32.184\,\mathrm{s},$$

which is equivalent to

$$\mathrm{TT} \approx \mathrm{GPS} + 51.184\,\mathrm{s}$$

With the advent of atomic clocks and the exact targeting of planetary spacecraft, measurements of time were further refined with the introduction of the Barycentric Dynamical Time, or TDB for short. It was adopted to take into account the relativistic time dilation (Sect. 2.3) when calculating orbits and astronomical ephemeris, and it applies to the solar-system-barycentric reference frame. The barycenter is the center of mass of two or more orbiting bodies. The

difference between TDB and yet another Barycentric Coordinate Time, or TCB, is about 16.6 s, and

$$\text{Mean ratio of the TCB second to the TDB second} = 1 - L_B,$$

where $L_B = 1.550519767 \times 10^{-8}$, an exceedingly small number.

In every day life we use Sun time, based on the solar day that is exactly 24 h long, in solar time. This solar day is the interval between two successive passages of the Sun across an observer's meridian. Astronomers also use star time, or sidereal time. A sidereal day is the time between successive passages of a star across the local meridian, and this star day is about 4 min less than a solar day.

1.7 Telling Time by the Stars

Astronomers use another sort of time, called sidereal time, to know when and how to point their telescopes to view a particular star or any other cosmic object. The term sidereal is derived from the Latin *sidus* meaning "star." As with solar time, this star time is based on the Earth's rate of rotation, but measured relative to the fixed stars rather than the Sun.

The sidereal day is the time it takes for a star – or any other celestial objects – to proceed from its highest point in the sky one day to its highest point the next day. The Earth makes one rotation about its axis in a sidereal day, but during that time it moves a short distance along its orbit around the Sun. At the end of a sidereal day, the Earth therefore needs to rotate a little more before the Sun reaches its highest point (Fig. 1.5). A solar day is therefore about 4 min longer than the sidereal day.

To be exact:

$$1 \text{ sidereal day} = 23\,\text{h}\ 56\,\text{m}\ 4.09\,\text{s} = 23.93447\,\text{h} = 86{,}164.1\,\text{s},$$

where the hours, minutes and seconds are in solar time.

Observatories have two kinds of clocks that tell either the local solar time or the local star time. The two kinds of time can also be determined using clocks or time simulators on the web.

At any moment, the Local Sidereal Time equals the right ascension, designated α, of a celestial object on the local meridian, so this time tells an observer when a given celestial object with a particular right ascension can be seen, provided its declination is in the observable range for the observer's latitude.

Example: When is a celestial object visible in your part of the sky?

For an object with an observable declination, the time at which it can be observed depends upon the object's right ascension, α, and the local sidereal time, abbreviated LST. The object crosses the local meridian when $\alpha = \text{LST}$, and may be visible for several hours before and after this time. However, a terrestrial clock or watch is geared to the Sun rather than the

stars, so it does not keep local sidereal time. A terrestrial clock is equal to the local sidereal time only at midnight of the Autumnal Equinox, about September 23, and thereafter the local sidereal time gains 2 h on the terrestrial clock for each succeeding month.

The geographic longitude, denoted λ, of the observer relates the Local Sidereal Time (LST) to the Greenwich Sidereal Time, or GST for short, at the Prime Meridian by

$$\mathrm{LST} = \mathrm{GST} - \lambda(\mathrm{observer}),$$

where the longitude is measured positive westward and can be converted into time using 24 h = 360°. The local meridian is an imaginary half circle stretching from the horizon due north, through the zenith, to the horizon due south. The zenith is directly overhead, an extension of a plumb line from the center of the Earth

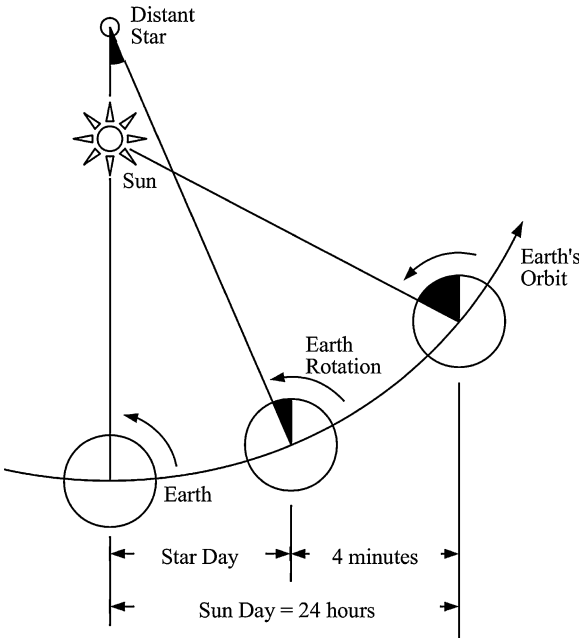


Fig. 1.5 Sun time and star time The Sun reaches its highest point in the daytime sky, its culmination, at noon, and this happens every 24 h in a solar day. A distant star returns to its highest point in the night sky every sidereal day of 23 h 56 min 04 s, which is the unit of star time. The Earth rotates once around its axis in one sidereal day, but during that time the Earth has moved along its orbit around the Sun (*bottom*). The star (*small circle, top*) and the Sun (*below star*) are at culmination, crossing the local meridian, at just one time (*left*). After a sidereal day, the Earth has to rotate for another 3 min 56 s before the Sun reaches its highest point (*right*). A solar day is therefore nearly 4 min longer than a sidereal day

through the observer to the celestial sphere. The Prime Meridian is the local meridian for the old Royal Observatory in Greenwich, England.

We define the Local Hour Angle (LHA) of any object on the celestial sphere to be the time since it last crossed the meridian, or

$$\text{LHA (object)} = \text{LST} - \alpha\,(\text{object}) = \text{GST} - \lambda(\text{observer}) - \alpha(\text{object}).$$

The Local Sidereal Time is equal to zero when the Vernal, or Spring, Equinox is on the local meridian.

The solar and star times are related by:

One mean solar day = 24 h 03 min 56.555 s of mean sidereal time

One mean solar day = 1.0027379 mean sidereal days,

One mean sidereal day = 23 h 56 min 04.09 s of mean solar time

One mean sidereal day = 0.99726957 mean solar days.

1.8 Optical Telescopes Observe Visible Light

Telescopes collect and magnify electromagnetic radiation from a cosmic object, and bigger telescopes provide two advantages. They gather more radiation than a smaller telescope, permitting the detection of fainter objects and providing a brighter image of any cosmic object for analysis. Big telescopes also provide greater angular resolution, which is the ability to see the separation between objects that are close together. Better resolution permits observation of finer detail on the object emitting the radiation. Kitchin (2013) provides a thorough discussion of telescopes and the techniques of using them to observe the cosmos.

Regardless of what cosmic object a telescope is pointed at, the object's radiation that carries information to the Earth travels in rays that are parallel to one another. A telescope's lenses and/or mirrors are used to focus and collect visible radiation, or light. They are described by the science of optics; therefore the study of visible light from cosmic objects is called optical astronomy. There are two types of optical telescopes, the refractor and the reflector, which respectively use a lens and a mirror to gather and focus optically visible light (Fig. 1.6). A telescope's lens bends the incoming rays by refraction, focusing them to a point where they meet, called the *focal point*. A curved mirror reflects the incoming rays, sending them to the focal point.

In a refractor, light is bent by refraction at the curved surface of a lens, called an objective, toward a focal point where the different rays of light meet. If we place a detector at the focal point, in the plane parallel to the lens, we can record an image of whatever the telescope is observing. The distance from the lens to the focal point is called the *focal length*, which determines the overall size of the image. The critical thing is the diameter, or aperture, of the light-gathering lens.

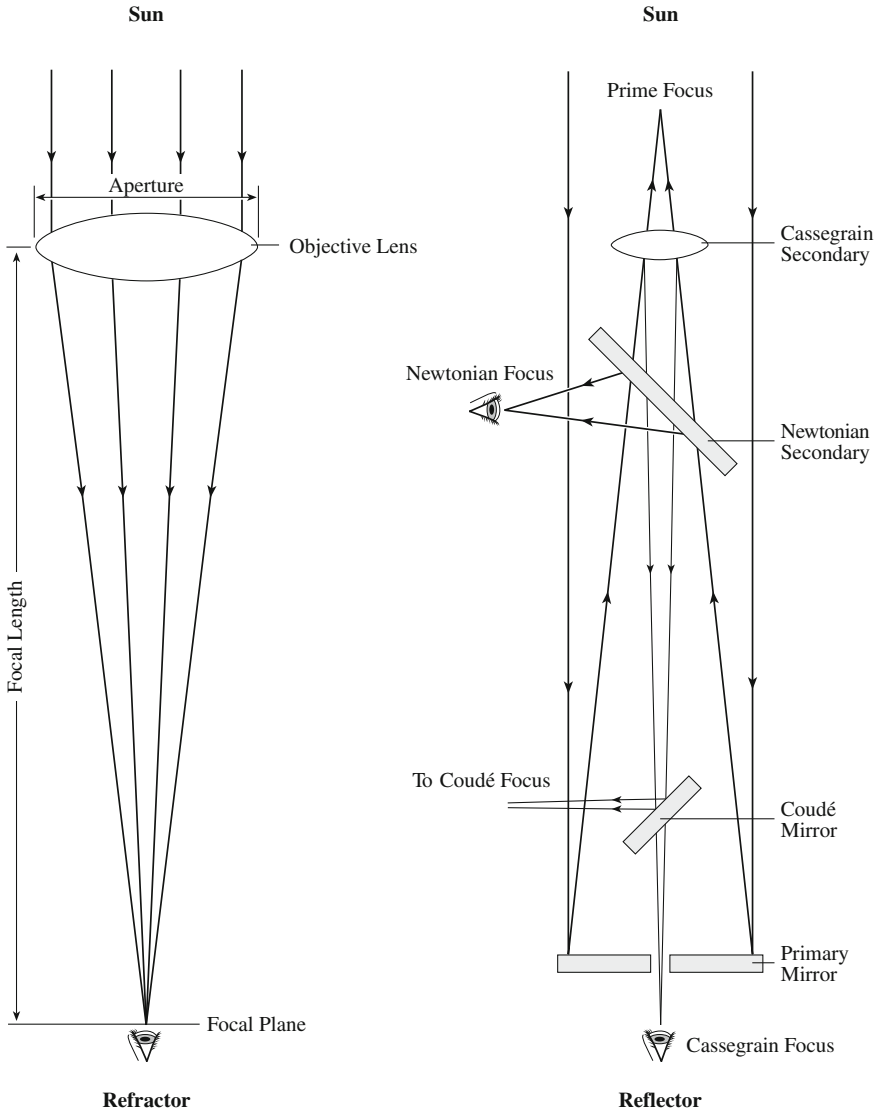


Fig. 1.6 Telescopes Light waves that fall on the Earth from a distant object are parallel to one another, and are focused to a point by the lens or mirror of a telescope. Early telescopes were refractors (*left*). The curved surfaces of the convex objective lens bend the incoming parallel light rays 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. In 1670, the English physicist Isaac Newton (1643−1727) constructed the first reflecting telescope (*right*), which used a large concave, or parabolic, primary mirror to collect and focus light. A small flat secondary mirror, inclined at an angle of 45° to the telescope axis, reflected 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

The larger the aperture, the more light is gathered and the finer the detail that can be resolved.

The other type of telescope, the reflector, uses a concave mirror with a parabolic shape to gather and focus the light. The prime focus is back in the path of the incoming light; so secondary mirrors are sometimes used to reflect the light to another place of observation. There are three types of secondary mirrors called the Cassegrain, Coudé, and Newtonian mirrors, which can focus light to different locations.

Professional astronomers place electronic detectors at the focal point of telescopes. These detectors generate digitized signals that are analyzed, manipulated and recorded in a computer. A Charge-Coupled Device, or CCD for short, might be used to efficiently detect the radiation and form an image. Nowadays, CCDs are used in this way in everything from digital cameras to the *Hubble Space Telescope*. In some cases, diffraction gratings are used to separate the incoming radiation into its component wavelengths, dispersing it into fine wavelength intervals to form a spectrum that can then be recorded by a CCD.

The diameter of this primary mirror determines the telescope's light-gathering ability and resolution. The amount of radiation that can be collected is proportional to the area of the mirror, and the square of its diameter.

Optical telescopes are even named by the diameter of their mirror. The 2.5 m (100 in.) Hooker Telescope at the Mount Wilson Observatory, California and the nearby 5 m (200 in.) Hale Telescope at the Palomar Observatory, California are of great historical importance. Recent large optical telescopes include the four Very Large Telescopes, each of 8.2 m (323 in.) effective aperture, located at the Paranal Observatory, Chile, the two 10 m (400 in.) Keck telescopes at the Mauna Kea Observatory in Hawaii, the 10.4 m (410 in.) Gran Telescopio Canarias on the Canary Islands and the Large Binocular Telescope located on Mount Graham in Arizona. It consists of two 8.4 m (330 in.) mirrors on a binocular mount.

The ability to resolve details is called the resolving power of a telescope. It is specified by the angular resolution, θ_{res}, a quantity that depends on the diameter, D_T, of the telescope lens or mirror and the wavelength, λ, of observation. The mathematical expression is:

$$\theta_{\text{res}} = \frac{\lambda}{D_T} \text{ rad},$$

or

$$\theta_{\text{res}} = 2.06265 \times 10^5 \, \frac{\lambda}{D_T} \text{ s of arc},$$

where 1 rad is equivalent to 57.2957795° and 206265″ (2.06265×10^5 s of arc), and there are 3,600 s of arc in a degree. This equation tells us that a bigger lens or

mirror provides finer angular resolution at a given wavelength, and that longer wavelengths require larger telescopes to achieve the same resolution as a smaller telescope at shorter wavelength.

Example: What is the smallest source detected on the Sun in optically visible light?

Due to atmospheric obscuration, the effective angular resolution of a typical ground-based optical telescope is about $1'' = 4.848 \times 10^{-6}$ rad. That is pretty good, for its comparable to seeing the details on a coin with a diameter of 0.5 cm from 1,000 m away. At the Sun's mean distance of 1.496×10^{11} m, this corresponds to structures that are 748 km across, about the distance from Boston to Washington, D.C. and about three-quarters the size of France.

The resolving power of a telescope operating at the wavelengths that we can detect with our eye, at a yellow wavelength of about 6×10^{-7} m, is about $0.124/D_T''$ if D_T is in meters. By way of comparison, the typical angular resolution of the unaided human eye is about $60''$, so the eye acts like a lens with a diameter of about 0.002 m, or only 2 cm. However, some people have sharper vision than the average.

Turbulence in the atmosphere limits the resolution of any telescope operating at visible wavelengths to about 1 s of arc; therefore, the angular resolution cannot be improved by building an optical telescope larger than about 0.12 m in diameter. Similar atmospheric variations cause the stars to "twinkle" at night. This atmospheric limitation to angular resolution at visible wavelengths is called seeing. The best seeing, of $0.2''$ in unusual conditions, is found at only a few sites in the world, and optical observatories are located in 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 the limits of the atmosphere.

Optical astronomy began about four centuries ago, in 1609, when Galileo Galilei (1564−1642) turned the newly invented spyglass, or telescope, toward the night sky, and discovered four previously unknown moons that circle Jupiter. He also resolved small craters on the Moon, and detected numerous stars in the Milky Way that cannot be seen by the unaided eye (Galilei 1610). His rudimentary telescope was a refractor with a lens whose diameter was only 0.04 m, or 1.6 inches and a little smaller than your hand. The angular resolution of his telescope at a visual wavelength of 6×10^{-7} m was $3.1''$, and an angular resolution more than about 10 times better than this cannot be achieved with any optical, or visible light, telescope on the planet.

A bigger lens or mirror also collects more light than a smaller one, permitting the detection of fainter sources. The human eye, for example, is severely limited by its inability to gather light. The eye, or rather the brain fed by the eye, can store the images for no more than a few tenths of a second. That's fortunate because if it stored an image for much longer, we couldn't watch movies. Light collected using telescopes can be stored for hours or more. This was first done using photographic plates and more recently by using electronic chips.

1.9 Telescopes that Detect Invisible Radiation

Telescopes of different designs are used to detect cosmic radiation outside the optically visible wavelengths. They are used to observe otherwise invisible x-ray, ultraviolet, infrared and radio wavelengths.

Relatively long radio waves are detected by radio telescopes, also known as radio antennas, whose shapes of are similar to the reflecting mirror of an optical telescope. The main reflector, called a dish, is a parabolic metal surface that gathers the incoming radio waves, reflecting and focusing them to an electronic receiver at the reflector focus. This receiving system converts the intensity of the incoming radio signal to numbers that are then transmitted to a computer. The data is stored in the computer as a matrix of numbers and is then manipulated to form images.

Our angular resolution equation also applies at radio wavelengths where very big telescopes are required to achieve significant angular resolution. At a radio wavelength of 0.1 m, an angular resolution of $1''$ requires a telescope with a diameter of 20 km. The advantage of radio signals is that the atmosphere does not distort them, or limit the angular resolution. We can observe the radio universe on a cloudy day, and during the day or night, just as radio signals are used to communicate with satellites at any time, even when it rains or snows outside.

Since radio waves are millions of times longer than those of light, a radio telescope needs to be at least a million times bigger than an optical telescope to obtain the same resolving power. For this reason, the first radio telescopes provided a very myopic, out-of-focus view. But this limitation was soon overcome when radio astronomers built successively larger telescopes, culminating in the 100 m, fully steerable parabolic dish at Effelsberg, West Germany. Its best angular resolution is about $10''$. This may be the largest steerable radio dish that can be built, but a novel way of building an even larger dish was to cover the floor of a valley with metal screen, producing a 305 m dish in Arecibo, Puerto Rico. This antenna relies on the rotation of the Earth to bring different regions of the sky into view.

Nowadays, relatively small radio telescope separated by large distances (called baselines) are combined and coordinated electronically. This results in radio images that are as sharp as optical ones (Fig. 1.7). Because it is spread out, an array of small telescopes has the property that is crucial for high resolving power – namely, great size relative to wavelength. The technique is known as interferometry because it analyzes how the waves detected at the telescopes interfere when they are added together, so the interferometer is an interference meter. A simple example combines the signals from a pair of telescopes with a computer to reconstruct the waves and create the image.

The sensitivity of an interferometric array is determined by the combined areas of the individual elements, and not by their separations. For example, the two-element interferometer can resolve details that are much finer than its component radio telescopes can by themselves. But the interferometer's collecting area is only twice that of the two individual components, so its sensitivity is just twice as great. Many

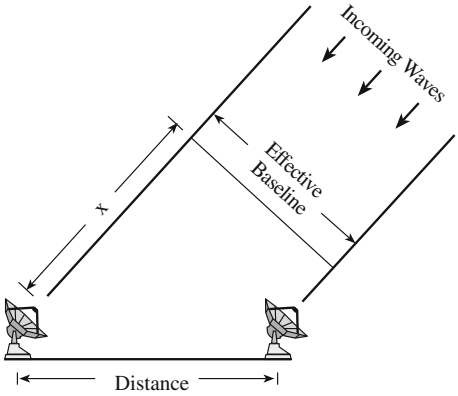


Fig. 1.7 Interferometer When incoming radiation approaches the Earth at an angle, the crests will arrive at two separated telescopes at slightly different times. This delay in arrival time is the distance X divided by the speed of light. If X is an exact multiple of the wavelength, then the waves detected at the two telescopes will be in phase and add up when combined. If not, they will be out of phase and interfere. The angular resolution of such an *interferometer*, or interference meter, is equal to the wavelength divided by the effective baseline. When the object being observed is directly overhead, the effective baseline is equal to the distance between the two telescopes

two-element telescope pairs are therefore combined in a full-fledged radio array to gather more radio radiation. The Very Large Array, abbreviated VLA, is an example of a modern interferometric radio array that is used to observe the cosmos. Quirrenbach (2001) provided a review of optical interferometry at visible wavelengths.

Even better angular resolution is obtained using radio signals recorded at widely separated radio telescopes when they are observing the same cosmic object at the same time. The recorded signals can be combined to effectively turn the Earth into a giant radio telescope with transcontinental interferometer baselines and the sharpest vision of any telescope on the Earth or in space.

Example: Very long baseline interferometry

At a radio wavelength of $\lambda = 0.1$ m, the diameter, D_T, of a telescope needed to obtain a resolution of $\theta_{\text{res}} = 1''$ is $D_T = 2.063 \times 10^5 \ \lambda/\theta_{\text{res}} \approx 20{,}000$ m $= 20$ km. But an interferometer with radio telescopes placed on opposite sides of the Earth will have a baseline of up to twice the planet's radius, or about 10^7 m. That would give it an angular resolution of about $0.0016''$ at a wavelength of 0.1 m.

Radio telescopes do not provide the only window on the cosmos. There are also invisible gamma rays, x-rays, ultraviolet and infrared telescopes. Radiation coming from celestial objects at these wavelengths is absorbed in our atmosphere and must be collected by telescopes in satellites that orbit the Earth above its

atmosphere. All of these space telescopes measure the intensity of the incoming signal and convert these measurements into radio transmissions that are sent to radio telescopes and receivers on the ground. NASA has, for example, launched four large *Space Telescopes*, the Great Observatories named *Hubble*, *Compton*, *Chandra* and *Spitzer*, which respectively operate at visible, gamma ray, x-ray and infrared wavelengths (Focus 1.4). The European Space Agency, abbreviated ESA, has built and launched the large infrared *Herschel* telescope that is used at far infrared and sub-millimeter wavelengths. The American, European and Japanese space agencies have sent a host of satellites into space with telescopes designed to observe specific cosmic phenomena, such as activity on the Sun that is observed from telescopes on *SOHO*, *Hinode*, *STEREO*, and *SDO*. The distances and motions of stars have been observed from *HIPPARCOS*, and the cosmic microwave background radiation delineated from *COBE*, *WMAP* and *PLANCK*.

Focus 1.4 The great observatories

NASA has sent four large telescopes into orbit around the Earth, designed to study the universe in both optically visible light and non-visible forms of radiation. The first in the series was the *Hubble Space Telescope*, or *HST* for short, launched on April 24, 1990 and operating mainly in visible light, the second was the *Compton Gamma Ray Observatory* (*CGRO*), launched into Earth orbit on April 5, 1991, the third was the *Chandra X-ray Observatory*, or *CXO*, launched on July 23, 1999, and the fourth was the infrared *Spitzer Space Telescope*, or *SST*, launched on August 23, 2003.

The *HST* is the largest visible-light, astronomical telescope ever put into space, with a 2.4 m (94.5 in.) primary mirror. The angular resolution of optically visible light telescopes in space is not limited by the Earth's atmosphere, so the *HST* has an angular resolution as good as $0.05''$. It has observed newly formed galaxies when the universe was less than half its present age, contributing to our understanding of the age and evolution of the Cosmos, watched super-massive black holes consuming the material around them, and helped astronomers determine how a mysterious "dark energy" has taken over the expansion of the universe. The telescope is named for the American astronomer Edwin Hubble (1889−1953) who demonstrated that spiral nebulae are galaxies like our own Milky Way and found that galaxies move away from us at speeds that increase with their distance.

Chandra investigates the high-energy, x-ray regions of the universe from objects such as active galactic nuclei, black holes, clusters of galaxies, dark matter, galaxies, neutron stars, pulsars, quasars, supernova remnants, supernovae, and white dwarfs. It is named after the Indian-American astrophysicist and Nobel laureate, Subrahmanyan Chandrasekhar (1910−1995).

Instruments aboard *Compton* detected thousands of energetic, brief gamma-ray bursts, as well as gamma rays from black holes, pulsars, quasars, and supernovae. The gamma ray observatory was named after the American physicist Arthur H. Compton (1892−1962).

The *Spitzer* telescope obtains images and spectra of the infrared energy, or heat, radiated by cosmic objects at wavelengths between 3 and 180 microns, where 1 micron is 10^{-6} m or one-million of a meter. The telescope has a 0.85 m (33.5 in.) primary mirror. It provides information on the formation, composition and evolution of planets, stars and galaxies, and is named after the American astrophysicist Lyman Spitzer, Jr. (1914−1997). Most satellite telescopes, including *Hubble*, circle our planet outside the Earth's atmosphere while remaining nearby to send observations down to the ground by radio signals, but *Spitzer* revolves around the Sun, trailing behind the Earth in its orbit, to avoid heat from the Earth and the Moon. A supply of liquid helium initially cooled the instrument to almost absolute zero so the telescope's heat radiation would not interfere with its detectors.

NASA's next Great Observatory, the *James Webb Space Telescope*, or *JWST* for short, is working on a launch date of 2018. It will have a large mirror, 6.5 m (21.5 feet) in diameter, and will operate at infrared wavelengths that will permit the observations of distant galaxies that formed in the early universe, as well as nearby planet-forming regions.

In the meantime, on May 14, 2009, the European Space Agency, abbreviated ESA, launched the *Herschel* infrared telescope; its primary mirror is 3.5 m (138 in.) in diameter. It is named for the English astronomer Sir William Herschel (1738−1822), the discoverer of the infrared spectrum and the planet Uranus, and his sister and collaborator Caroline (1750−1848). *Herschel* is used to see deep into star-forming regions, galactic centers and planetary systems.

Astronomy from space has several advantages over ground-based observations. The weather in space is always perfectly clear and the atmosphere does not blur images obtained from telescopes in space. Furthermore, a large telescope is not needed to observe the short ultraviolet and x-ray wavelengths with high angular resolution. The aperture must be only 0.002 m across to achieve an angular resolution of $1''$ at an extreme ultraviolet wavelength of 10^{-8} m and only 0.00002 m for the same resolution at soft x-rays of 10 keV in energy and a wavelength of 10^{-10} m. Moreover, in space the sky is truly dark, and observations do not need to be limited to the night.

Example: The size of optical, x-ray and radio telescopes, and resolving features on the Sun or Moon

The angular resolution, θ, of a telescope with a diameter D_T operating at a wavelength λ is given by $\theta = \lambda/D_T$ rad, where 1 rad $= 2.06265 \times 10^5\,''$ and $''$ denotes second of arc with $1'' = 4.848 \times 10^{-6}$ rad. The angular resolution of most ground-based, visible light, or optical, telescopes is limited by turbulence in the Earth's atmosphere to about $1''$, or under excellent seeing conditions at remote mountain tops to perhaps $0.25'' = 1.212 \times 10^{-6}$ rad,

which is also the angular resolution of the Solar Optical Telescope on the *Hinode* mission. The size of a telescope required to give this angular resolution at a yellow wavelength of $\lambda = 580$ nm is $D_T = \lambda/\theta \approx 0.5$ m. The *Hinode* mission also has an x-ray instrument, operating at a wavelength of $\lambda = 1.24 \times 10^{-10}$ m, corresponding to 10 keV in photon energy (see Sect. 2.6 for definition of photon energy). The size of the x-ray telescope for the same angular resolution is $D_T = \lambda/\theta \approx 0.0001$ m. A ground-based radio telescope is not limited in angular resolution by the atmosphere, but to achieve an angular resolution of $0.25''$ at a radio wavelength of $\lambda = 1$ m, the telescope diameter would have to be $D_T = \lambda/\theta \approx 10^6$ m or a million meters, comparable in size to a large country and about one sixth the radius of the Earth.

The smallest linear size, L, that can be resolved on the Sun with any telescope with this angular resolution is $L = D \times \theta$, where the mean distance between the Earth and the Sun is $D = 1$ AU $= 1.496 \times 10^{11}$ m, and therefore $L \approx 181$ km for an angular resolution of $0.25'' = 1.212 \times 10^{-6}$ rad. By way of comparison, the angular resolution of the human eye is about $60''$ so the smallest sunspot that the eye can resolve has a size of $L = \text{AU} \times 60'' \times 4.848 \times 10^{-6} \approx 4.35 \times 10^7$ m $\approx 6.8\ R_E$ where the radius of the Earth $R_E = 6{,}378$ km. The tallest mountain that can be observed at the limb, or edge, of the Moon with the unaided human eye will have a height $H = D \times \theta$ where the mean distance of the Moon is 3.844×10^8 m and for $\theta = 60''$ we have a mountain height of $H = 1.12 \times 10^5$ m, much taller than Mount Everest, whose elevation is 8.848×10^3 m. In other words, we could not see mountains on the Moon until telescopes were used.

1.10 Units Used by Astronomers and Astrophysicists

By any terrestrial standard, the scale of astronomical objects is enormous in mass, luminosity, distance, size and age. Astronomers and astrophysicists use the Sun's values of these quantities as benchmark units that reflect their large amount. Any solar value is denoted by a subscript symbol $\odot$, a circle with a dot in the center.

The mass of the Sun, denoted by the symbol $M_\odot$, is often used as the unit of celestial mass. Its value is

$$M_\odot = 1.989 \times 10^{30}\,\text{kg},$$

where kg denotes a kilogram, or 1,000 g. The mass of most stars lies in a narrow range of between 0.1 $M_\odot$ and 100 $M_\odot$, just as the mass of newborn children varies by a relatively small amount. Galaxies typically contain about 100 billion, or 100 thousand million, stars, so the stellar mass of a galaxy is about $10^{11}\ M_\odot$, but a

galaxy often contains even more mass in invisible dark matter that lies beyond the visible stars.

The rate at which radiation carries energy away from a cosmic object is known as its luminosity, designated by L. Luminosity has the units of energy per unit time, which is also the unit of power. The SI unit of luminosity is a joule per second (J s^{-1}), where joule is the unit of energy, and one watt of power is equal to one joule per second, or 1 J s^{-1} = 1 W.

The unit of luminosity used by astronomers is often the Sun's luminosity, denoted $L_\odot$. Its value is:

$$L_\odot = 3.828 \times 10^{26}\,\mathrm{J\,s^{-1}},$$

where J s^{-1} denotes joule per second. Stars vary by many orders of magnitude in their luminosity, from 0.001 $L_\odot$ to a million $L_\odot$ or from 10^{-3} $L_\odot$ to 10^6 $L_\odot$. The luminosity of a galaxy is roughly 10^{11} $L_\odot$.

In astronomy and astrophysics, temperatures are measured on the kelvin scale, named after Lord Kelvin (1824−1907) who proposed it (Kelvin 1848). This temperature unit is written kelvin, with a lower case k, and assigned the unit symbol capital K. The freezing temperature of water is 273.15 K and the boiling temperature of water at sea level on Earth is 373.15 K. The kelvin scale is an absolute, thermodynamic temperature scale where absolute zero is the temperature at which all thermal motion ceases. Nothing can move at a temperature of 0 K. For conversion to the degrees Centigrade, denoted by C, and degrees Fahrenheit, abbreviated by F, we have K = C + 273.15 = (5F/9) + 255.22, with C = K − 273.15 and F = (9 K/5) − 459.67 = (9C/5) + 32.

Astronomers use the astronomical unit (AU) as the unit of distance within the solar system. It is the mean distance between the Earth and the Sun, with a value of

$$1\,\mathrm{AU} = 149597870691\,\mathrm{m} \approx 1.496 \times 10^{11}\,\mathrm{m}.$$

To be exact, astronomers now use the speed of light, denoted by the lower case letter c, as a defining constant for distance, with

$$\text{Speed of light} = c = 299792458\,\mathrm{m\ s^{-1}} \approx 2.9979 \times 10^8\,\mathrm{m\ s^{-1}},$$

with the derived value of the light travel time, τ_A, for 1 AU

$$\tau_A = 499.0047863852\,\mathrm{s} \approx 499\,\mathrm{s}.$$

This is the time it takes for radiation to travel from the Sun to the Earth.

The unit of stellar size is the radius of the Sun, denoted $R_\odot$, given by

$$R_\odot = 6.955 \times 10^8\,\mathrm{m}.$$

The supergiant stars can be as large as 1,000 $R_\odot$ in radius, giant stars are about 10 times smaller, and the smallest stars that shine by nuclear reactions are about 0.1 $R_\odot$ in radius. Collapsed white dwarf stars are about as big as the Earth, whose radius is 6.378 × 10^6 m and about 0.01 $R_\odot$.

The distances between stars are expressed in a unit called the parsec, or pc for short, where

$$1\text{pc} = 3.08567758128 \times 10^{16}\,\text{m} = 206265\,\text{AU} \approx 3.086 \times 10^{16}\,\text{m},$$

which is also equivalent to

$$1\text{pc} = 3.261564\,\text{light-year} \approx 3.26\,\text{light-year},$$

where a light-year is the distance light travels in one year at the speed of light, c, or

$$1\,\text{light-year} = 9.4607304726 \times 10^{15}\,\text{m} \approx 9.461 \times 10^{15}\,\text{m}.$$

As we shall subsequently see, the term parsec is derived from the parallax method of determining distance, where one parsec is a parallax of one second of arc.

Example: How far away and long ago was starlight emitted?

The Sun is located at a mean distance of 1 AU $= 1.496 \times 10^{11}$ m. Traveling at the speed of light $c = 2.9979 \times 10^{8}$ m s^{-1}, it takes a time $\tau = \text{AU}/c = 499$ s for radiation to travel from the Sun to the Earth, so the sunlight we see this very moment was emitted 499 s ago. The nearest star other than the Sun is Proxima Centauri, and it takes 4.286 light-years for starlight to travel from this star to the Earth. Since 1 year $= 3.156 \times 10^{7}$ s, the ratio of the distance to Proxima Centauri and the distance to the Sun is about $4.286 \times 3.156 \times 10^{7}/499 \approx 271{,}000$. The first stars were formed shortly after the big bang, which occurred about 13.7 billion years ago. So these first stars are located at a distance of about 13.7 billion light-years. Using the conversion of 1 parsec $=$ 1 pc $=$ 3.26 light-years and l light-year $= 9.461 \times 10^{15}$ m, these first stars are located at a distance of about 4.2×10^{9} pc and 1.3×10^{26} m, almost 1 million billion times further away than the Sun.

The extent of a galaxy is measured in units of kiloparsec, or kpc for short, where 1 kpc $= 10^{3}$ pc. The distance between the Sun and the center of our Galaxy is, for example, about 8.5 kpc. Nearby galaxies are separated by about a million parsecs, denoted as a megaparsec and abbreviated Mpc, where 1 Mpc $= 10^{6}$ pc $= 3.0857 \times 10^{22}$ m. The nearest large spiral galaxy, Andromeda or M 31, is located at a distance of 0.78 Mpc, while a very remote galaxy might be at a distance of a billion parsec, denoted as a gigaparsec and abbreviated as Gpc, where 1 Gpc $= 10^{9}$ pc.

Astronomers use the second, abbreviated by the lower case letter s, for small time scales and the year, or yr for short, for large ones. The orbital period of the Earth around the Sun is one year, with a value of

$$1\,\text{yr} = 3.156 \times 10^{7}\,\text{s} = 365.25\,\text{d},$$

Table 1.2 Principal SI units and their conversion to corresponding c.g.s. units

Quantity	SI units	Conversion to c.g.s. units
Length[a]	Meter (m)	100 centimeters (cm)
Mass[b]	Kilogram (kg)	1,000 grams (g)
Speed[c]	Meter per second (m s^{-1})	100 centimeters per second (cm s^{-1})
Energy[d]	Joule (J)	10,000,000 erg = 10^7 erg
Power	Watt (W) = J s^{-1}	10,000,000 erg s^{-1} = 10^7 erg s^{-1}
Temperature	Kelvin (K)	degrees Centigrade = C = K−273
Magnetic flux density	Tesla (T)	10,000 gauss (G) = 10^4 G
Force	Newton (N) = kg m s^{-1}	100,000 dyn = 10^5 dyn
Pressure	Pascal (Pa) = N m^{-2}	10 dyn cm^{-2}

[a] One nanometer (nm) is 1 nm = 10^{-9} m; 1 Å = 1 Å = 10^{-10} m; 1 mile = 1.609 km; and 1.0 in. = 2.54 cm

[b] 1 ton = 2,240 lb = 1.016 047 × 10^3 kg ≈ 10^3 kg

[c] Speed is the magnitude of velocity. 1 m s^{-1} = 3.600 km h^{-1} = 2.237 miles per hour, and 1 mile per hour = 44.704 cm s^{-1}

[d] The energy of high-energy particles and x-ray radiation are often expressed in units of kilo-electron volts, or keV, where 1 keV = 10^3 eV = 1.602 × 10^{-16} J, or MeV = 1,000 keV, with 1 eV = 1.602 176 487 × 10^{-19} J ≈ 1.602 × 10^{-19} J

and 1 day = 1 d = 86,400 s. A pulsar might rotate with a period of 1 s, the Earth is 4.6 billion years, or 4.6 Gyr, old and the expanding universe originated about 14 billion years ago. The International System of Units (Système International, abbreviated SI, is used in this book. It includes the length unit of meter (m), the mass unit of kilogram (kg), and the time unit of second (s). The SI units of energy, luminosity, temperature, and magnetic field strength, are joule (J), watt (W, or J s^{-1}), kelvin (K), and tesla (T), respectively.

Many astronomers and astrophysicists have often used, and still use, the c.g.s. units of centimeter (cm), gram (g), second (s) in their professional papers. Conversions from the SI units to the c.g.s. units are given in Table 1.2.

1.11 Physical Constants

Since the fundamental laws of physics apply throughout the universe, the physical constants used in the equations that describe these laws are thought to be universal and unvarying in space or time. These constants include the speed of light, c, the Newtonian gravitational constant, designated G, the Boltzmann constant, denoted by k and the Planck constant, designated h.

The speed of light, c, is independent of the frame of reference in space or time. It provides an upper speed limit affecting any object in the entire universe, and has a value of

$$c = \text{speed of light} = 299{,}792{,}458\,\text{m s}^{-1} \approx 2.9979 \times 10^8\,\text{m s}^{-1}$$

The gravitational constant G enters into Newton's universal law of gravitation, expressing the force of gravity $F = GMm/D^2$ between two bodies of mass M and m separated by a distance D. If the law of gravitation is universal, applying to all objects in the universe, than G must be independent of time, position, mass and the nature of the bodies. It has a value of

$$\begin{aligned} G &= \text{gravitational constant} = 6.67428 \times 10^{-11}\ \text{N m}^2\ \text{kg}^{-2} \\ &\approx 6.674 \times 10^{-11}\ \text{N m}^2\ \text{kg}^{-2}. \end{aligned}$$

The Boltzmann constant k and the Planck constant h are quantum constants used to describe the macroscopic properties of exceedingly small things. Atoms in "thermal" equilibrium are characterized by a single temperature T, and an energy $E \approx kT$, where

$$k = \text{Boltzmann constant} = 1.3806504 \times 10^{-23}\ \text{J K}^{-1} \approx 1.3806 \times 10^{-23}\ \text{J K}^{-1}.$$

Astrophysicists measure temperature on the kelvin scale, where the symbol K denotes degrees kelvin. The Boltzmann constant k appears in the statistical description of the velocities of atoms that are in thermal equilibrium and in the ideal gas law that specifies the pressure of a gas at a given temperature.

The Planck constant h is used to describe the particle, or photon, nature of radiation, specifying the photon energy $E = h\nu$ for radiation of frequency ν, where

$$h = \text{Planck constant} = 6.626069 \times 10^{-34}\ \text{J s} \approx 6.6261 \times 10^{-34}\ \text{J s}.$$

This constant appears in the description of thermal (blackbody) radiation, and when describing the interaction of radiation with matter.

Additional universal constants, listed at http://physics.nsit.gov/, include the electric and magnetic constants:

$$\begin{aligned} \varepsilon_0 &= \text{electric constant} = \text{permittivity of vacuum} = (10^{-9}/(36\pi)) \\ &= 8.8542 \times 10^{-12}\ \text{F m}^{-1} \end{aligned}$$

$$\begin{aligned} \mu_0 &= \text{magnetic constant} = \text{permeability of vacuum} = 4\pi \times 10^{-7} \\ &= 1.2566 \times 10^{-6}\ \text{N A}^{-1}. \end{aligned}$$

Chapter 2
Radiation

2.1 Electromagnetic Waves

The physical perception of the universe is governed almost solely by the electromagnetic radiation received from cosmic objects. This radiation carries energy and moves through space in periodic waves at the speed of light, designated by the lower case letter c. The speed of light in empty space is a universal constant, independent of reference in space and time. The radiation is called *electromagnetic* because it propagates by the interplay of oscillating electric and magnetic waves.

Our understanding of electricity and magnetism is founded upon the experimental investigations of the English scientist Michael Faraday (1791–1867), who invented the first rotating electric motor and discovered electromagnetic induction, the principle behind the electric transformer and generator (Faraday 1843). His experiments led Faraday to propose that electromagnetic forces extend into empty space around charged bodies, electrical conductors, and magnets; these invisible forces are now called *electromagnetic fields*.

The Scottish mathematician and theoretical physicist James Clerk Maxwell (1831–1879) was able to express Faraday's results in a precise mathematical form, now known as *Maxwell's equations* (Maxwell 1865). These four partial differential equations depend on variations of the force fields in four dimensions – three for space and one for time.

In regions with no charge or currents, such as a vacuum, Maxwell's equations describe sinusoidal *electromagnetic waves* (Focus 2.1). The waves described by this electromagnetic wave equation have a speed equal to the speed of light, leading Maxwell to comment, "light is an electromagnetic disturbance propagated through the field according to electromagnetic laws." The changing magnetic field creates a changing electric field that, in turn, creates a changing magnetic field. The electric and magnetic field directions are orthogonal to each other and to the direction of travel.

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_2, © Springer-Verlag Berlin Heidelberg 2013

Focus 2.1 Plane waves of electromagnetic radiation

For the electric field, $\boldsymbol{E}$, and the magnetic field, $\boldsymbol{B}$, in free space, Maxwell's equations take the form (Maxwell 1865):

$$\begin{aligned} \nabla \cdot E &= 0 \\ \nabla \cdot B &= 0 \\ \nabla \times E &= -\frac{\partial B}{\partial t} \\ \nabla \times B &= \mu\varepsilon \frac{\partial E}{\partial t}, \end{aligned} \tag{2.1}$$

where $\nabla\cdot$ is the divergence operator, with units of m^{-1}, $\nabla\times$ is the curl operator, with units of s^{-1}, and $\partial/\partial t$ is the partial derivative with respect to time, t. These equations can be written as second-order partial differential equations:

$$\begin{aligned} \left(\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2}\right) &= 0 \\ \left(\nabla^2 B - \frac{1}{c^2}\frac{\partial^2 B}{\partial t^2}\right) &= 0 \end{aligned} \tag{2.2}$$

that describes the propagation of electromagnetic waves through a medium or a vacuum. Here c is the speed of light in the medium. In a vacuum:

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = 2.99792458 \times 10^8\,\mathrm{m\,s^{-1}}, \tag{2.3}$$

where the electric constant, or vacuum permittivity, $\varepsilon_0 = 8.854187817 \times 10^{-12}\ \mathrm{F\,m^{-1}} \approx 8.854 \times 10^{-12}\,\mathrm{F\,m^{-1}}$, and the magnetic constant, or vacuum permeability, $\mu_0 = 1.256632061 \times 10^{-6}\,\mathrm{N\,A^{-2}} \approx 1.257 \times 10^{-6}\,\mathrm{N\,A^{-2}}$. In a medium with refractive index n:

$$c = \frac{1}{n\sqrt{\varepsilon_0 \mu_0}} = \frac{1}{\sqrt{\mu\varepsilon}}, \tag{2.4}$$

where ε is the electric permittivity of the medium and μ is the magnetic permeability of the medium.

There are sinusoidal, plane-wave solutions of these equations written as:

$$\begin{aligned} E(r,t) &= E_0 \cos(\omega t - k \cdot r) \\ B(r,t) &= B_0 \cos(\omega t - k \cdot r), \end{aligned} \tag{2.5}$$

where t is the time variable, and the angular frequency, ω, is related to the wavenumber, k, by the dispersion relation

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda}, \tag{2.6}$$

where λ is the wavelength, $\nu = \omega/2\pi = c/\lambda$ is the frequency in s^{-1}, and the constant $\pi \approx 3.154159$.

The energy flux, S, and energy density, U, in the plane wave are

$$S = \frac{c}{8\pi}\sqrt{\frac{\varepsilon}{\mu}}E_0^2 \tag{2.7}$$

which is directed along the direction of wave propagation, and

$$U = \frac{\varepsilon E_0^2}{4\pi}. \tag{2.8}$$

Maxwell realized that c equals the speed of light, which others had previously measured, and concluded that light is a form of electromagnetic radiation.

In common with any wave, electromagnetic radiation has a wavelength, usually denoted by the lowercase Greek letter lambda, λ. The wavelength is the distance between successive crests or successive troughs (Fig. 2.1). Different types of electromagnetic radiation differ in their wavelength, although they propagate at the same speed. Like waves on water, electromagnetic waves have crests and troughs; but, unlike water waves, electromagnetic waves can propagate in vacuous empty space.

In SI units, the wavelength is measured in meters, abbreviated m. Other units of wavelength are the nanometer, or nm for short, where 1 nm $= 10^{-9}$ m, the Ångström, abbreviated Å where 1 Å $= 10^{-10}$ m $= 0.1$ nm, and the micron, denoted μ where 1 μ $= 10^{-6}$ m. Radio astronomers might specify the wavelength in meters or centimeters, abbreviated cm where 1 cm $= 10^{-2}$ m.

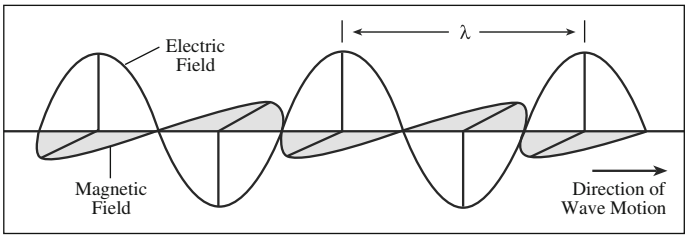


Fig. 2.1 Electromagnetic waves All forms of radiation consist of electrical and magnetic fields that oscillate at right angles to each other and to the direction of travel. They move through empty space at the speed of light. The separation between adjacent wave crests is called the *wavelength* of the radiation and usually is designated by the lowercase Greek letter lambda, λ

Sometimes radiation is described by its frequency, denoted by the lower case Greek letter nu written v. The frequency indicates how fast the radiation oscillates, or moves up and down. The frequency of a wave is the number of wave crests passing a stationary observer every second, measured in Hertz, abbreviated Hz. One Hertz is equivalent to one cycle per second, or $1\ \text{Hz} = 1\ \text{s}^{-1}$. Radio astronomers use a frequency unit of megahertz, abbreviated MHz, where $1\ \text{MHz} = 10^6$ Hz. Radio stations that transmit frequency modulated, or FM, signals, are denoted by their call letters and the frequency of their broadcasts in MHz. The frequency range of FM radio broadcasts is 88–108 MHz. Amplitude modulated, or AM, radio signals are broadcast in several bands of frequency between 0.150 and 30 MHz.

Electromagnetic waves all travel through empty space at the same constant speed – that is, the *speed of light* in a vacuum $c = 299{,}792{,}458\ \text{m s}^{-1}$, or about $2.9979 \times 10^8\ \text{m s}^{-1}$. The product of wavelength, λ, and frequency, v, is equal to the speed of light, c, or

$$\lambda \times v = c. \tag{2.9}$$

So, radiation at shorter wavelengths has a higher frequency and a longer wavelength corresponds to a lower frequency. Any electromagnetic wave, regardless of wavelength or frequency, travels though empty space at the speed of light, and it is the maximum speed possible (Focus 2.2).

Focus 2.2 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. This was first inferred from observations of Jupiter's moon Io in the 17th century. The King of France had directed Giovanni Domenico Cassini (1625–1712), director of the Paris Observatory, to use such observations to improve knowledge of terrestrial longitude and maps of France. Both the Danish astronomer Ole Rømer (1644–1710), who also worked at the observatory, and Cassini noticed a varying time between eclipses of Io by the Jupiter (Rømer 1677). Although the time between Io eclipses was approximately 42 h, it varied by an amount of up to 22 min.

Both astronomers concluded that it was not the orbit of Io around Jupiter that changed, but the time it took Jupiter's light to travel from Io to the Earth, which depended on the Earth's position in its orbit around the Sun. When the Earth was on the side of its orbit that is closest to Jupiter, the observed eclipse period for Io was shortest, and when the Earth was on the opposite side of its annual orbit around the Sun, Io's apparent eclipse period was largest.

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 between the longest and shortest observed Io period, or a velocity of

$c = 2$ AU/22 min and approximately 2.27×10^8 m s^{-1}, where 1 AU $= 1.496 \times 10^{11}$ m is the mean distance between the Earth and the Sun. At Cassini's time this distance was not well known (see following Sect. 2.5).

Jupiter orbits the Sun at a mean distance of 5.203 AU and Jupiter's natural satellite Io orbits Jupiter with a period of 1.769 Earth days. It is eclipsed by the planet with that period. Observations of changes in the eclipse period were interpreted as differences, Δt, in the time, t that light takes to travel from Jupiter to Earth. When Jupiter is furthest from Earth, its distance will be 6.203 AU, since the Earth is 1.00 AU from the Sun, and when Jupiter is closest to the Earth, the giant planet's distance will be 4.203 AU. So the total change in Io's apparent orbital period, from longest to shortest, will be $\Delta t = (6.203–4.203)/c = 2$ AU$/c = 998$ s $= 16.63$ min $= 0.0116$ Earth days, where $c = 2.9979 \times 10^8$ m s^{-1} is the speed of light. This is a relatively small change in the Io's actual orbital period of 1.769 Earth days.

The English astronomer James Bradley (1693–1762) unexpectedly discovered the aberration of starlight about half a century later. It is a change in the observed position of a star that depends on the ratio of the velocity of the Earth and the speed of light. Using then current estimates for the Earth's orbital motion around the Sun, Bradley used his aberration measurements to infer a speed of light of about 294,500 m s^{-1} (Bradley 1728).

More refined laboratory measurements during subsequent centuries indicated that light is always moving at a constant speed with the precise velocity of $c = 299{,}792{,}458$ m s^{-1}. Light emitted by any star moves at this speed through empty space for all time. It never stops or slows down, and it never comes to rest. Nothing outruns light; it is the fastest thing around.

Electromagnetic radiation has no way of marking time, and it can persist forever. As long as its rays pass through empty space and encounter no atoms or charged particles like electrons, it will survive unchanged. Radiation emitted from any star or galaxy today might therefore travel for all time in vacuous space, bringing its message forward to the end of the universe. Astronomers on Earth intercept just a small part of this radiation, which is streaming away from both known and unknown objects located throughout the cosmos.

2.2 The Electromagnetic Spectrum

Most of us remember the colorful display of a rainbow, which is sunlight bent into separate wavelengths by droplets of water. In the mid-17th century, the English scientist Isaac Newton (1642–1727) showed that sunlight could also be broken into its colors using a prism – a specially cut chunk of glass (Newton 1671, 1704). Furthermore, each color could not be divided into other colors. A crystal

Table 2.1 Approximate wavelengths of colors[a]

Color	Wavelength (nm = 10^{-9} m = 10 Å)
Violet	420
Blue	470
Green	530
Yellow	580
Orange	610
Red	660

[a] Approximate wavelengths good to about 10 nm

chandelier or compact disk also displays the spectrum of visible light, arranging the colors by their different wavelengths.

From short to long waves, the colors in the spectrum of visible light correspond to violet, blue, green, yellow, orange and red (Table 2.1). Their wavelengths might be specified in nanometers, abbreviated nm, where 1 nm = 10^{-9} m or in Ångströms, abbreviated Å, where 1 Å = 0.1 nm = 10^{-10} m. Light from the Sun or an incandescent light bulb often is called white light, because it contains all of the colors, whereas black denotes the absence of color when we see no light.

The *electromagnetic spectrum* describes the types and wavelengths of electromagnetic radiation (Fig. 2.2). From short wavelengths to long ones, this spectrum includes gamma rays, x-rays, ultraviolet radiation, visible light, infrared radiation and radio waves (Table 2.2).

Our eyes detect a narrow range of wavelengths, which include the visible colors. It comprises just one small segment of the much broader electromagnetic spectrum. This band of light is also termed visible radiation, to distinguish it from invisible radiation that cannot be seen with the eye. The radiation we can see is also known as optically visible radiation, since the science of optics is used to describe the lenses and mirrors used to detect the light. The most intense radiation of the Sun and many other stars is emitted at these optically visible wavelengths, and our atmosphere permits it to reach the ground. Other types of radiation, like the invisible x-rays, are absorbed in our atmosphere and do not reach the Earth's surface.

The invisible domains include infrared and radio waves – with wavelengths longer than that of red light – and the ultraviolet (UV) rays, x-rays, and gamma (γ) rays, whose wavelengths are shorter than violet light. They all are *electromagnetic waves* and part of the same family, and they all move in empty space at the speed of light, but we cannot see them.

Gamma rays are the shortest and most energetic electromagnetic waves. Their wavelengths are as small as the nucleus of an atom, or about 10^{-15} m, and their waves are so energetic that they can pass through a thick iron plate.

The *x-ray* region of the electromagnetic spectrum extends from a wavelength of 100 billionth (10^{-11}) of a meter, which is about the size of an atom, to the short-wavelength side of the ultraviolet. The German physicist Wilhelm Röntgen (1845–1923) discovered x-rays, producing them with an electrical discharge in a

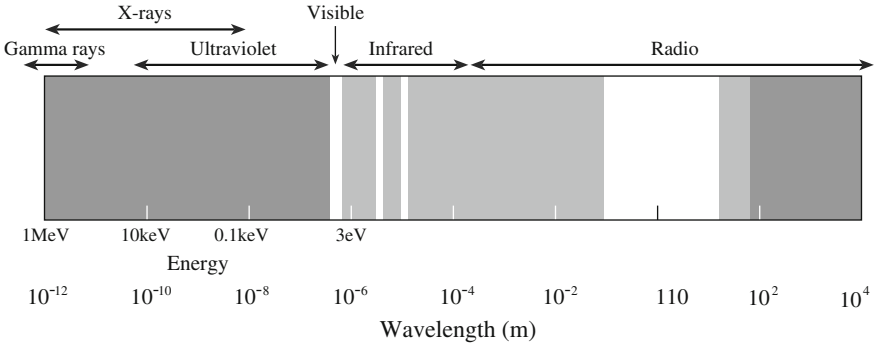


Fig. 2.2 Electromagnetic spectrum Radiation from cosmic objects can be emitted at wavelengths from less than 10^{-12} m to greater than 10^{4} m, where m denotes meters. The visible spectrum that we see with our eyes is a very small portion of the entire range of wavelengths. Lighter shading indicates a greater transparency of the Earth's atmosphere to cosmic radiation. It only penetrates the Earth's atmosphere at visible and radio wavelengths, respectively represented by the narrow and broad white areas. Electromagnetic radiation at short gamma ray, X-ray and ultraviolet wavelengths, represented by the dark areas, is absorbed in our atmosphere. The universe is now observed in these spectral regions from above the atmosphere in Earth-orbiting satellites

Table 2.2 The electromagnetic spectrum

Region	Wavelength range (m)
Radio	10^{-3}–10^{3}
Microwave	10^{-3}–1
Infrared	7×10^{-7}–10^{-3}
Visible	4×10^{-7}–7×10^{-7}
Ultraviolet	10^{-8}–4×10^{-7}
x-ray	10^{-11}–10^{-8}
Gamma ray	Less than 10^{-11}

glass vacuum tube (Röntgen 1896). He used the energetic x-rays to penetrate skin and muscle, detecting human bones and revolutionizing medicine.

The wavelength of *ultraviolet radiation*, abbreviated UV, is just a bit longer, between 10^{-8} and 4×10^{-7} m, with extreme ultraviolet radiation, denoted EUV, lying in the short wavelength part of this range. Most of the ultraviolet radiation from the Sun is absorbed in our air, but prolonged exposure to the amount that reaches the ground can burn your skin.

The infrared part of the electromagnetic spectrum is located at wavelengths between 7×10^{-7} and 10^{-3} m. The German-born English astronomer William Herschel (1738–1822) discovered infrared radiation when he put a beam of sunlight through a prism to spread it into its spectral components. He noticed that an unseen portion of sunlight warmed a thermometer placed beyond the red edge of the visible spectrum (Herschel 1800). The thermometer recorded higher

temperatures in the invisible infrared sunlight than in normal visible sunlight. Herschel called them *calorific rays* because of the heat they generated. The term *infrared* did not appear until the late 19th century.

Humans "glow in the dark," emitting infrared radiation, but we cannot see the heat; it is outside our range of vision. Soldiers can locate the enemy at night by using night-vision goggles with infrared sensors that detect their heat, and spy satellites use infrared telescopes to detect heat radiated by rocket exhaust and by large concentrations of troops and vehicles.

Atmospheric molecules such as carbon dioxide and water vapor absorb infrared radiation. So the air that looks so transparent to our eyes is opaque to much of the infrared radiation coming from outer space. Telescopes located above part of the atmosphere, on the tops of mountains in dry climates, can catch some of the incoming infrared radiation before it is completely absorbed. The atmosphere similarly blocks the heat radiation from the Earth's surface, keeping it warmer than it would otherwise be. This warming of the ground is known as the greenhouse effect.

The atmosphere effectively absorbs most of the ultraviolet and infrared radiation from cosmic objects and all of their x-rays and gamma rays, which never reach the ground. To look at the universe at these invisible wavelengths, we must loft telescopes above the atmosphere. This was done first by using balloons and sounding rockets, followed by Earth-orbiting satellites with telescopes that view the cosmos at invisible ultraviolet, infrared, x-ray and gamma ray wavelengths.

Radio waves are between 0.001 and 1,000 m long, too long to enter the eye and not energetic enough to affect vision. The German physicist Heinrich Hertz (1857–1894) discovered *radio waves* by building equipment to both produce and detect the invisible electromagnetic signals (Hertz 1887). The unit of frequency $v = c/\lambda$ is now named the *Hertz* in his honor; this unit is abbreviated Hz. *Microwaves* have wavelengths in the short part of the radio-wave region, between 0.001 and 1.0 m.

Radio waves are the only type of invisible radiation that is not absorbed in the Earth's atmosphere. Radio waves even can pass through rain clouds; therefore, the radio universe can be observed on cloudy days and in stormy weather, just as a home or car radio works even when it is raining or snowing. Cosmic radio waves that are longer than about 10 m are nevertheless reflected by an ionized layer in the Earth's atmosphere, called the ionosphere; so these longer radio waves cannot reach the ground and must be observed from space.

2.3 Moving Perspectives

Motion changes our perspective, and observations depend on our relative motion with respect to the object being observed. These moving perspectives are described using inertial frames of reference, which move at a constant velocity, never accelerating or decelerating. The Dutch physicist Hendrik A. Lorentz (1853–1928)

derived the coordinate transformation of Maxwell's equations from one inertial system to another, showing that the equations are invariant when subjected to this transformation. The Lorentz transformation utilizes a parameter γ, now known as the Lorentz factor, which is given by (Lorentz 1904):

$$\gamma = \left[1 - \frac{V^2}{c^2}\right]^{-\frac{1}{2}} = \left[1 - \beta^2\right]^{-\frac{1}{2}}, \tag{2.10}$$

where V is the relative velocity of the two inertial frames of reference, $\beta = V/c$ and $c = 2.99792458 \times 10^8$ m s^{-1} is the speed of light.

The German-born physicist Albert Einstein (1879–1955) generalized the Lorentz transformation in the *Principle of Relativity* (Einstein 1905a, b), which states that the laws of nature and the results of experiments performed in an inertial frame are independent of the uniform velocity of the system. Einstein additionally proposed that there exists in nature a limiting, invariant speed, the speed of light, c, now known as a universal constant.

The unvarying speed of light was first demonstrated in the late 19th century 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 ether, a wind would be generated, and the observed speed of light would vary, like the speed of a sailboat going with or against the wind. 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 (Michelson and Morley 1887). So the experiment meant that there was no light-carrying ether. It also implied that the speed of light is constant, exactly the same in all directions and at all seasons, and independent of the motion of the observer (Focus 2.3).

Focus 2.3 The Michelson-Morley experiment

Many experiments have been carried out to confirm the unvarying speed of light, but the most famous one was conducted in a basement laboratory at the Case School of Applied Science in Cleveland, Ohio in 1887, when the American scientists Albert A. Michelson and Edward W. Morley attempted to use an interferometer to precisely measure how the speed of light depends on the Earth's motion through space.

Scientists of that time firmly believed in an imaginary luminiferous ether, an invisible, frictionless, and unmoving medium that was supposed to permeate all of space. Its presence explained how light waves could travel at high speed through the apparent emptiness of space, providing the medium in which they propagate. Light was supposed to be transmitted in space by the vibrations of the hypothetical, invisible ether.

If the Earth moved through the stationary ether, a wind would be generated and the observed speed of light would vary, like the speed of an airplane moving with or against the wind. But Michelson and Morley found that there was no detectable difference in the interference pattern produced when a beam of light was sent into the ether wind in the direction of the Earth's motion or directed at right angles to it. Moreover, there was no difference in the measured speed of light when the Earth was traveling toward the Sun and away from it half a year later. That is, Michelson and Morley could measure no difference, Δc, in the speed of light, c, in two perpendicular paths of equal length, in the direction of the Earth's motion or transverse to it, with a precision of $\Delta c/c \leq 0.0001$ (Michelson 1881; Michelson and Morley 1887). Roy J. Kennedy, at the California Institute of Technology, subsequently refined the experiment and improved the measurement precision by a factor of ten (Kennedy 1926; Kennedy and Thorndike 1932).

So the Michelson-Morley experiment meant that there was no light-carrying ether. It also meant that the speed of light is always constant and everywhere the same. In 1907 Michelson was awarded the Nobel Prize in Physics for his optical precision instruments and the spectroscopic and metrological investigations carried out with their aid.

The speed of light, c, enters into Albert Einstein's (1879–1955) *Special Theory of Relativity* through the Lorentz factor $\gamma = [1 - (V/c)^2]^{-1/2}$ for an object moving at an observed 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 moving 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, mass is also relative, and it increases with the speed by the same infamous γ factor.

In the *Special Relativity*, motions and events are described by coordinates in space (x, y, z) and time, t, within an inertial frame of reference that moves at a constant velocity. The length of an object moving with the reference frame of an observer is called the proper length, and the time read in a clock in that frame is the proper time.

If proper time, t, and time interval, Δt, between two events at one location are measured in system K, then the time interval, $\Delta t'$, between the events as measured in system K' moving with uniform velocity V is:

$$\Delta t' = \Delta t \sqrt{1 - \frac{V^2}{c^2}} = \frac{\Delta t}{\gamma}. \tag{2.11}$$

A moving clock will therefore appear to go slower to an observer in the moving system, which is known as *time dilation*.

Time dilation can prolong the decay time of fast-moving, unstable cosmic ray particles by several orders of magnitude, and noticeably lengthen the lifetime of elementary particles produced in man-made particle accelerators.

Atomic clocks have been flown around the world, first eastward and then westward, and compared with the time recorded by a reference atomic clock on the ground. As predicted by *Special Relativity*, the flying clocks lost time (aged slower) during the eastward trip, in the direction of the Earth's rotation, and gained time (aged faster) during the westward trip (Hafele and Keating 1972).

Lengths are also diminished at high speed, shrinking in the direction of motion. For proper spatial separation or length, Δx, in system K, there is a *Lorentz contraction* or shortening, $\Delta x'$, in the moving K' system given by:

$$\Delta x' = \Delta x \sqrt{1 - \frac{V^2}{c^2}} = \frac{\Delta x}{\gamma}. \tag{2.12}$$

Thus, both space and time are relative in the *Special Theory of Relativity*.

Mass is also relative, for it increases with the speed. If a particle or object has rest mass, m_0, in a non-moving frame, the mass increases in the moving one to m' given by:

$$m' = \frac{m_0}{\sqrt{1 - \frac{V^2}{c^2}}} = \gamma m_0. \tag{2.13}$$

The rest-mass energy $E = m_0 c^2$ increases in the moving frame to the energy E' given by:

$$E' = \frac{m_0 c^2}{\sqrt{1 - \frac{V^2}{c^2}}} = \gamma m_0 c^2. \tag{2.14}$$

This expression has been verified in high-energy particle experiments that demonstrate that the energy of a subatomic particle can increase with its speed. The equation also shows that an infinite amount of work would be required to accelerate a particle to the speed of light, with $V = c$, implying that no physical object can move faster than the speed of light in an inertial frame. The mass grows without bound when an object moves as fast as light, and there is nothing that can propel it so fast.

Light is especially difficult to describe using this theory, for any specification of mass, size, or time intervals are undefined when moving at light's speed.

In the *Special Theory of Relativity*, which applies to the non-accelerating and non-gravitational laws of physics, distance is measured by a metric, or line element, ds, that combines space, x, y, z, and time, t. It was first proposed by Einstein's former teacher, Hermann Minkowski (1864–1909) and is given by Minkowski (1908):

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2, \tag{2.15}$$

where the speed of light, c, is used to give the units of space from time and the all-important negative sign indicates time passing. In spherical coordinates r, θ, ϕ the metric ds is written

$$ds^2 = -c^2dt^2 + dr^2 + r^2d\theta^2 + r^2 \sin^2 \theta d\phi^2. \tag{2.16}$$

In this description, two events don't have a uniquely defined separation in either space or time. Instead, they are separated in space-time. So, the concepts of space and time are interwoven.

Space and time manage to join together in a detectable way when objects move exceptionally fast, approaching the speed of light. In these special circumstances, there is no space without time and no time without space; they are fused together. If the motion is fast enough, it will change the size and shape of things, or slow the passing of time. But these effects only become significant at exceptionally high speeds, close to the speed of light. We never encounter these experiences in normal circumstances, and they are not directly applicable to our everyday lives.

2.4 Thermal (Blackbody) Radiation

An ideal thermal radiator is known as a *blackbody*. By definition, a blackbody absorbs all the radiation that falls upon it and reflects none – hence the term black. A black shirt will similarly absorb most of the visible sunlight falling on it and reflects no colors.

Thermal radiation is emitted by a gas in thermal equilibrium, and arises by virtue of an object's heat, or temperature. A single temperature characterizes *thermal radiation*.

Any hot gas that is in thermal equilibrium, with a temperature above absolute zero, will attempt to radiate its energy away. The emission from such a thermal radiator is found at all wavelengths, or frequencies, but with a varying intensity that depends on the temperature (Fig. 2.3). As the temperature increases, more energy is radiated at all wavelengths. Moreover, the wavelength of maximum radiation shifts toward the shorter wavelengths when the temperature rises.

Since the emission of thermal radiation is present at all wavelengths, astronomers say it emits a continuum spectrum. A display of its radiation intensity as a function of wavelength, known as the spectrum, shows no gaps, breaks or sudden increases or decreases. It is an unbroken continuum ascending to peak intensity and then dropping again as the wavelength increases.

No real object emits a perfect thermal, or blackbody, spectrum, but the Sun shines with roughly such a spectrum. It closely matches the radiation spectrum of a blackbody at a temperature of 5,780 K.

The German physicist Max Planck (1858–1947) derived the formula for the spectrum of a perfect absorber, or blackbody, introducing the idea that it radiates energy in fundamental indivisible units, which he called *quanta*, whose energy is

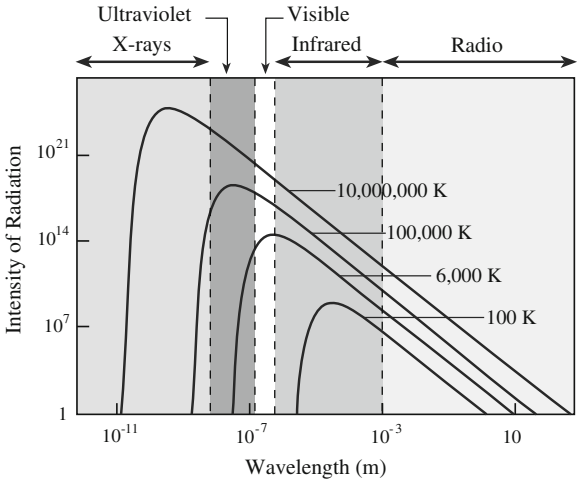


Fig. 2.3 Blackbody radiation The spectral plot of blackbody radiation intensity as a function of wavelength depends on the temperature of the gas emitting the radiation. The German physicist Max Planck (1858–1947) derived the formula that describes the shape and peak of this spectrum in 1900. He proposed that the radiation energy was quantized, which provided a foundation for quantum theory. At higher temperatures the wavelength of peak emission shifts to shorter wavelengths, and the thermal radiation intensity becomes greater at all wavelengths. At a temperature of 6,000 degrees on the kelvin scale, or 6,000 K, the thermal radiation peaks in the visible, or V, band of wavelengths. A hot gas with a temperature of 100,000 K emits most of its thermal radiation at ultraviolet, or UV, wavelengths, whereas the emission peaks in X-rays when the temperature is 1 million to 10 million K

proportional to the frequency of the radiation (Planck 1901, 1910, 1913). The constant of proportionality between the frequency and energy of the radiation is now known as the Planck constant, designated by the lower case letter h. It has a value of $h = 6.626\ 069\ 57 \times 10^{-34}$ J s, or about $h \approx 6.626 \times 10^{-34}$ J s. This marked the beginning of quantum physics, whose history is discussed by Kragh (2002).

Planck found that a blackbody with temperature T emits a *continuum spectrum* of radiation characterized by a brightness distribution, $B_\nu(T)$, which depends only on the frequency ν and temperature T and is given by:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp[h\nu/(kT) - 1]} \text{ J s}^{-1} \text{ m}^{-2} \text{ Hz}^{-1} \text{ steradian}^{-1} \tag{2.17}$$

where the Planck constant $h \approx 6.626 \times 10^{-34}$ J s, the Boltzmann constant $k \approx 1.381 \times 10^{-23}$ J K^{-1}, and a steradian is the dimensionless SI unit of solid angle, which is related to the area an angle cuts out. The solid angle of a full sphere is 4π and that of a hemisphere is 2π, where $\pi = 3.13159$.

The Planck distribution can also be written per unit wavelength, $B_\lambda(T)$, where the wavelength $\lambda = c/\nu$ and:

$$B_\lambda(T) = B_\nu \left|\frac{d\nu}{d\lambda}\right| = B_\nu(T)\frac{c}{\lambda^2} = \frac{2hc^2}{\lambda^5}\frac{1}{\exp\left[\left(\frac{hc}{\lambda kT}\right) - 1\right]} \text{ J s}^{-1}\text{ m}^{-2}\text{ m}^{-1}\text{ steradian}^{-1}. \tag{2.18}$$

The blackbody spectrum is markedly asymmetric. It falls off very rapidly with decreasing wavelength on the short wavelength side of the maximum and decreases gradually with increasing wavelength at long wavelengths. At short wavelengths, or high frequencies, we have the so-called Wien tail of the distribution, derived by the German physicist Wilhelm Wien (1864–1928) near the end of the 19th century. It is given by (Wien 1893):

$$B_\nu(T) = \frac{2h\nu^3}{c^2}\exp\left(\frac{-h\nu}{kT}\right) \text{ for } h\nu \gg kT. \tag{2.19}$$

Two English physicists, Lord Rayleigh (John Strutt, 1842–1919) and James Jeans (1877–1946) derived an expression for the brightness of thermal radiation at long wavelengths, or low frequencies. For wavelength λ this *Rayleigh-Jeans law* is (Rayleigh 1900, 1905; Jeans 1905, 1909):

$$B_\lambda(T) = \frac{2ckT}{\lambda^4} \text{ for } hc \ll \lambda kT \tag{2.20}$$

or at frequency ν:

$$B_\nu(T) = \frac{2\nu^2 kT}{c^2} \text{ for } h\nu \ll kT, \tag{2.21}$$

These equations are applicable at radio wavelengths or frequencies for most temperatures.

The *Rayleigh–Jeans law* agrees with experimental results at large wavelengths, with $\lambda \gg hc/(kT)$, or, equivalently, at low frequencies $\nu \ll kT/h$, but strongly disagrees at the short ultraviolet wavelengths (or high frequencies). This inconsistency between observations and the predictions of classical physics is commonly known as the *ultraviolet catastrophe*; Planck (1901) explained the inconsistency when he introduced radiation quanta.

The blackbody, or thermal, spectrum has a maximum intensity at a wavelength, $\lambda_{\max}$, which can be found by taking the derivative of the Planck distribution and setting the equation to zero, or from $dB_\lambda(T)/d\lambda = 0$, giving:

$$\lambda_{\max} = \frac{b}{T} = \frac{0.00289777}{T} \text{ meters} \approx \frac{0.0029}{T} \text{ meters}, \tag{2.22}$$

where b is the Wien displacement constant and T is the temperature on the kelvin scale, abbreviated K. This expression is called the *Wien displacement law*, after the German physicist Wilhelm Wien (1864–1928) who formulated the relationship based on a thermodynamic argument (Wien 1893). It is known as a displacement law because the wavelength peak $\lambda_{\max}$ is displaced when the temperature, T, is

changed. The expression indicates that colder objects radiate most of their energy at longer wavelengths, and that hotter objects are most luminous at shorter wavelengths. In other words, as the temperature of a gas increases, most of its thermal radiation is emitted at shorter and shorter wavelengths.

Example: The most intense thermal radiation at different temperatures

The Sun radiates its most intense radiation in the visible colors. At an orange wavelength of $\lambda = 500$ nm, the effective temperature of the Sun's photosphere is $T = 0.0029/(5.00 \times 10^{-7}) \approx 5{,}800$ K. The average body temperature of a human is about $T = 310$ K. From the Wien displacement law, the wavelength of maximum thermal radiation at this temperature is $\lambda_{max} = 0.0029/310 \approx 9.35 \times 10^{-6}$ m, corresponding to infrared wavelengths. This heat radiation can be detected by rattlesnakes and by night-vision goggles. The primary mirror of the *Spitzer Space Telescope* has a diameter of $D_T = 0.85$ m, and its angular resolution θ at this infrared wavelength is $\theta = \lambda/D_T \approx 1.1 \times 10^{-5}$ rad $\approx 2.27''$, where 1 rad $= 2.06265 \times 10^5\,''$. Suppose this telescope was pointed down at the ground to act as a spy satellite from a geosynchronous orbit where the orbital period is equal to the Earth's rotation period of 24 h. The semi-major axis of such an orbit is equal to $a = 42{,}164$ km, so the altitude H above the ground in $H = a - R_E = 3.579 \times 10^7$ m, where the radius of the Earth is $R_E = 6{,}371$ km (see Sect. 4.1). The smallest feature this telescope could resolve on the ground would have a linear size of $L = H \times \theta \approx 394$ m, bigger than a human but comparable to a convoy of vehicles. An x-ray telescope operating at a wavelength of $\lambda = 1.24 \times 10^{-9}$ m would detect the thermal radiation of a gas at a temperature of $T = 0.0029/\lambda \approx 2.3 \times 10^6$ K, or about 2 million K. In contrast, the cosmic microwave background radiation has a temperature of $T = 2.725$ K, and the wavelength at which its emission is most intense is $\lambda_{max} = 0.0029/2.725 \approx 0.001$ m or 1 mm.

The Wien displacement law helps explain why stars have different colors. Since red wavelengths, at about 660 nm, are longer than blue wavelengths, at around 470 nm, you would expect that the visible disk of a red star would be cooler than the disk of a blue star. The Wien displacement law yields effective disk temperatures of about 4,400 K for the red star and roughly 6,200 K for the blue star.

However, the radiation from exceptionally hot stars, which peaks at short, invisible ultraviolet wavelengths, also enhances the radiation intensity at adjacent blue wavelengths. A star that is most intense at an unseen ultraviolet wavelength of 30 nm might have a disk temperature as great as 100,000 K, and such a star will also emit more radiation in blue visible light than a cooler star (Fig. 2.4). Careful spectral calibration of stellar colors indicates that blue stars, of spectral class O, can indeed have disk temperatures as high as 280,000 K, while the red stars of spectral class M can be about 100 time cooler, at 2,800 K (Sect. 10.10).

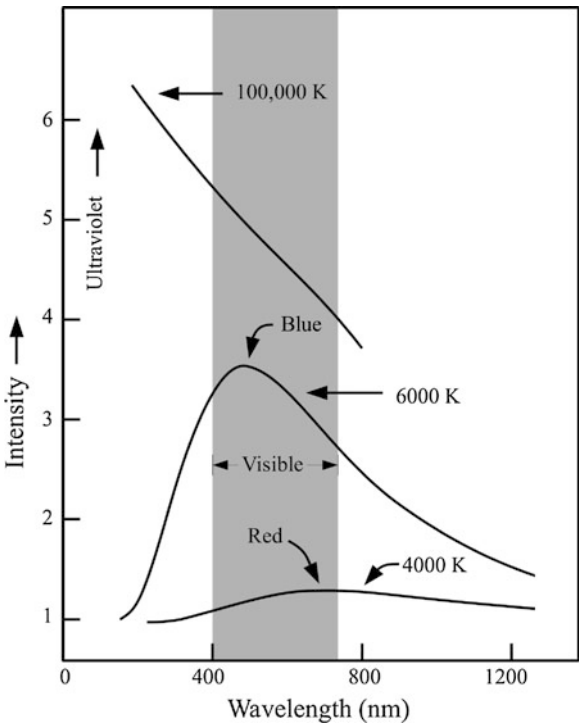


Fig. 2.4 Ultraviolet overflow The continuum spectrum of a star's thermal radiation changes with the effective temperature of the stellar disk, and this results in different star colors within the visible range of wavelengths, from 400 to 700 nm (*middle*). They range from blue stars, at relatively short visible wavelengths to red stars, at the longer wavelengths detected by our eyes. The thermal radiation of a star with an effective disk temperature of about 4,000 degrees on the kelvin scale, denoted K, peaks at the red wavelengths, and a hotter star with a temperature of about 6,000 K emits its most intense emission at blue wavelengths. A much hotter star at 100,000 K will be most intense at invisible ultraviolet wavelengths (*left*), but because the total energy emitted by a star increases dramatically with temperature, the very hot star will also appear bright at blue wavelengths

In terms of frequency, Wien's displacement law for the maximum frequency, $\nu_{\max}$, is determined from $dB_\nu(T)/d\nu = 0$ and is given by: $\nu_{\max} \approx 2.8\,kT/h \approx 5.8 \times 10^{10}\,T$ Hz.

Because the spectrum of blackbody radiation per unit frequency interval, $B_\nu(T)$, differs from the Planck distribution per unit wavelength, $B_\lambda(T)$, the $\nu_{\max}$ does not equal $c/\lambda_{\max}$.

The energy density, $u_\nu(T)$, of blackbody radiation, per unit frequency interval, is

$$u_\nu(T) = \frac{4\pi}{c} B_\nu(T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{\exp\left[\left(\frac{h\nu}{kT}\right) - 1\right]}. \tag{2.23}$$

Table 2.3 Radiation constants

a = Radiation density constant = $8\pi^5 k^4/(15c^3h^3) = 4\sigma/c = 7.5657 \times 10^{-16}$ J K^{-4} m^{-3}
σ = Stefan-Boltzmann constant = $2\pi^5 k^4/(15c^2h^3) = ac/4 = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}
c_1 = First radiation constant = $2\pi hc^2 = 3.741771 \times 10^{-16}$ J s^{-1} m^2
c_2 = Second radiation constant = $hc/k = 0.0143877$ m K
$b = \lambda_{max}T$ = Wien displacement law constant = 0.002897768 m K $\approx$ 0.002898 m K

The energy density has SI units of J m^{-3} Hz^{-1}. The radiation is isotropic, or the same in all directions, and the solid angle of a full sphere is 4π sr, where the constant $\pi \approx 3.14159$. When this expression is integrated over all frequencies, we obtain the total energy density, u, of a blackbody:

$$u = \int_0^\infty u_\nu(T)d\nu = aT^4, \tag{2.24}$$

where the radiation constant a is given by

$$a = \frac{8\pi^5 k^4}{15c^3h^3} \approx 7.57 \times 10^{-16}\,\mathrm{J\,K^{-4}\,m^{-3}} \tag{2.25}$$

The radiant flux, $f_\nu(T)$, of energy flowing out of the blackbody over π sr, or over the hemisphere facing an observer, is

$$f_\nu(T) = \pi B_\nu(T)\,\mathrm{J\,s^{-1}\,m^{-2}\,Hz^{-1}}, \tag{2.26}$$

in units of energy per unit time per unit area per unit frequency interval. The radiant flux is what is observed from astronomical objects. When integrating the flux over all frequencies one obtains the total radiant output per unit area, f, given by:

$$f = \int_0^\infty f_\nu(T)d\nu = \int_0^\infty \pi B_\nu(T)\,d\nu = \frac{ac}{4}T^4 = \sigma T^4, \tag{2.27}$$

where the Stefan-Boltzmann constant, σ, is given by: $\sigma = ac/4 = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}.

This and other radiation constants are given in Table 2.3.

We can add up, or integrate, the contributions to the blackbody spectrum at every wavelength to obtain the total luminosity of a thermal radiator. This results in the Stefan-Boltzmann law in which the luminosity increases with the square of the radius and the fourth power of the effective temperature. Luminosity is intrinsic to a star, establishing its power and energy output per unit time.

The Stefan-Boltzmann law states that the total power, or intrinsic luminosity L, at the visible disk of a star or other astronomical object with radius, R, and effective temperature, T_{eff}, is:

$$L = 4\pi\sigma R^2 T_{\mathrm{eff}}^4\,, \tag{2.28}$$

where $\pi = 3.1416$ and the Stefan-Boltzmann constant $\sigma = 2\pi^5 k^4/(15c^2h^3) = 5.6704 \times 10^{-8}\,\mathrm{J\,s^{-1}\,m^{-2}\,K^4}$. The effective temperature, T_{eff}, is the disk temperature that the object would have if it were a perfect blackbody radiating at luminosity L.

The unit of energy is the joule, and the unit of luminosity is joule per second, abbreviated $\mathrm{J\,s^{-1}}$. Power is often expressed in units of watts, where 1 watt = $1\ \mathrm{W} = 1\ \mathrm{J\,s^{-1}}$.

The Stefan-Boltzmann law indicates that at a given effective temperature, bigger stars have a greater luminosity than smaller stars, and at the same size, hotter stars are intrinsically more luminous than cooler stars. The Austrian physicist Joseph Stefan (1835–1893) obtained this law using experimental measurements made by the English physicist John Tyndall (1820–1893), and Stefan's student Ludwig Boltzmann (1844–1906) derived it from theoretical considerations, using thermodynamics (Stefan 1879; Boltzmann 1872).

The intensity of radiation striking a unit area decreases as the radiation spreads out into an increasing volume. The area of an imaginary sphere located at a distance, D, from the Sun or any other star is given by $4\pi D^2$, so the intensity per unit area, designated by l, is given by $l = L/(4\pi D^2)$, which falls off as the inverse square of the distance. You can notice this effect when watching the increased brightness of a car's headlight when the car approaches you and its distance decreases, or when watching the car's taillights dim as it moves away to greater distance.

The radiant flux, f, of a blackbody, thermal radiator of radius R and absolute luminosity, L, and temperature T, observed at a distance, D, is

$$f = \frac{L}{4\pi D^2} = \frac{\sigma R^2 T_{\mathrm{eff}}^4}{D^2}. \tag{2.29}$$

2.5 How Far Away is the Sun, and How Bright, Big and Hot is it?

2.5.1 Distance of the Sun

How far away is the Sun? The mean distance separating the Earth and the Sun is known as the astronomical unit, abbreviated AU, and it provides the crucial unit of planetary distance. Yet, for a very long time no one knew exactly how big it was. We now know that it is about 149.6 million km.

By the end of the 17th century, astronomers and other scientists had a good understanding of how the planets move around the Sun, but they could produce a scale model of the solar system that only provided relative distances of the planets from the Sun. The true distances and speeds of motion of the planets remained unknown.

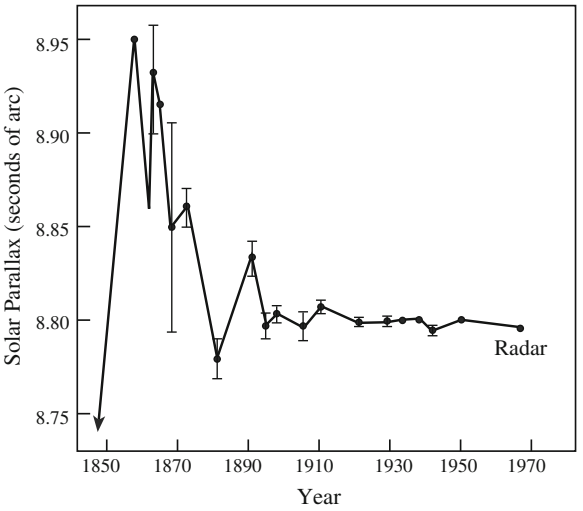


Fig. 2.5 Distance to the Sun Values of the solar parallax obtained from measurements of the parallaxes of Venus, Mars, and the asteroid Eros between 1850 and 1970. The solar parallax, designated by $\pi_{\odot}$, is half the angular displacement of the Sun viewed from opposite sides of the Earth. The error bars denote the probable errors in the determination; the points for 1941, 1950 and 1965 all have errors smaller than the plotted points. In the 1960s, the newly developed radar (i.e., radio detection and ranging) technology enabled determination of the Sun's distance with an accuracy of about 1,000 m. The radar value of the solar parallax is 8.79405 s of arc

It is no wonder then that obtaining a precise value for the Sun−Earth distance played an important role in the astronomy of the 18th and 19th century. The quest for accurately measuring that distance involved hundreds of trips to remote countries, tens of thousands of observations and photographs, and the lifetime work of several astronomers. They first determined the separations of the Earth and a nearby planet, such as Venus or Mars, and then used this planetary distance to infer the separation of the Earth and the Sun.

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

Since angular measurements were involved, the astronomical unit was naturally specified by an angle called the solar parallax, which is defined as half the angular separation of the Sun as viewed from opposite sides of the Earth. More than a century of estimates for the solar parallax are shown in Fig. 2.5 and discussed in Focus 2.4 – also see Hirshfeld (2001) and Van Helden (1985).

Focus 2.4 The solar parallax and the Sun's distance

The distance separating the Earth and the Sun, known as the astronomical unit or AU for short, is determined by first estimating the distance between the Earth and a nearby planet. This planetary distance then can be used to specify the AU. The distance of Venus from the Sun, for example, is equal to one half of the distance between the Earth and Venus when it is closest and farthest away, on the other side of the Sun. When the Venus−Sun distance is known, we can infer the distance of any other planet from the Sun using Kepler's third law (see Sects. 3.1, 3.2), which relates the orbital periods and orbital distances of the planets.

For more than a century, the distances of Venus and Mars were determined by triangulation from different points on the Earth. It involved measurements of the parallax, or angular difference in the apparent direction of the planet, as observed from widely separated locations. The solar parallax, designated by the symbol $\pi_\odot$ was then inferred. It is defined mathematically by $\sin \pi_\odot = R_E/AU$, where the equatorial radius of the Earth is $R_E = 6.378 \times 10^8$ m. The ratio of R_E and the AU provided an angle in radian units, and one radian is equivalent to 2.06265×10^5 $''$ where the symbol $''$ denotes a second of arc or an arc second.

In 1672, Giovanni Domenico Cassini (1625–1712), an Italian astronomer and the first director of the Paris Observatory, obtained an early triangulation of Mars, combining his observations from Paris with those taken by his colleague Jean Richer (1630–1696) from Cayenne, French Guiana. The planet was then in opposition, at its closest approach to the Earth. From the two sets of observations of Mars, made from opposite sides of the Earth and about 7,200 km apart, it was possible to estimate the distance to Mars and to infer an approximate value of $9.5''$ for the solar parallax (Van Helden 1985).

Astronomers in the 18th and 19th century 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 and 1769, with an estimate for the solar parallax of $8.57'' \pm 0.04''$ and in 1874 and 1882 with a wide range of results between $8.76''$ and $8.88''$ from world-wide observations. The method also involved comparison of observations from widely separated locations to determine the distance by triangulation.

In 1877, David Gill (1843–1914), an unemployed Scottish astronomer with no university degree, traveled to the small island of Ascension near the equator where he could use the Earth's rotation to view the near approach of Mars from different directions, obtaining a solar parallax of $8.78'' \pm 0.01''$.

Subsequent determinations of the distance to the nearby asteroid named 433 Eros, during its closest approaches to the Earth in 1900–1901 and 1930–1931, resulted in respective estimates for the solar parallax of $8.807'' \pm 0.0027''$ and $8.790'' \pm 0.001''$.

Significant improvements in the precision of planetary distances came in the late 1960s by bouncing pulsed radio waves off of Venus and timing the echo. The round-trip travel time – about 276 s when Venus is closest to the Earth – was measured using atomic clocks, and a precise distance to Venus then was obtained by multiplying half of the round-trip time by the speed of light.

The distance of Venus from the Sun is equal to one half of the difference between the Earth and Venus when it is closest and farthest away, on the other side of the Sun. The resulting radar value for the solar parallax was $8.79405''$. The corresponding value of the astronomical unit, inferred from the radar determination of the distance of Venus, is 149,597,870 km, with an accuracy of about 1 km, or for the accuracy required in most astronomical calculations 1 AU $= 1.496 \times 10^{11}$ m (Ash et al. 1967; Muhleman 1969).

Nowadays the accuracy of the mean Earth−Sun distance is fixed by the exact value for the speed of light. The Earth–Sun light travel time, τ_{AU} – or the time for light to travel across 1 AU – is given as a primary astronomical constant and has the value $\tau_{AU} = 499.0047863852$ s, with a derived value for the mean Earth-Sun distance of $1\text{AU} = c\tau_{AU} = 1.495978707 \times 10^{11}$ m, where the speed of light $c = 299792458$ m s^{-1}. The derived value of the solar parallax is $\pi_{\odot} = 8.7941433''$ where $''$ denotes second of arc.

The time for light to travel from the Sun to the Earth is used now as a primary astronomical constant. It is approximately 499 s, which corresponds to an AU of about 149.6 million km or 1.496×10^{11} m, and approximately 10,000 times the diameter of the Earth. Once scientist's determined the Sun's distance, they could determine the Earth's mean orbital velocity, by assuming – to a first approximation – a circular orbit and dividing the Earth's orbital circumference by its orbital period of $P_E = \text{one year} = 3.1557 \times 10^7$ s. The Earth's velocity is 2π AU$/P_E =$ 29,800 m s^{-1}, which is equivalent to approximately 107,000 km per hour, much faster than the fastest airplane or car.

By way of comparison, the light travel time from the next nearest star other than the Sun to the Earth is 4.24 years, or approximately 134 million s. This star is called Proxima Centauri and it is located at a distance of 4.01×10^{16} m and 268,000 AU.

Therefore, the Sun is about a quarter million times closer to the Earth than the next nearest star. Because of this closeness, the Sun is approximately 100 billion times brighter than any other star. This brilliance and proximity permit detailed investigations that are not possible for any other star. As a result, studies of the Sun provide the foundation and benchmark for an understanding of other stars.

Although our unaided eyes can see about six thousand stars in the night sky, and telescopes reveal hundreds of billions of them in the Milky Way, our own daytime

star, the Sun, is a special star. It is the source of all our power. Its radiation energizes our planet, warms the ground and sea, lights our days, strengthens our bodies, and sustains life on Earth.

The life-sustaining Sun also links us to the other stars, and to understand how the Sun or any other star operates, we must examine their radiation, which spreads out and carries energy in all directions.

2.5.2 How Big is the Sun?

Any incandescent body shines because it is hot. The wire filament in an incandescent light bulb is, for example, heated to a white-hot temperature of about 3,000 K to produce its luminous glow. As it turns out, the visible solar disk is just about twice that hot, and it owes its much greater luminosity to its vastly larger size.

The solar radius, denoted $R_\odot$, can be determined from observations of the Sun's angular size and distance, and these measurements indicate that the solar radius $R_\odot = 6.955 \times 10^8$ m, which is 109 times the radius of the Earth (see Sect. 1.4).

2.5.3 The Unit of Energy

The joule is the unit of energy in the International System of units, abbreviated by SI from the French Système International d'unitès. The SI unit of energy is named after the English physicist James Prescott Joule (1818–1889), who described the relationship of heat to mechanical work (Joule 1847), leading to the theory of conservation of energy. When an SI unit is spelled out in English, it begins with a lower case letter, like joule, but it is abbreviated with a capital version of the first letter, such as J.

A joule is the work required to produce one watt of power for one second, so a power of 1 J s^{-1} is equivalent to one watt.

One joule is twice the kinetic energy of a mass of one kilogram, abbreviated 1 kg, moving at a speed of one meter per second, or 1 m s^{-1}. This amount of energy is a very small number as far as the mass and speed of cosmic objects are concerned. The Sun, for example, has a mass of about 2,000 billion billion billion kg, or 2×10^{30} kg, and moves though space at a speed of about 220,000 m s^{-1}.

Even an ordinary table lamp with a 100 watt light bulb uses just 100 J s^{-1}, whereas the Sun liberates a lot more power, some 382.8 million billion billion J s^{-1} written 3.828×10^{26} J s^{-1}.

2.5.4 The Sun's Luminosity

The Sun emits radiation in all directions, and as the solar radiation spreads out into space, it is dispersed into an ever-increasing volume. The distant Earth therefore collects only a small fraction of the total energy radiated by the Sun. The solar constant specifies the amount of the Sun's radiation that arrives at our planet. It is denoted by the symbol $f_\odot$, and is precisely 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. (Any physical parameter of the Sun is denoted by a subscript $\odot$, a circle with a dot at the center.)

Artificial satellites have been used to accurately measure the Sun's total irradiance, or radiant flux, just outside the Earth's atmosphere, establishing the value of the solar constant (Kopp et al. 2005):

$$f_\odot = 1{,}361\,\mathrm{J\,s^{-1}\,m^{-2}}. \tag{2.30}$$

We can use the solar constant and Earth-Sun distance to determine the total amount of energy radiated by the Sun every second. At the Earth's mean distance of 1 AU from the Sun, the solar radiation per unit area is diminished by $4\pi\,(\mathrm{AU})^2$, the surface area of a sphere at this distance. We therefore infer the Sun's luminosity, denoted $L_\odot$, by multiplying the solar constant with this area to obtain:

$$L_\odot = 4\pi f_\odot (\mathrm{AU})^2 = 3.828 \times 10^{26}\,\mathrm{J\,s^{-1}}. \tag{2.31}$$

where $1\ \mathrm{AU} = 1.496 \times 10^{11}$ m.

2.5.5 Taking the Sun's Temperature

The Sun, like any incandescent body, shines because it is hot. How hot? Once we know the radius and luminosity of the Sun, we can determine the temperature of the Sun's visible disk. Using the Stefan-Boltzmann law, the effective temperature, $T_{\mathrm{eff}\odot}$, of the visible solar disk is given by:

$$T_{\mathrm{eff}\odot} = \left[\frac{L_\odot}{4\pi\sigma R_\odot^2}\right]^{1/4} \approx 5{,}780\,\mathrm{K}, \tag{2.32}$$

where the Stefan-Boltzmann constant $\sigma = 5.670 \times 10^{-8}$ J m $\mathrm{K^{-1}}$ $\mathrm{s^{-1}}$, and the Sun's radius is $R_\odot = 6.955 \times 10^8$ m. At this temperature, all elements in the Sun are present in gaseous form. The Sun is only about twice as hot as the wire filament in an incandescent light bulb, so its much greater luminosity is due to its vastly larger size.

Astronomers use the kelvin temperature scale that starts from absolute zero, the temperature at which atoms and molecules cease to move. The unit for this scale is written kelvin, without a capital K, or just denoted by a capital K. Water freezes at

273 K and boils at 373 K, and to convert to degrees Celsius, abbreviated by C, just subtract 273, or C = K − 273. The conversion to degrees Fahrenheit, denoted by F, is more complicated, with F = (9 K/5) − 459.4.

2.5.6 How Hot are the Planets?

Solar radiation warms a planet's surface, 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 a planet by assuming that the surface of a terrestrial planet or the cloud tops of a giant planet 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, $T_{\rm ep}$, from the relation $T_{\rm ep} = 279\ (\mathrm{AU}/D_p)^{1/2}$ K, where D_p is the planet's distance from the Sun and the mean distance between the Earth and the Sun is 1 AU = 1.496×10^{11} m.

To derive this expression, notice that the radiant energy per unit time, L_P that a planet of radius R_P receives from the Sun is:

$$L_P = \pi R_P^2 f = \frac{R_p^2 L_\odot}{4D_p^2} = \pi R_P^2 \frac{\sigma R_\odot^2 T_{e\odot}^4}{D_P^2}\ \mathrm{J\,s^{-1}}, \tag{2.33}$$

where R_P is the radius of the planet, f is the total amount of radiant solar energy perunit 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 $\sigma = 5.670 \times 10^{-8}$ J m^{-2} K^{-4} s^{-1}, the solar radius $R_\odot = 6.955 \times 10^8$ m, and the effective temperature of the visible solar disk is $T_{e\odot} = 5{,}780$ K.

The Stefan-Boltzmann law can also be applied to a planet, giving its radiant luminosity, L_P, or the amount of radiation energy lost per unit time, as a function of its radius and effective temperature, $T_{\rm ep}$.

$$L_P = 4\pi\sigma R_P^2 T_{\rm ep}^4\ \mathrm{J\,s^{-1}}. \tag{2.34}$$

Assuming thermal equilibrium between energy lost and received, so $L_P = \pi R_P^2 f$, and combining equations we obtain:

$$T_{\rm ep}^4 = \frac{L_p}{4\pi\sigma R_p^2} = \frac{L_\odot}{16\pi D_P^2 \sigma} = \frac{R_\odot^2 T_{e\odot}^4}{4\,D_P^2} \approx 1.35 \times 10^{32}\left(\frac{1}{D_P^2}\right) \approx 6.03 \times 10^9 \left(\frac{\mathrm{AU}}{D_P}\right)^2, \tag{2.35}$$

or

$$T_{\text{ep}} \approx 279\left(\frac{\text{AU}}{D_P}\right)^{1/2} \text{K}, \tag{2.36}$$

where 1 AU = 1.496×10^{11} m is the mean 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 star's disk temperature and the square root of the star's radius, as well as the planet's distance from the star, or on the star's absolute luminosity and the planet's distance from the star. This is of interest in determining the habitable zone, in which the planet surface temperature might permit liquid water, at temperatures between 273 and 373 K; it is located closer to a star that is less luminous.

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 0.0–1.0. Rocky bodies like the planet Mercury or the Earth's Moon 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 we detect with our eyes.

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

$$T_{\text{ep}} = 279(1 - A)^{1/4}\left(\frac{\text{AU}}{D_P}\right)^{1/2} \text{K}. \tag{2.37}$$

There are two kinds of albedo, the Bond albedo (Bond 1863), which measures the total proportion of electromagnetic energy reflected, and the visual geometric

Table 2.4 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} (K)	Mean Temperature[b] (K)
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	134
Uranus	19.19	0.51	53	76
Neptune	30.07	0.41	45	73

[a] Distance and mean temperature from the Jet Propulsion Laboratory. Effective temperatures are calculated from the visual geometric albedos, which are from http://ssd.jpl.nasa.gov

[b] The mean surface temperatures for the terrestrial planets and the mean cloud-top temperatures for the giant planets

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 our formula is applied to the Earth we obtain T_{ep} (Earth) $\approx$ 256 K using the Bond albedo and T_{ep} (Earth) $\approx$ 249 K 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 K.

The effective temperatures of the planets are compared to their mean observed surface or cloud-top temperatures in Table 2.4. 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 (Focus 2.5). The giant planets are also hotter, due to the heat left over from their formation or to helium raining down inside them.

Focus 2.5 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 our atmospheric blanket might warm the Earth was suggested by the French mathematician Jean-Baptiste Fourier (1768–1830) and developed by the Irish scientist John Tyndall (1820–1893). 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 (Fourier 1824, 1827).

Tyndall built an instrument to measure the heat-trapping properties of various gases, examining the transmission of infrared radiation through them. He found that the main constituents of our atmosphere – oxygen and nitrogen – were transparent to both visible and infrared radiation. Oxygen molecules, denoted O_2, account for 21 % of our atmosphere, while nitrogen molecules, designated N_2, accounts for 78 %. These diatomic, or two-atom, molecules are incapable of absorbing any noticeable amounts of infrared heat radiation.

Tyndall also found that water vapor, designated H_2O, and carbon dioxide, denoted CO_2, absorb significant heat even though they are minor ingredients of the Earth's atmosphere (Tyndall 1861, 1863). As Tyndall realized, these

gases are transparent to sunlight, which warms the ground, but partially opaque to the infrared rays, which are trapped near the surface and warm our globe. Water vapor and carbon dioxide 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.

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.

As Tyndall pointed out, our environment would be much colder at nighttime in the absence of the greenhouse effect, and the Earth might otherwise be covered with frost. The warming is crucial to life on Earth. If the Earth had no atmosphere, it would be directly heated by the Sun's light to temperatures below the freezing point of water. Fortunately, the extra heat from the greenhouse effect keeps the oceans, lakes and streams from turning into ice.

Nevertheless, humans have increased *global warming* by burning coal, oil, and gas and releasing carbon dioxide into the atmosphere. This extra warming has been rising ever since the industrial revolution. The effect was suggested by the Swedish scientist Svante Arrhenius (1859–1927), realized as an environmental threat by the American scientists Roger Revelle (1909–1991) and Hans E. Suess (1909–1920), and documented by Charles D. Keeling's (1928–2005) measurements of the atmospheric carbon dioxide (Arrhenius 1896; Revelle and Suess 1957; Keeling 1960, 1978, 1997). Weart (2008) describes the discovery of global warming.

The Nobel Peace Prize was awarded in 2007 jointly to the Intergovernmental Panel on Climate Change and to 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 film entitled *An Inconvenient Truth* (2006) documents Gore's campaign to make the issue of global warming, by human emissions of heat-trapping gases, a recognized problem.

2.6 The Energy of Light

When radiation moves in space from one place to another, it will behave like trains of waves. But when radiation is absorbed or emitted by atoms, it behaves not as a wave but as a package of energy, or like a particle, a *photon*. A photon is a discrete

quantity of energy associated with electromagnetic radiation. Thus, light has a wave-particle duality; it can act light a wave and a particle depending on the situation (De Broglie 1923).

Photons have no electric charge and travel at the speed of light. They are created whenever a material object emits electromagnetic radiation, and they are consumed when matter absorbs radiation. And each atom, ion, or molecule can only absorb and radiate at a very specific set of photon energies. (An ion is an atom that has lost one or more electrons.)

The ability of radiation to interact with matter is determined by the energy of its photons.

Photon energy depends on the wavelength or frequency of the radiation. Waves with shorter wavelengths, or higher frequencies, correspond to photons with higher energy. That is, the energy, E, transported by a particular photon is directly proportional to the radiation frequency, ν, and inversely proportional to the radiation wavelength, λ. The photon energy, E, is given by:

$$E = h\nu = \frac{hc}{\lambda}\,, \tag{2.38}$$

where h is the Planck constant with the value $h = 6.626 \times 10^{-34}$ J s, the frequency is given in Hz or s^{-1}, and the wavelength is in m.

The idea that light acts like a particle, the photon, when interacting with matter originated when Albert Einstein (1879–1955) explained the *photoelectric effect*, in which some metals release a current of electrons when light shines on them. Measurements of this effect indicated that the kinetic energy of the individual escaping electrons increases with the frequency of the incoming light wave. Einstein explained the observations by supposing that individual electrons are not hit by a continuous stream of light energy, but by an individual photon of light with an energy $h\nu$ (Einstein 1905a, b).

Einstein was awarded the 1921 Nobel Prize in Physics for his services to theoretical physics, and especially for his discovery of the law of the photoelectric effect. The American scientist Robert A. Millikan (1868–1953) subsequently endorsed the photon interpretation, despite his initial reservations, and used the effect to measure the value of Planck's constant h (Millikan 1916); he received the 1924 Nobel Prize in Physics for his work on the elementary charge of electrons and on the photoelectric effect.

The amount of energy transported by a single photon is quite small. For yellow light, the wavelength $\lambda = 580$ nm, so the frequency $\nu = 5.17 \times 10^{14}$ Hz, and the photon energy, E, is only 3.42×10^{-19} J. A hundred-watt light bulb radiates a power of 100 J s^{-1}, so it sends out an incredible 2.9 million million million, or 2.9×10^{18}, photons every second.

Radio waves have even smaller photon energy, when compared with the photons of visible light. The low energies of the radio photons cannot easily excite the atoms of our atmosphere, so radio photons easily pass through the air. Visible

radiation can also slip through the Earth's atmosphere with little trouble. Its photons are too energetic to resonate with molecular vibrations and they are too feeble to excite atoms.

Ultraviolet photons are sufficiently energetic to tear off electrons from atoms and many molecules in the Earth's atmosphere, particularly in the ozone layer. That's a good thing, since most of these ultraviolet photons cannot reach the ground. If they did they would cause lots of damage to our skin and eyes.

Astronomers often describe energetic, short-wavelength radiation, such as x-rays or gamma rays, in terms of their energy rather than their wavelength or frequency. At the atomic level, the natural unit of energy is the electron volt, or eV. One electron volt is the energy an electron gains when it passes across the terminals of a 1-volt battery. A photon of visible light has an energy of about two electron volts, or 2 eV. Much higher energies are associated with nuclear processes; they are often specified in units of millions of electron volts, denoted MeV. A somewhat lower unit of energy is 1,000 electron volts, called kilo-electron volts and abbreviated keV; it is often used to describe x-ray radiation. For conversion between energy units, $1\ \text{eV} = 1.602 \times 10^{-19}$ J and $1\ \text{keV} = 1.602 \times 10^{-16}$ J.

The x-ray region lies between 1 and 100 keV of energy. There are soft x-rays with relatively low energy and modest penetrating power, with energies of 1–10 keV. The hard x-rays have higher energy and greater penetrating power, at 10–100 keV. Gamma rays are even more energetic than x-rays, exceeding 100 keV in energy.

2.7 Radiation Scattering and Transfer

2.7.1 Why is the Sky Blue and the Sunsets Red?

Our atmosphere is a colorless gas, as you can see in looking at the air in your room, but the sky is usually blue and sunsets are red. The incident sunlight contains all colors, but molecules in our atmosphere scatter blue light from the Sun more than they scatter red sunlight. John Tyndall (1820–1893) discovered the effect when passing light through a clear fluid holding small particles in suspension (Tyndall 1861), and Lord Rayleigh (1842–1919) derived the relevant equations for atmosphere molecules a decade later. When the Sun is overhead, the light that reaches us is mostly scattered sunlight, and this causes the sky to appear blue. When the Sun sets, its rays pass through a maximum amount of atmosphere, and most of the blue light is scattered out before it reaches us. The setting Sun is therefore reddened, and atmospheric dust also contributes to its apparent red color.

2.7.2 Rayleigh Scattering

The scattering of radiation by a particle depends on the size, a, of the particle and the wavelength, λ, of the radiation. When the particle is much smaller in size than the wavelength, or $a \ll \lambda$, then the effect is known as *Rayleigh scattering*, named after Lord Rayleigh (1842–1919). It applies to gas molecules that scatter visible sunlight, explaining why the sky is blue and why a sunset red.

The intensity of Rayleigh scattered radiation, I, by a spherical particle of radius a at wavelength λ for an incident wave of intensity I_0 is (Rayleigh 1871, 1899):

$$I \approx \frac{8\pi^4 a^6}{D^2 \lambda^4} \left[\frac{n^2 - 1}{n^2 + 2}\right]^2 I_0 \left(1 + \cos^2 \theta\right), \tag{2.39}$$

where D is the distance from the sphere to the observation point, the scattering angle θ is the angle between the direction of propagation of the incident wave and the direction of observation, and n is the relative index of refraction $n = [\varepsilon_2 \mu_2 / (\varepsilon_1 \mu_1)]^{1/2}$ between the sphere, denoted by subscript 2, and the surrounding medium, labeled with subscript 1, the ε denotes the dielectric constant and μ is the magnetic permeability.

The amount of Rayleigh scattering from a single particle can also be expressed as a scattering cross section, σ_S given by (Rayleigh 1871):

$$\sigma_S = \frac{128\pi^5 a^6}{3\lambda^4} \left[\frac{n^2 - 1}{n^2 + 2}\right]^2. \tag{2.40}$$

The major molecular constituent in our atmosphere, nitrogen, has $\sigma_S = 5.1 \times 10^{-31}$ m^2 in green light at a wavelength of $\lambda = 530$ nm.

The strong wavelength dependence of the Rayleigh scattering, which varies as λ^{-4}, means that the shorter blue wavelengths are scattered much more than the longer, red wavelengths. Since the molecules in our atmosphere are much smaller than the wavelengths of colored light, the blue component of sunlight is more strongly scattered down to our eyes than the other colors, creating our bright blue sky. At sunset the Sun's rays pass through a maximum amount of atmosphere; most of the blue sunlight is then scattered out of our viewing direction, and the setting Sun is colored red. Dust in the air also helps redden the sunset.

Dust particles are larger than molecules, and comparable in size to the wavelength of visible light. The equations that describe the scattering are then more complicated; it is known as Mie scattering after the German physicist Gustav Mie (1869–1957) who first published its mathematical equations. They are used to describe the scattering of starlight by interstellar dust, which reddens the light of distant stars and has a relatively weak dependence on wavelength, varying as λ^{-1}.

2.7.3 Thomson and Compton Scattering

Electromagnetic radiation can be scattered by a free electron, which is unattached to an atom. The electric field of the incident wave accelerates the electron, which moves in the direction of the oscillating electric field and emits radiation at the same wavelength, or frequency, as the incident wave. The scattering is described by the *Thomson scattering* cross section, which is independent of the wavelength, or frequency, of the incident radiation.

Thomson scattering is very important deep within the Sun, where the temperatures are high enough to ionize the atoms, producing numerous free electrons that scatter radiation produced by nuclear fusion reactions in the solar core and determine how that radiation works it way out to the visible disk of the Sun (see Sect. 8.5). It also establishes the upper limit to the luminosity of a star, known as the Eddington luminosity, which is related to the largest mass a star may have (Sect. 10.1, Focus 10.2).

The English physicist Joseph John Thomson (1856–1940) first provided the expression for the total scattered power, P, or the energy scattered per unit time in all directions (Thomson 1903, 1906), which is given by:

$$P = \sigma_T c U, \tag{2.41}$$

where U is the energy density of the incident radiation, c is the speed of light, and the Thomson scattering cross section, σ_T, is given by:

$$\sigma_T = \frac{8\pi}{3} r_e^2 = \frac{8\pi}{3}\left(\frac{e^2}{4\pi\varepsilon_0 m_e c^2}\right)^2 = 6.6525 \times 10^{-29}\ \mathrm{m}^2, \tag{2.42}$$

where the classical electron radius $r_e = 2.818 \times 10^{-15}$ m, the electron charge $e = 1.602 \times 10^{-19}$ C, the electric constant in vacuum $\varepsilon_0 = 8.854 \times 10^{-12}$ F m^{-1}, the rest mass of the electron $m_e = 9.1094 \times 10^{-31}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. Notice that the free electron acts as if it had the classical electron radius when interacting with radiation.

The Thomson scattered radiation is polarized along the direction of the electron's motion, or along the direction of the oscillating electric field of the incident radiation. The power scattered per unit solid angle, $dP/d\Omega$, therefore depends on the angle θ between the direction of the electron's motion and the direction of the observer, or:

$$\frac{dP}{d\Omega} = \frac{3}{8\pi}\sigma_T U \sin^2\theta. \tag{2.43}$$

Shortly after the big bang origin of the expanding universe, it was so hot that the universe was completely opaque to electromagnetic radiation as the result of Thomson scattering. The cosmic microwave background radiation, dating back to shortly after the big bang, is thought to be linearly polarized as a result of Thomson scattering (see Sect. 15.2).

The Thomson scattering cross section is applicable whenever the incident photon energy is much less than the rest mass energy of the electron, for radiation frequency $v \ll m_e c^2/h \approx 10^{20}$ Hz. When the photon energy of the incident electromagnetic radiation is comparable to, or larger than, the rest mass energy of the free electron, or for frequencies $v \geq 10^{20}$ Hz, the incident radiation transfers energy to the electron, and the scattered photon has less energy, or a lower frequency and longer wavelength, than the incident one. The effect is named *Compton scattering* or the *Compton effect*, after the American physicist Arthur H. Compton (1892–1962) who first observed and explained it, receiving the 1927 Nobel Prize in Physics for his discovery. The change in wavelength, $\Delta\lambda$, caused by Compton scattering from an electron that is at rest, or not moving, is given by (Compton 1923a, b):

$$\Delta\lambda = \lambda_2 - \lambda_1 = \frac{h}{m_e c}(1 - \cos\theta) = \lambda_C(1 - \cos\theta), \tag{2.44}$$

where λ_1 is the wavelength of the incident radiation, λ_2 is the wavelength of the Compton scattered radiation, h is the Planck constant, m_e is the rest mass of the electron, c is the speed of light, and θ is the scattering angle, or the angle by which the incident radiation is deflected. The quantity $\lambda_C = h/(m_e c) = 2.426 \times 10^{-12}$ m is known as the *electron Compton wavelength*. The Compton wavelength for any other subatomic particle is given by the same expression with m_e replaced by the mass of the particle.

In the *inverse Compton effect*, the electrons are not at rest, and may be moving at high speeds. These high-energy electrons scatter low energy photons, and the photons now gain energy in the Compton interaction and the electrons lose energy. When the electron's speed is large, approaching that of light, the scattered frequency, v_2, for incident radiation of frequency v_1 is given by:

$$v_2 \approx \gamma^2 v_1 \text{ for } \gamma h v_1 \ll m_e c^2, \tag{2.45}$$

the scattering cross section is $\sigma_S = \gamma^2 \sigma_T$ for Thomson scattering cross section σ_T, and the total energy radiated per unit time, P, by an electron passing through radiation of energy density U is given by

$$P \approx \gamma^2 \sigma_T c U, \tag{2.46}$$

where the energy of the electron is $\gamma m_e c^2$ and γ is the Lorentz factor $\gamma = [1 - (V/c)^2]^{-1/2}$ for an electron moving at velocity V.

The scattered radiation from high-energy electrons with Lorentz factors $\gamma = 1{,}000$ has a frequency that is a million times that of the incident radiation. Thus radio radiation becomes ultraviolet radiation, infrared radiation becomes x-rays and optical radiation becomes gamma rays.

When the electron velocity is high and $\gamma h v_1 \gg m_e c^2$, all of the electron energy is transferred into the scattered radiation regardless of the incident photon frequency and $v_2 = \gamma m_e c^2/h$ for $\gamma h v_1 \gg m_e c^2$.

Inverse Compton scattering can quench the synchrotron radiation of cosmic radio sources, which is emitted by high-speed electrons, and the effect can be important in the x-ray radiation of relativistic electrons being accreted by a black hole. Longair (2011) has provided applications of scattering formulae to astronomical objects in high-energy situations.

2.7.4 Radiation Transfer

Once radiation is emitted from an astronomical object, it must pass through intervening space before it reaches the observer. The radiation can be absorbed when passing through a layer or cloud of matter, and the same material can also emit radiation. The material's effect on the radiation is therefore characterized by an absorption coefficient per unit length α_ν at frequency ν and emission coefficient ε_ν at frequency ν. For matter in thermodynamic equilibrium at temperature T,

$$\varepsilon_\nu = \alpha_\nu B_\nu(T), \tag{2.47}$$

a result derived by Gustav Kirchhoff (1824–1887) and hence known as *Kirchhoff's law* (Kirchhoff 1860). Here $B_\nu(T)$ is the Planck distribution for thermal, or blackbody, radiation discussed in the previous Sect. 2.4, and the emission coefficient ε_ν is the power per unit volume per unit frequency interval per unit solid angle.

The optical depth, denoted by τ_ν, characterizes the absorption capability of the intervening matter at radiation frequency ν. It is given by:

$$\tau_\nu = \int_0^L \alpha_\nu dx \tag{2.48}$$

for a cloud or layer of thickness L in the x direction and an absorption coefficient per unit length α_ν. For radiation of incident intensity $I_\nu(0)$, the intensity of radiation, $I_\nu(L)$, on leaving the cloud will be:

$$I_\nu(L) = I_\nu(0)\exp(-\tau_\nu) + \frac{\varepsilon_\nu}{\alpha_\nu}[1 - \exp(-\tau_\nu)], \tag{2.49}$$

and the intensity $I_\nu(\text{cloud})$ of the thermal radiation emitted by the cloud is given by:

$$I_\nu(\text{cloud}) = \int_0^L \varepsilon_\nu \exp(-\alpha_\nu x)dx \approx B_\nu(T)[1 - \exp(-\tau_\nu)], \tag{2.50}$$

with $B_\nu(T) = \varepsilon_\nu/\alpha_\nu$. If the cloud or layer has negligible radiation at frequency ν, its emission coefficient is effectively zero and:

$$I_\nu(L) = I_\nu(0)\exp(-\tau_\nu). \tag{2.51}$$

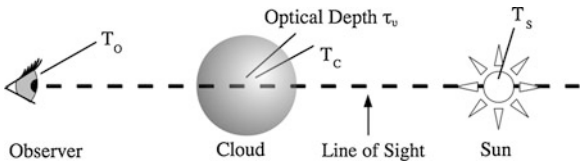


Fig. 2.6 Looking through a cloud When an interstellar cloud happens to lie along the line of sight to a star, the observed temperature, denoted T_O, can differ from the star's temperature, abbreviated T_S. That is because the cloud, with temperature T_C, will shine like any hot gas, emitting its own radiation, and the cloud can also absorb and scatter the star's radiation that is passing through it. A thick, dense cloud can absorb all the incident star's radiation, so you don't even see the star; just the cloud is detected. In contrast, a thin, rarefied cloud can be transparent; therefore, you look right through it, detecting the star as if the cloud wasn't even there. The cloud is characterized by its optical depth, denoted by the symbol τ_v, which depends on the substance in the cloud, the thickness of the cloud along the line of sight, and the observation frequency, designated by v. An optically thin cloud is transparent to the radiation at this frequency and the cloud optical depth $\tau_v \ll 1$. An optically thick cloud, with $\tau_v \gg 1$ is opaque, and at this frequency we cannot observe anything behind the cloud

The term optical depth implies that we are talking about radiation at the visible wavelengths we detect with our eyes. If the optical depth $\tau_v \gg 1$, along a ray path through a cloud or layer, then that cloud or layer is known as optically thick. On the other hand, a transparent cloud or layer is known as optically thin if $\tau_v \ll 1$. It follows that an optically thick object extinguishes the light of a source behind it, whereas an optically thin object absorbs negligible amounts of light passing through it. More generally, the terms optically thick and optically thin roughly mean opaque and transparent at the wavelength or frequency of electromagnetic radiation we are considering.

When the emission coefficient is not zero, the brightness, $B_{Cv}(T)$, of the thermal emission from the intervening cloud or layer is given by:

$$\begin{aligned} B_{Cv}(T) &= B_v(T)[1 - \exp(-\tau_v)] \\ &= B_v(T) \text{ if } \tau_v \gg 1 \text{ (optically thick)} \\ &= \tau_v B_v(T) \text{ if } \tau_v \ll 1 \text{ (optically thin)}, \end{aligned} \tag{2.52}$$

where $B_{Cv}(T)$ denotes the cloud brightness at frequency v and temperature T.

The observed brightness and temperature of a source that lies behind a cloud will depend on the temperature of the source, the temperature of the cloud, and the optical depth of the cloud (Fig. 2.6). When the cloud is transparent at the observation wavelength or frequency, which corresponds to the completely optically thin situation, the source temperature is observed. If the cloud is opaque, or optically thick, the cloud's temperature is observed.

If a source of brightness, $B_S(T)$, at frequency, v, and temperature, T_S, is irradiating a cloud or layer of temperature, T_C, the total observed brightness $B_{Ov}(T_{\mathrm{TOT}})$ is given by

$$B_{Ov}(T_{\mathrm{TOT}}) = B_{Sv}(T_S)\exp(-\tau_v) + B_{Cv}(T_C), \tag{2.53}$$

where

$$T_{\mathrm{TOT}} = T_S\exp(-\tau_v) + T_C[1 - \exp(-\tau_v)]. \tag{2.54}$$

For the optically thin case $\tau_v \ll 1$ the total observed temperature would be given by:

$$T_{\mathrm{TOT}} = T_S(1 - \tau_v) + \tau_v T_C. \tag{2.55}$$

Chapter 3
Gravity

3.1 Ceaseless, Repetitive Paths Across the Sky

Look up at the Sun as it glides across the bright blue sky, or watch the Moon's nightly voyage. On dark, moonless nights you also might notice a bright planet traveling against the stars.

Ancient astronomers thought that the Moon, Sun, and planets all moved in circles, forever wheeling around the central, unmoving Earth, and the Moon does indeed revolve about our planet. But the Earth and other planets revolve about the Sun, and the Sun does not revolve around the Earth.

So, motion is a matter of perspective. It is always relative, perceived only in relation to something else, by comparison with another object that is either at rest or moving in a different way.

The earliest Sun-centered theories of planetary motion had one fatal flaw; they also initially assumed that the planets move in circular orbits. This explanation couldn't be reconciled with careful observations of the changing positions of the planets in the sky that were meticulously carried out by the Danish astronomer, Tycho Brahe (1546–1601). Johannes Kepler (1571–1630), Brahe's assistant and eventual successor, found that the architecture of the solar system had to be described by noncircular shapes.

After 8 years of computations, Kepler found in 1609 that the observed planetary orbits could be described by ellipses with the Sun at one focus (Kepler 1609). This ultimately became known as *Kepler's first law* of planetary motion. Although the planetary orbits are nearly circular, they are slightly elliptical in shape.

At about the same time, Kepler described how a planet moves at different speeds as it travels along its elliptical orbit. He was able to state the relationship in a precise mathematical form now called *Kepler's second law*, which can be explained with the help of Fig. 3.1. 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 also is known as the *law of equal areas*. During the three equal time intervals shown in Fig. 3.1, the

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_3, © Springer-Verlag Berlin Heidelberg 2013

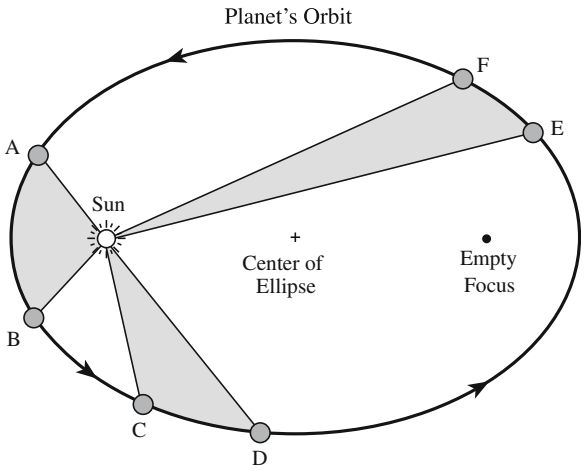


Fig. 3.1 Kepler's first and second laws The German astronomer Johannes Kepler (1571–1630) published his first two laws of planetary orbital motion in 1609. His first law states that the orbit of a planet about the Sun is an ellipse with the Sun at one focus. The other focus of the ellipse is empty. According 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, and 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

planet moves through different arcs because its orbital speed changes, but the areas swept out are identical.

So, a planet moves faster when it is closer to the Sun, and the modern explanation for this involves one of the fundamental concepts of physics, known as the *conservation of angular momentum* (Focus 3.1).

Focus 3.1 Moving along an elliptical trajectory

According to *Kepler's first law*, the planets move in elliptical orbits (Fig. 3.2). A 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 yearlong orbital period permits the determination of the Sun's mass and the Earth's orbital velocity, but only after astronomers had found out how large the AU is.

The distances to the solar focus and the shape of an ellipse are determined by its eccentricity, e. At perihelion the distance between the planet and the Sun is $[a\ (1 - e)]$ and at aphelion that distance is $[a\ (1 + e)]$. 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. 3.2, with an eccentricity of about $e = 0.5$.

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, at 0.467 AU.

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 at speed or velocity, V, and a distance, D,

$$\text{Angular momentum} = M \times V \times D. \tag{3.1}$$

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

The conservation law states that as long as no outside force is acting on a planet, its angular momentum cannot change. This means that a planet continues 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 from the Sun increases so the speed must decrease.

Kepler's third law took another 10 years of work to discover (Kepler 1619). In this musical pattern, each planet produces its own unique "note" as it moves around the Sun, with an orbital period that increases with a planet's distance from the Sun. Kepler's harmonic relationship states that the squares of the planetary periods are in proportion to the cubes of their average distances from the Sun. If P_P denotes the orbital period of a planet measured in years and a_P describes its semi-major axis measured in AU, then Kepler's third law states that $P_P^2 = a_P^3$. This expression is illustrated in Fig. 3.3, for the major planets and for the brighter moons of Jupiter. It also implies that a more distant planet moves with a slower speed. For a circular orbit, the planet's uniform velocity $V_P = 2\pi a_P / P_P = \text{constant} \times a_P^{-1/2}$, which falls off as the inverse square root of a_P.

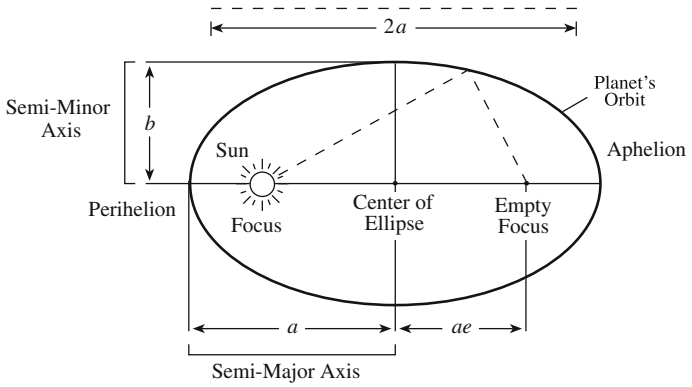


Fig. 3.2 Ellipse Each planet moves in an ellipse with the Sun at one focus. The length of a line drawn from the Sun, to a planet and then to the empty focus, denoted by the *dashed line*, is always $2a$, or twice the semi-major axis, a. The eccentricity, or elongation, of the planetary ellipse has been greatly overdone in this figure; planetary orbits look much more like a circle

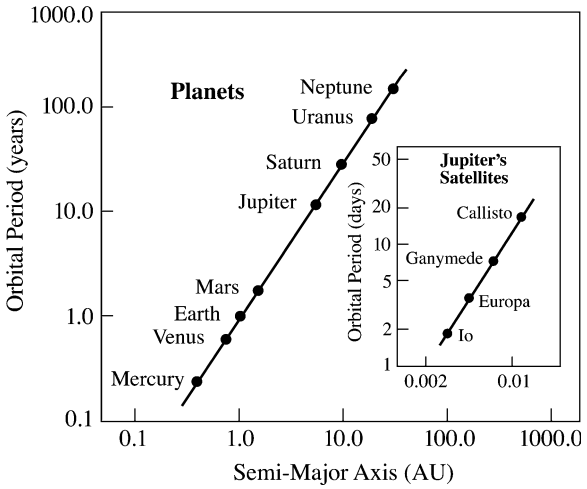


Fig. 3.3 Kepler's third law The orbital periods of the major planets in years are plotted against the semi-major axes of their elliptical orbits in astronomical units (*AU*), using a logarithmic scale. The straight line that connects the points has a slope of $3/2$, thereby verifying Kepler's third law that states that the square of the orbital periods increase with the cubes of the planetary distances. The German astronomer Johannes Kepler (1571–1630) published this third law in 1619. This type of relation applies to any set of bodies in elliptical orbits, including Jupiter's four largest satellites shown in the inset, with a vertical axis in units of days and a horizontal axis that gives the distance from Jupiter in AU units

These were amazing discoveries, but no one yet had explained what holds up the Moon and planets in their orbits. The explanation awaited the discovery of gravity, a principle that rules the universe.

3.2 Universal Gravitational Attraction

What moves the planets within their well-defined orbits? Kepler supposed that some invisible magnetic force emanated from the rotating Sun, and that this force pushed the planets through space. The farther the planet is from the Sun, the weaker the solar force and the slower a planet's motion – as described by Kepler's harmonic relationship.

Roughly half a century later, the great English scientist Isaac Newton (1643–1727) proposed another unseen agent, the invisible gravitational force of the Sun. Newton showed that the pull of gravity is universal, with an unlimited range and capacity to act on all matter, thereby holding the Moon, comets, and planets in their orbits.

Gravitation is the driving and organizing force of the universe (see Mac Dougal 2013); that is why it is known as *universal gravitation*. It binds stars and galaxies together and is responsible for their formation. The pull of gravity keeps our feet on the ground, so we rotate with the spinning Earth and stay on it. The atmosphere and oceans similarly are held close to the planet by its relentless gravitational pull.

It is gravity that explains why and how things fall. We might suppose, as Aristotle once did, that a heavy object will fall faster than a lighter one, in direct proportion to its weight, but that is not the case. Unless some outside force is involved, such as wind, all objects fall at the same rate, regardless of their weight. Galileo Galilei (1564–1642) stated the idea in his *Discoursi* (Galilei 1638), and apparently the same idea was stated 17 centuries before that by the Roman poet Lucretius (c. 99 BC–c. 55 BC) in *De rerum natura* (Lucretius 55 BC).

Galileo also proposed that any undisturbed body will fall with uniform acceleration, and he showed that the distance traveled by an object falling from rest is proportional to the square of the elapsed time. The distance, d, after time, t, is given by $d = gt^2/2$, where $g \approx 9.8\ \text{m s}^{-2}$ is the local acceleration of gravity on the Earth. In these ways Galileo provided a scientific foundation for Newton's subsequent theory of universal gravitation.

Newton realized 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 its orbit.

Newton showed that motions everywhere, whether in the celestial heavens on the ground, are described by the same concepts and that all material objects are subject to gravitation. Therefore, everything in the observable universe moves in predictable and verifiable ways. The basic ideas are that a moving body will continue to move in a straight line, unless acted on by an outside force, and that

every object attracts every other object as the result of universal gravitation. These insights resulted in Sir Isaac Newton becoming the first person in England to be knighted for his scientific work.

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* (Newton 1687). It was presented to the Royal Society of London in 1686, which withdrew from publishing it due to insufficient funds; Halley, a wealthy man, paid for the publication the following year.

The enormous reach of gravity can be traced to two causes. First, gravitational force decreases relatively slowly with distance, which gives gravity a much greater range than other natural forces, such as the strong force that holds the nucleus of an atom together. Second, gravitation has no positive and negative charge, like electricity, or opposite polarities like magnets. This means that there is no gravitational repulsion between masses. That is, the force of gravity acting between two objects always pulls them together and never pushes them apart. The attractive forces among unlike electrical charges in an atom cancel one another, 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, which possess inertia, the tendency to resist any change in motion. Mass is an intrinsic aspect of an object. It is different from weight, which decreases with distance from the main source of gravity. An astronaut, for example, weighs less after leaving the Earth, but his or her mass is just the same.

As expected, the strength of the gravitational force decreases with increasing distance, and Newton used Kepler's relationship between a planet's orbital period and distance to show that the force of gravity falls off as the inverse square of the distance from the center of the main source of gravity, the Sun.

Newton also demonstrated that the force of gravity at the Earth's surface is the same as the force, diminished by distance, which holds our Moon in place during its endless journey around the Earth. In effect our planet's gravity is forever pulling on the Moon, so it is perpetually falling toward the Earth while maintaining the same mean distance from it. Without the Earth's gravitational pull, the Moon would not orbit our planet but instead would travel out into space, never returning to Earth. The Sun's gravity similarly deflects the moving planets into their curved paths, so they forever revolve around the Sun (Newton 1687).

The gravitational power of an individual object depends on its mass and diminishes with distance from it. Expressed mathematically, any mass, M_1, produces a gravitational force, F_G, on another mass, M_2, given by the expression:

$$F_G = \frac{GM_1M_2}{D^2}, \tag{3.2}$$

where the universal gravitational constant denoted G, has the value $G \approx 6.674 \times 10^{-11}$ N m^2 kg^{-2}, and D is the distance between the centers of the two masses. This expression for the force is sometimes called an *inverse square law*, since the force of gravity is inversely proportional to the square of the distance or separation. The SI unit of force is appropriately called the newton, abbreviated N, and it is equal to the amount of net force required to accelerate a mass of 1 kg at a rate of 1 m every second squared, so 1 N = 1 kg m s^{-2}, and we can express the universal constant of gravitation with the units $G \approx 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

Any two masses attract one another with a gravitational force that varies in proportion to the product of the masses and the inverse square of the separation between their centers. The constant of proportionality – the *universal gravitational constant* G – was not measured until 71 years after Newton's death, and then indirectly by Henry Cavendish (1731–1810). Cavendish's aim was to determine the mass density of the Earth and because the planet's radius was known, he could effectively weigh the world. After nearly a year of meticulous observations, Cavendish (1798) announced that the Earth has a mass density of $\rho_E = 5{,}488 \pm 33$ kg m^{-3} (when corrected for a small arithmetical error in his paper). His result meant that the mass of the Earth is $M_E = 4\pi R_E^3 \rho_E / 3 \approx 6 \times 10^{24}$ kg, where the approximate radius of the Earth, $R_E \approx 6.4 \times 10^6$ m, was known at the time. (Distances part way around the surface of the Earth had been found by the surveying technique of triangulation, and combined to determine the Earth's circumference and a radius of about 6,400 km.) In Cavendish's time, mass and weight were assumed to be equal and, as he stated in his correspondence, he succeeded in weighing the world. It weighed in at a little more than 6 billion trillion metric tons. (A metric ton is 1,000 kg or 2,205 pounds).

Although he didn't specifically determine the gravitational constant, the value implied from Cavendish's work is $G = 6.754 \times 10^{-11}$ m^3 kg^{-1} s^{-2} (Brush and Holton 2001). A very precise value of G, accurate to more than the third decimal place, is still unknown, since gravity is a relatively weak force when compared to other forces that might act on the relevant experimental apparatus (Heyl 1930; Rose et al. 1969; Luther and Towler 1982; Gillies 1997; Fixler et al. 2007). The currently accepted value is:

$$G = 6.67428 \times 10^{-11}\,\mathrm{N\,m^2\,kg^{-2}} = 6.67428 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}, \tag{3.3}$$

with an uncertainty of 1 part in 10^4. For computations involving the orbits of either the natural or the artificial satellites around planets, and the trajectories of spacecraft visiting them, astronomers use the product of G and the planet's mass M_P, since GM_P is known more accurately than either term alone. The geocentric gravitational constant, GM_E, is, for example, a primary astronomical constant:

$$GM_E = 3.986004391 \times 10^{14}\,\mathrm{m^3\,s^{-2}}, \tag{3.4}$$

where $M_E = 5.9736 \times 10^{24}$ kg is the mass of the Earth, which is given together with other physical properties of the planet in Table 3.1.

Example: How fast are the Moon and planets moving, and how do we measure the mass of the planets?

Assuming a circular orbit at a distance D with period P, the average orbital speed will be $V = 2\pi D/P$, where $\pi = 3.14159$. The mean distance of our Moon from the Earth is $D = 384{,}400$ km $= 3.844 \times 10^8$ m and its orbital period around the Earth is $P = 27.3$ Earth days, where 1 day = 86,400 s, and its average orbital speed is about 1.02 km s^{-1}. For the Earth's orbit around the Sun, $D = 1$ AU $= 1.496 \times 10^{11}$ m and P = 1 yr $= 3.156 \times 10^7$ s, with an average orbital speed of $V = 2\pi D/P = 29.78$ km s^{-1}; since one hour 3,600 s, the Earth is moving at about 107,200 km/h, a lot faster than a vehicle on the highway. Jupiter is located at a distance of about 5.2 AU from the Sun, so from Kepler's third law, in which the square of the orbital periods scale as the cubes of the planetary distance, the orbital period P_J of Jupiter about the Sun will be $P_J = (5.2 \text{ AU}/1.0 \text{ AU})^{3/2} = 11.86$ years. Its average orbital speed is about 13 km s^{-1}, which is about three times slower than the Earth's orbital speed.

We can estimate the mass of a planet, M, from the motion of one of its moons, or natural satellites, using Kepler's third law, $M = 4\pi^2 D^3/(GP^2)$, where the Newtonian constant of gravitation $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. For our Moon, with the distance D and orbital period P given just above, we infer a mass of the Earth $M_E \approx 6.0 \times 10^{24}$ kg. The orbital parameters for Jupiter's natural satellite Io are $D = 421{,}700$ km $= 4.217 \times 10^8$ m and $P = 1.77$ Earth days $= 1.53 \times 10^5$ s, and with these parameters we obtain the mass of Jupiter $M_J \approx 1.9 \times 10^{27}$ kg $\approx 318\, M_E$.

Any object has a gravitational potential stored within it due to its efforts at overcoming relentless gravity. Two separated objects, for example, have worked against the gravitational attraction that pulls them together, achieving a reserve of energy and a potential for future action.

According to the conservation of energy, a fundamental law of physics, energy cannot be created or destroyed, just transformed. So the energy that went into overcoming the pull of gravity is stored in any object, and this stored potential energy can be converted into the kinetic energy of motion.

This *gravitational potential energy* is due to an object's position and is associated with the gravitational force. It depends on the height of the object, its mass, and the strength of the gravitational field it is in. For a very small mass m, as tiny as a point, the gravitational potential energy U when separated by a distance r in the gravitational field of another point mass M is:

$$U = -\frac{GMm}{r}. \tag{3.5}$$

Table 3.1 Earth's orbital and physical properties

Orbital characteristics

Po = orbital period of Earth about Sun = 365.25636 days = 1.000 sidereal year

Vo = average orbital speed of Earth about Sun = 2.9783×10^4 m s^{-1} = 107,200 km h^{-1}

a_E = AU = astronomical unit = mean Earth–Sun distance = $1.495\ 9787 \times 10^{11}$ m

$\pi_\odot$ = solar parallax = arcsin $(a_e/\text{AU}) \approx a_e/\text{AU} \approx 8.794143$ seconds of arc (for Earth's equatorial radius $a_e = 6.3781 \times 10^6$ m)

e = eccentricity = 0.01671

$a_E\,(1 + e)$ = aphelion = $1.52\ 098\ 232 \times 10^{11}$ m = 1.01671388 AU

$a_E\,(1 - e)$ = perihelion = $1.47\ 098\ 290 \times 10^{11}$ m = 0.98329134 AU

Physical characteristics

Age = 4.6×10^9 year

M_E = mass = 5.9736×10^{24} kg

$M_\odot/M_E$ = inverse mass = 332 946

a_e = equatorial radius = 6.3781×10^6 m

a_p = polar radius = 6.3568×10^6 m

$f = (a_e - a_p)/a_p$ = flattening = 0.0033528 = 1/298.25642

R_E = mean radius = $(a_e^2 a_p)^{1/3} \approx 6.371 \times 10^6$ m

ρ_E = mean mass density = $3M_E/\left(4\pi R_E^3\right) = 5515$ kg m^{-3}

GM_E = geocentric gravitational attraction = 3.986×10^{14} m^3 s^{-2}

g_E = equatorial gravitational acceleration = $GM_E/a_E^2 = 9.780\,\text{m}\,\text{s}^{-2}$

V_{escE} = surface escape velocity of Earth = $(2GM_E/R_E)^{1/2} \approx 1.1186 \times 10^4\,\text{m}\,\text{s}^{-1} \approx 11.2\,\text{km}\,\text{s}^{-1}$

B = magnetic field strength of Earth (equator to poles) = 0.3–0.6 G = (3–6) $\times 10^{-5}$ T (magnetic field poles reverse every 250,000 years)

P_r = rotation period = 24 h = 8.64×10^4 s

dP_r/dt = slow down of rotation = 0.002 s century^{-1}

ω = angular velocity of rotation = 7.292×10^{-5} radians s^{-1}

V_r = equatorial rotation velocity = 465.12 m s^{-1} = 1,674.4 km h^{-1}

A = albedo (Bond) = 0.306 or albedo (geometric) = 0.367

T = mean surface temperature = 287.2 K

Atmosphere

P = mean surface pressure at sea level = 1 bar = 1.01×10^5 Pa

N_2 = nitrogen molecule = 78.08 % by volume

O_2 = oxygen molecule = 20.95 % by volume

Ar = argon = 0.92 %

CO_2 = carbon dioxide = 0.038 %

H_2O = water vapor $\approx$ 1 % variable

The negative sign is a convention, not important for most physical purposes where differences in energy are used.

This expression can be used to determine the escape velocity from the gravity of an object of radius R; just equate the kinetic energy $mV^2/2$ to GMm/R to get the velocity V of escape, or V_{esc}, given by:

$$V_{esc} = \left(\frac{2GM}{R}\right)^{1/2}, \tag{3.6}$$

which is independent of the small mass m.

For a self-gravitating sphere of uniform mass density, rather than a point mass, the gravitational potential energy is given by integrating the potential energy over all parts of the sphere, resulting in:

$$U = \frac{-3GM^2}{5R}, \tag{3.7}$$

where R is the radius of the sphere and the mass, M is given by

$$M = \frac{4}{3}\pi R^3 \rho, \tag{3.8}$$

for a mass density ρ.

The *gravitational binding energy* of a sphere held together by its gravity is $3GM^2/(5R)$, without the minus sign; it is the amount of energy required to pull all of the material apart and the amount of energy released, mainly by heat, during its formation.

Because a precise value of the gravitational constant, G, is only known to three significant figures, the orbits of the planets are calculated using the *Gaussian constant of gravitation*, denoted by the symbol k, first proposed by the German mathematician Carl Friedrich Gauss (1777–1855). It is given by (Gauss 1809):

$$k^2 = \frac{4\pi^2 a_E^3}{P_E^2(M_E + M_\odot)}, \tag{3.9}$$

where a_E is the semi-major axis of the Earth's orbit about the Sun, the orbital period of the Earth, P_E, is one year, and M_E and $M_\odot$ respectively denote the mass of the Earth and the Sun. Here $a_E =$ AU is the astronomical unit, while the symbol a_e with a lowercase subscript e, is the equatorial radius of the Earth. The $a_E = 1.496 \times 10^{11}$ m and $M_\odot = 1.989 \times 10^{30}$ kg. By using this constant Gauss was able to simplify the calculation of planetary orbits; he had previously used it in his 1801 prediction of the orbit of the first asteroid, Ceres, which had been lost from view.

The Canadian-American astronomer Simon Newcomb (1835–1909) determined the value of k with such great precision (Newcomb 1895) that it is still used in computing the planetary ephemerides and is one of the defining astronomical constants. This value is:

$$k = 0.01720209895(\text{AU})^{3/2} M_\odot^{-1/2} (D)^{-1}, \tag{3.10}$$

where AU denotes the astronomical unit and the mean solar day $D = 86{,}400$ s. Thus, k^2 is the Newtonian constant of gravitation expressed in units of the astronomical unit, the solar mass, and the day. The derived, heliocentric, or Sun-centered, gravitational constant, $GM_\odot$, is given by $GM_\odot = (\text{AU})^3 k^2/D^2 = 1.327\,244 \times 10^{20}$ m^3 s^{-2}.

For objects near the Earth, the local acceleration of gravity g can be considered to be approximately constant and the expression for the gravitational potential energy relative to the Earth's surface becomes:

$$U = mgh, \tag{3.11}$$

where h is the height above the Earth's surface and g is the surface value of the acceleration of gravity, or $g = 9.780\ \text{m s}^{-2}$ at the Earth's equator. The local acceleration of gravity, g, determines how things fall. For an object of mass, M, and radius, R, we have:

$$g = \frac{GM}{R^2}, \tag{3.12}$$

where the universal constant of gravitation $\text{G} = 6.674 \times 10^{-11}\ \text{m}^3\ \text{kg}^{-1}\ \text{s}^{-2}$. For the Earth, the local acceleration of gravity at the equator is $g_E = 9.780\ \text{m s}^{-2}$, increasing to about $9.832\ \text{m s}^{-2}$ at the poles. The detailed mathematical expressions for the variation of g with altitude and latitude on the Earth are given in Focus 3.2.

Focus 3.2 The Earth's gravity

The gravitational acceleration of the Earth, g, depends on the distance from the planet's center. The value g_H at altitude H above the Earth's mean radius, R_E, is given by:

$$g_H = \frac{g_0 R_E^2}{(R_E + H)^2}, \tag{3.13}$$

where the standard gravity $g_0 = 9.8331\ \text{ms}^{-2}$ at $R_E = 6.371 \times 10^6$ m.

As we previously discussed in Chap. 1, Focus 1.2, the surface of the Earth is not perfectly round, being extended at the equator and squashed at the poles. The surface radius r at latitude ϕ of the Earth geoid is:

$$r = a_e\left(1 - f\sin^2\phi\right), \tag{3.14}$$

where the equatorial radius $a_e = 6.378\ 140 \times 10^8$ m, and the flattening factor f is given by

$$f = \frac{a_e - a_p}{a_p} = \frac{3}{2}J_2 + \frac{1}{2}m = 0.0033528 = 1/298.25642. \tag{3.15}$$

The polar radius $a_P = 6.356755 \times 10^6$ m; the dynamical form factor J_2 for the Earth is given by

$$J_2 = 0.0010826359. \tag{3.16}$$

The effective gravity of the Earth is reduced by its rotation, and this reduction is greatest at the equator. The ratio m of centrifugal acceleration at the equator to the gravitational acceleration at the equator is given by:

$$m = \frac{\omega^2 a_e^3}{GM_E} = 0.00346 \tag{3.17}$$

where the angular velocity of the Earth's rotation $\omega = 2\pi$ radians/ 86,400 s $= 7.292 \times 10^{-5}$ rad s^{-1}. Therefore, the effective gravity of the Earth is reduced by rotation, but at most by about 3 % and near the equator.

At sea level, then we can estimate the surface gravitational acceleration, g, at latitude, ϕ, from the formula derived by the French astronomer and mathematician Alexis Claude de Clairault (1713–1765). Known as *Clairault's theorem*, it is (Clairault 1743):

$$g = g_e\left[1 + \left(\frac{5m}{2} - f\right)\sin^2\phi\right] \tag{3.18}$$

where the surface equatorial acceleration of gravity is given by:

$$g_e = \frac{GM_E}{a_e^2}\left(1 + \frac{3}{2}J_2 - m\right) = 9.780\,327 \text{ m s}^{-2}. \tag{3.19}$$

Expressed numerically, Clairault's theorem becomes:

$$g = 9.780327\left(1 + 0.0053024\sin^2\phi - 0.0000058\sin^2 2\phi\right) \text{m s}^{-2}, \tag{3.20}$$

at latitude ϕ.

3.3 Mass of the Sun

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 (see Chap. 4):

$$P_P^2 = \frac{4\pi^2}{G(M_P + M_\odot)}a_P^3, \tag{3.21}$$

where the constant $\pi = 3.14159$, the universal gravitational constant $G \approx 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the 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 $(M_P + M_\odot)$ is therefore, to the first approximation, a constant equal to the Sun's mass, $M_\odot$, regardless of the planet under consideration, and Kepler's third law becomes:

$$P_P^2 = \frac{4\pi^2}{GM_\odot} a_P^3. \tag{3.22}$$

Since any planet mass, M_p, is much smaller than the Sun's mass, $M_\odot$, we have:

$$M_\odot = \frac{4\pi^2 a_P^3}{GP_P^2}, \tag{3.23}$$

where and $\pi = 3.141592$.

We can use the Earth's orbital motion at a mean distance of one astronomical unit, or $1\ \text{AU} = 1.496 \times 10^{11}$ m and its orbital period of one year, or $1\ \text{year} = 3.1556926 \times 10^7$ s to infer the mass of the Sun, $M_\odot$ from:

$$M_\odot = 4\pi^2(\text{AU})^3 \Big/ \left[G(1\text{yr})^2\right] = 1.989 \times 10^{30}\,\text{kg}. \tag{3.24}$$

It is the benchmark unit for specifying the mass of the stars and galaxies.

The ratio of the mass of the Earth, M_E, to the mass of the Sun, $M_\odot$, which is independent of G, is given by

$$\frac{M_E}{M_\odot} = \frac{1}{332{,}946}. \tag{3.25}$$

The Sun is about 333,000 times more massive than the Earth, and contains more than 99.9 % of the mass of the entire solar system, so our assumption that our planet's motion is controlled by the massive Sun is amply justified.

The first person to estimate the mass of the Sun was Newton, in the *Principia*, where he calculated that the ratio of the mass of the Earth to the mass of the Sun was 1/28,700. After an improved value for the distance to the Sun was available, he revised his result to obtain a ratio of 1/169,282 in the third edition of the *Principia*. The modern value is of 1/332,946 is a result of improved determinations of the AU.

3.4 Tidal Effects

3.4.1 The Ocean Tides

While walking along the beach we might notice that the waves are rising farther and farther up the shore, steadily advancing and enlarging the bounds of the sea. The tide is flooding the beach. But several hours later it retreats and goes down again. We say that the tide is rising and falling, while the sea runs in and out, and Newton showed that the Moon's attraction is the main cause of the *ocean tides*.

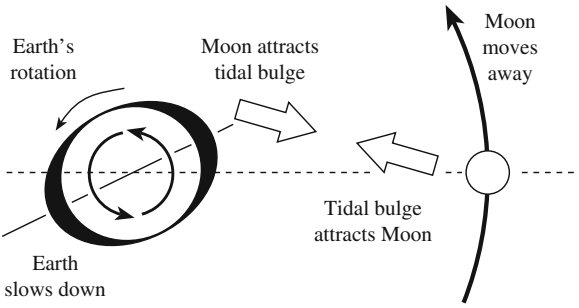


Fig. 3.4 Cause of the Earth's ocean tides The Moon's gravitational attraction causes two tidal bulges in the Earth's ocean water, one on the closest side to the Moon and one on the farthest side. The Earth's rotation twists the closest bulge ahead of the Earth−Moon line (*dashed line*), which produces a lag between the time the Moon is directly overhead and the time of highest tide. The Moon pulls on the nearest tidal bulge, slowing down the Earth's rotation. 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

Table 3.2 Orbital and physical properties of the Moon

Orbital characteristics

D_M = mean distance of Moon = 3.844×10^8 m = 384,400 km

V_M = mean orbital speed of Moon = 1.022×10^3 m s^{-1} = 1.022 km s^{-1}

V_{escE} = escape velocity of Earth at Moon's distance = $2GM_E/D_{moon})^{1/2} = 1.44 \times 10^3$ m s^{-1}

dD_M/dt = rate of increase of Moon's distance = 0.0382 ± 0.0007 m year^{-1}

P_M = orbital period of Moon = 27.3216 days = sidereal month = fixed star to fixed star (time from new moon to new moon is 29.530589 days = synodic month)

e = eccentricity = 0.0549

$D_M\,(1 + e)$ = apogee = 405,410 km

$D_M\,(1 - e)$ = perigee = 362,570 km

Physical characteristics

Age = 4.6×10^9 year

M_M = mass = 7.348×10^{22} kg

$M_\odot/M_M$ = inverse mass = 27,068,708.7

M_E/M_M = Earth-Moon mass ratio = 81.30056

$\mu = M_M/M_E$ = Moon-Earth mass ratio = 0.0123

R_M = mean radius = 1.737×10^6 m

ρ_M = mean mass density = $3M_M/\left(4\pi R_M^3\right) = 3346.4$ kg m^{-3}

g_M = equatorial surface gravity = 1.622 m s^{-2}

V_{escM} = escape velocity from surface = $(2GM_M/R_M)^{1/2} \approx 2.38 \times 10^3$ m s^{-1} = 2.38 km s^{-1}

P_r = sidereal rotation period = 27.3216 days (synchronous)

A = albedo = 0.136

T = mean equatorial surface temperature = 220 K (Temperature range at lunar equator 100–390 K)

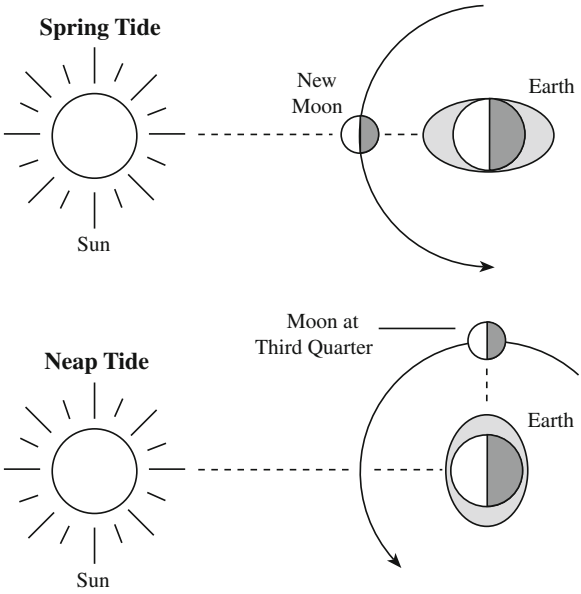


Fig. 3.5 Earth's spring and neap ocean tides The height of the tides and the phase of the Moon depend on the relative positions of the Earth, Moon, and Sun. When the tide-raising forces of the Sun and the Moon are in the same direction, they reinforce one another, making the highest high tides and the lowest low tides. These spring tides (*top*) occur at either 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 the Moon are at right angles to one another. The tidal forces are then in opposition, producing the lowest high tides and the highest low tides, or the neap tides (*bottom*). In this diagram, the height of the tides is greatly exaggerated in comparison to the size of the Earth

Because the Moon's gravitational force decreases with increasing distance, the Moon pulls hardest on the ocean facing it and least on the opposite ocean, whereas the Earth between is pulled with an intermediate force. In this way, the Moon's gravity draws out the ocean into the shape of an egg and creates two high tides. As the Earth's rotation carries the continents past the two tidal humps, we experience the rise and fall of water, the ebb and flow of the tides, twice every day (Fig. 3.4).

To understand the Moon's tidal producing force, we need to know its mass and distance, and they are given with other physical information about the Moon in Table 3.2.

The Moon creates most of the ocean waves, but the Sun also contributes to the size and rhythm of the waves. Although more massive than the Moon, our Sun also is much farther away; as a result the tide-producing force of the Moon is about

2.2 times that of the Sun. Near both full and new Moons, the tide-raising forces of the Sun and the Moon are in the same direction, producing the spring tide (Fig. 3.5). They reinforce one another's tides and produce high tides that can be a few times higher than normal. Two weeks later, the two tidal forces are in opposition and interfere with one another, and the range of these neap tides is then lower than any others.

Example: Moon tides and Sun tides

We can estimate the tide-producing capability of the Moon from the difference ΔF_M of its gravitational force, F_M, on the near and far sides of the Earth from:

$$\Delta F_M = \frac{GM_M M_E}{(D_M - R_E)^2} - \frac{GM_M M_E}{(D_M + R_E)^2} = \frac{4GM_M M_E R_E}{D_M^3}, \tag{3.26}$$

assuming $D_M \gg R_E$, where the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the mass of the Moon $M_M = 7.348 \times 10^{22}$ kg, the mass of the Earth $M_E = 5.974 \times 10^{24}$ kg, the mean radius of the Earth $R_E = 6.371 \times 10^6$ m, and the average distance of the Moon from the center of the Earth is $D_M = 3.844 \times 10^8$ m.

The ratio of the Moon's tide-producing force ΔF_M to the Sun's tide-producing force $\Delta F_\odot$ is:

$$\frac{\Delta F_M}{\Delta F_\odot} = \frac{M_M}{M_\odot}\left(\frac{D_\odot}{D_M}\right)^3 = 2.18, \tag{3.27}$$

where the mass of the moon is $M_M = 7.348 \times 10^{22}$ kg, the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, the mean distance of the moon is $D_M = 3.844 \times 10^8$ m, and the mean Sun-Earth distance is $D_\odot = 1$ AU $= 1.496 \times 10^{11}$ m. In other words, the Moon tides are about twice the Sun tides when they are in the same direction in the sky.

The Moon is now moving away from the Earth at a rate of 0.038 m year^{-1}, so its tide-producing force is gradually weakening, while the Sun remains at the same mean distance from the Earth with an unchanging tidal effect. The Moon tides will be equal to the Sun tides when the Moon has moved out to a distance of $D_M \approx 4.98 \times 10^8$ m, or an additional 1.14×10^8 m beyond its current distance. At the current rate of recession, that will happen about 3 billion years from now.

3.4.2 Tidal Locking into Synchronous Rotation

A planet's gravitational force pulls any natural satellite, or moon, into a slightly elongated shape along an axis pointing toward the planet. That is, a planet's gravitation produces two tidal bulges in the solid body of the satellite; one on the closest side to the planet and one on the satellite's farthest side. If the satellite's rotation twists the closest bulge ahead of the planet-satellite line, the planet pulls back on it. As a result, one hemisphere of the satellite always faces the planet, and the satellite takes as long to rotate as it does to orbit the planet. Then we say that the satellite has been tidally locked into synchronous rotation with the planet.

The Moon is in *synchronous rotation* about the Earth, so the Moon's rotation period is the same as the time it takes for the Moon to orbit the Earth, which is 27.32 Earth days. As a result, from Earth we always see the Moon's near side and never its far side. Only when a spacecraft passes beyond the Moon and looks back at it can we see the far side of the Moon. Most of the major moons, or large natural satellites, in the solar system have synchronous rotation with their planet.

If the mass of two orbiting bodies is comparable and their physical separation is relatively small, they both may be tidally locked to one another. This is the case for Pluto and its nearby large moon, Charon. Mutual tidal locking also occurs for close binary stars.

3.4.3 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, slows down the planet.

When the ocean tides flood and ebb, they create eddies in the water, producing friction on the ocean floor, which heats the water ever so slightly and dissipates energy at the expense of the Earth's rotation. The tides therefore act as brakes on the spinning Earth, slowing it by friction. As a result of this *tidal friction*, the rotation of the Earth is slowing down and the day is becoming longer at a rate of 2 ms, or 0.002 s, per century (Focus 3.3). In other words, the days are getting longer at the rate of 1 s every 50,000 years, and tomorrow will be 60 billionths of a second longer than today.

Focus 3.3 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 m 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, per unit area, ΔA, is

$$\frac{\Delta E}{\Delta t \Delta A} = \gamma \rho V^3 \approx 2\,\mathrm{J\,s^{-1}\,m^{-2}}, \tag{3.28}$$

where the density of sea water is $\rho \approx 1{,}000\,\mathrm{kg\,m^{-3}}$, a typical velocity $V \approx 1\,\mathrm{m\,s^{-1}}$, 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 $\Delta E/\Delta t \approx 10^{11}\,\mathrm{J\,s^{-1}}$ for the Irish Sea alone, and in the following year Sir Harold Jeffreys (1891–1989) estimated that the total rate of energy loss by tidal friction in the shallow seas surrounding Europe, Asia, and North and South America is $\Delta E/\Delta t \approx 10^{13}\,\mathrm{J\,s^{-1}}$ (Taylor 1919; Jeffreys 1920). This is comparable to the estimate obtained by considering the flux of energy convected into the shallow seas by tidal currents. Jeffreys (1975), Munk and MacDonald (1960, 1975), and Lambeck (1978, 1980) have discussed both the flux and bottom friction methods.

The lost energy comes from the Earth's rotational energy, which is equal to

$$\frac{1}{2} M_E V_{rot}^2 = \frac{2\pi^2 M_E R_E^2}{P_E^2} \tag{3.29}$$

where the mass of the Earth $M_E = 5.9736 \times 10^{24}$ kg, $V_{rot} = 2\pi R_E/P_E$, the constant $\pi \approx 3.14159$, the mean radius of the Earth $R_E = 6.371 \times 10^6$ m, and the rotation period of the Earth $P_E = 24\,\mathrm{h} = 8.640 \times 10^4$ s. For a period change ΔP in time interval Δt, the loss in rotational energy is:

$$2\pi^2 M_E R_E^2 \left[\frac{1}{P_E^2} - \frac{1}{(P_E + \Delta P_E)^2}\right] \approx \frac{4\pi^2 M_E R_E^2 \Delta P_E}{P_E^3}. \tag{3.30}$$

Setting this equal to the ΔE and collecting terms, we obtain:

$$\frac{\Delta P_E}{\Delta t} = \frac{P_E^3}{4\pi^2 M_E R_E^2} \frac{\Delta E}{\Delta t} \approx 6.7 \times 10^{-13}\,\mathrm{s\,s^{-1}} \approx 0.002\,\mathrm{s\ per\ century}, \tag{3.31}$$

for $\Delta E/\Delta t = 10^{13}\,\mathrm{J\,s^{-1}}$ and one century 100 years $= 3.156 \times 10^9$ s.

The long term increase in the length of the day of roughly 2 ms per century has been documented over the past 2,700 years from historical records of occultations of stars by the Moon and solar and lunar eclipses (Stephenson and Morrison 1984).

Paleontologists have made indirect historical measurements of the Earth's rotation through studies of fossil corals. The growth patterns of these corals consist of annual bands and fine daily ridges, produced by the effects 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 h 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 h long about 2 billion years ago (Wells 1963; Mazzullo 1971).

3.4.4 The Moon is Moving Away from the Earth

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 3.4). When we do the arithmetic, we find that the change of 0.002 s per century in the length of a day implies an outward motion of the Moon, amounting to about 0.04 m yr^{-1}. Small as it is, this value is just measurable with the laser light sent to corner reflectors, called corner cubes, placed on the Moon by the *Apollo* astronauts. Pulses of light are sent from the Earth to the tiny reflecting mirrors on the Moon, and the time for the light to travel to the Moon and return to Earth is measured. The distance to the Moon is the product of this round-trip light travel time and the speed of light. The lunar laser ranging data indicate that the Moon is moving away from the Earth at a rate of 0.0382 ± 0.0007 m $year^{-1}$ (Dickey et al. 1994).

Focus 3.4 Conservation of angular momentum in the Earth-Moon system

According to one of the unbreakable conservation laws, the angular momentum, or 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:

$$\text{Conservation of Angular Momentum} = M \times V \times R = \text{constant}.$$

This means that the angular momentum that the Earth loses in slowing down will be transferred to the Moon. For the Earth, the angular momentum is

rotational, with $V = 2\pi\ R_E/P_E$, where P_E is the Earth's rotation period of one day and the subscript E denotes the Earth. So, we have:

$$\text{Earth's Rotational Angular Momentum} = 2\pi M_E R_E^2/P_E.$$

Since the length of the Earth's day is increasing by the amount ΔP_E as time goes on, the Earth's rotational angular momentum is decreasing by the amount:

$$\begin{aligned}\text{Decrease in Rotational Angular Momentum} &= 2\pi M_E R_E^2 \left[\frac{1}{P_E} - \frac{1}{(P_E + \Delta P_E)}\right] \\ &= \frac{2\pi M_E R_E^2\, \Delta P_E}{P_E^2}. \qquad (3.32)\end{aligned}$$

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:

$$\text{Moon's Orbital Angular Momentum} = M_M \times V_M \times D_M, \qquad (3.33)$$

where 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 from Kepler's third law with the mass of the Earth and the Moon's distance from it, or by:

$$V_M = 2\pi D_M/P_M \approx (GM_E/D_M)^{1/2}, \qquad (3.34)$$

where $G = 6.674 \times 10^{-11}\ \text{m}^3\ \text{kg}^{-1}\ \text{s}^{-2}$ is the universal gravitational constant. Substituting this velocity expression into the angular momentum relation, we obtain:

$$\text{Moon's Orbital Angular Momentum} = M_M D_M (GM_E/D_M)^{1/2}. \qquad (3.35)$$

Since the mass of the Moon and the mass of the Earth do not change, the Moon's distance has to increase by an amount ΔD_M to provide an increase in the angular momentum.

$$\text{Increase in Orbital Angular Momentum} = M_M\, \Delta D_M \left(\frac{GM_E}{D_M}\right)^{1/2}. \qquad (3.36)$$

Setting the loss in rotational angular momentum equal to the gain in orbital angular momentum and collecting terms we obtain

$$\Delta D_M = \frac{2\pi M_E\, R_E^2\, \Delta P_E}{M_M\, P_E^2 \left(\frac{GM_E}{D_M}\right)^{1/2}} \approx 1.8 \times 10^{-9}\text{m s}^{-1} \approx 0.057\,\text{m yr}^{-1}, \qquad (3.37)$$

where $M_E = 5.974 \times 10^{24}$ kg, $R_E = 6.371 \times 10^6$ m, $\Delta P_E \approx 6.7 \times 10^{-13}$ s s^{-1}, $M_M = 7.348 \times 10^{22}$ kg, $P_E = 24$ h $= 8.640 \times 10^4$ s, $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, $D_M = 3.844 \times 10^8$ m and 1 yr $= 3.156 \times 10^7$ s. This is the approximate amount, of 0.04 m yr^{-1}, measured by sending laser pulses from the Earth to the corner reflectors left on the Moon by astronauts (Dickey et al. 1994).

Example: How close was the Moon to the Earth in their youth?

The mean distance of the Moon is now $D_{moon} = 3.844 \times 10^8$ m, and laser signals to the Moon's corner reflector mirrors indicate that the Moon is moving away from the Earth at the rate of $dD_{moon}/dt = 0.0382$ m yr^{-1}. The age of the Earth is 4.6 billion year. Assuming that the Moon was formed when the Earth was in its youth, and that the Moon has always been moving away from the Earth at the presently observed rate, then the Moon has moved over a distance of $\Delta D_{moon} = 0.0382 \times 4.6 \times 10^9 = 1.76 \times 10^8$ m, or about half its present distance. If the Moon moved away from the Earth at a faster rate when it was young, then it could have been formed by a collision with the newly formed Earth.

Will the Moon's outward motion carry it away from the Earth altogether? Only the intrusion of a massive third body could achieve that. 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.

3.4.5 A Planet's Differential Gravitational Attraction Accounts for Planetary Rings

One might expect the numerous particles of a planetary ring to have accumulated long ago into larger satellites. But the interesting feature of these rings – and a clue to their origin – is that they do not coexist with large moons. Planetary rings are also usually closer to the planets than their large satellites.

The rings normally are confined to an inner zone where the planet's tidal forces would stretch a large moon until it fractured and split apart, while also preventing small bodies from coalescing to form a larger moon (Fig. 3.6). The outer radius of this zone in which rings are found is called the *Roche limit* after the French mathematician Édouard A. Roche (1820–1883), who described it (Roche 1849, 1850, 1851).

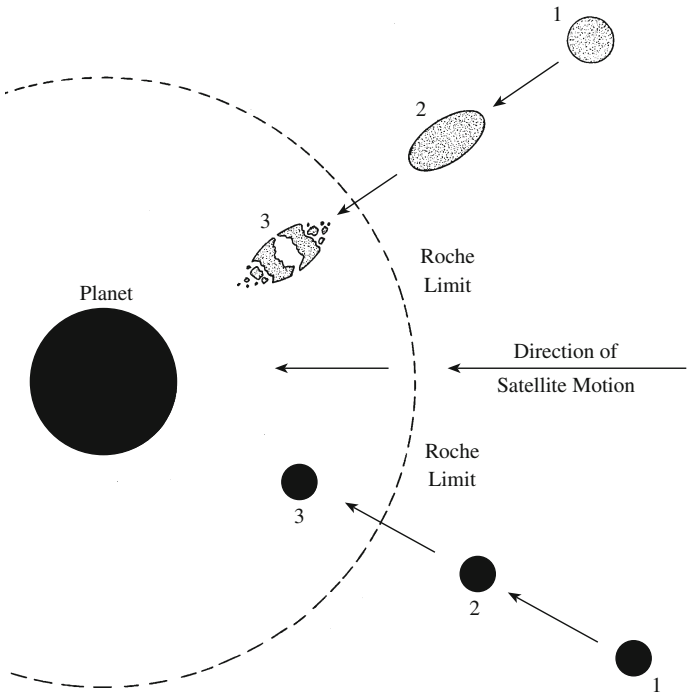


Fig. 3.6 Roche limit A large satellite (*top*) that moves well within a planet's Roche limit (*dashed curve*) will be torn apart by the tidal force of the planet's gravity. This was first investigated in 1849 by the French mathematician Édouard A. Roche (1820–1883). The side of the satellite closer to the planet feels a stronger gravitational pull than the side farther away, and this difference works against the self-gravitation that holds the body together. A small solid satellite (*bottom*) can resist tidal disruption because it has significant internal cohesion in addition to self-gravitation

For a rigid satellite with the same mass density as its planet, the Roche limiting distance from the center of the planet is 1.26 times the planet's radius (Focus 3.5). The Roche limit for a solid body is 1.38 times that radius (Aggarwal and Oberbeck 1974), and Roche's initial calculation, for fluid objects, was 2.446 times the planetary radius. Anywhere inside this distance, a large moon can no longer remain intact, but instead gets torn apart by planetary tides. Nevertheless, because of their material strength and cohesion, small moons less than 100 km across can exist inside the Roche limit without being tidally disrupted, just as the ring particles can.

The sharp edges of planetary rings can be formed by small satellites, which can pass within the Roche limit with enough internal cohesion to withstand the planet's differential gravitational forces. These smaller bodies can also confine ring particles within narrow boundaries. Goldreich and Tremaine (1982) have reviewed the dynamics of planetary rings.

Focus 3.5 The Roche limit

To visualize the significance of the Roche limit, consider two particles of equal mass, m, separated by a distance, R, and located at a distance, D, from a planet of mass, M_P. The gravitational pull of the planet on the particle closest to it will be greater than the pull on the more distant particle. If the difference in pull on the near and far particles, the tidal force, exceeds the mutual gravitational attraction between the two particles, they cannot remain close to each other and will disperse. The outcome of the tug-of-war between the tidal force and the mutual attraction is primarily decided by the particles' distance from the planet. At distances less than the Roche limit, D_{Roche}, particles are pulled apart, and this prevents the accumulation of larger moons. The tidal force will also tear apart any large moon-like object that ventures within the Roche limit.

The gravitational force, F_P, of a planet of mass, M_P, on a smaller mass, m, whose center is located at a distance, D, from the center of the planet is:

$$F_P = \frac{GM_P m}{D^2}, \tag{3.38}$$

where G is the universal constant of gravitation. The planet will pull harder on the side of the object that is closer to it and less hard on the side that is further away. The difference, ΔF, between the force felt by one side and the center of the mass, m, is

$$\Delta F = \frac{GM_P m}{2}\left[\frac{1}{(D-R_m)^2} - \frac{1}{(D+R_m)^2}\right] \approx \frac{2GM_P m}{D^3} R_m, \tag{3.39}$$

where the factor of ½ arises because the center of mass is located midway between the two forces, and R_m is the radius of the smaller object. If it approaches the planet, D becomes smaller and this tidal disruptive force will

increase, eventually pulling the object apart at a critical distance D_{Roche} from the center of the planet.

The gravitational binding force, F_B, which attracts the opposite sides of the object and holds it together, is Gm/R_m^2 per unit mass, or for the total mass

$$F_B = \frac{Gm^2}{R_m^2}. \tag{3.40}$$

The Roche limit is reached when the tidal disruptive force, ΔF, equals the binding force, F_B, and when we set these two expression equal and collect terms we obtain:

$$D_{Roche} = \left(\frac{2M_P}{m}\right)^{1/3} R_m. \tag{3.41}$$

This result is expressed in terms of R_m, the radius of the small object, but by using the mass densities ρ_P and ρ_{m} for the planet and small mass respectively, with $M_P = 4\pi\rho_P R_P^3/3$ and $m = 4\pi\rho_m R_m^3/3$, we obtain the Roche limit in terms of the planet radius, R_P:

$$D_{Roche} = \left(\frac{2\rho_P}{\rho_m}\right)^{1/3} R_P, \tag{3.42}$$

which for a planet and smaller object of the same mass density becomes $D_{Roche} = 1.26\ R_P$ (Jeans 1917).

The calculation by Roche used liquid objects whose shapes can distort continuously, and his result is (Roche 1849, 1850, 1851):

$$D_{Roche} = 2.446\left(\frac{\rho_P}{\rho_m}\right)^{1/3} R_P \approx 2.446\, R_P. \tag{3.43}$$

For a satellite with no internal strength and whose density is the same as the planet, the Roche limit is 2.446 times the planetary radius, or about 175,000 km for Jupiter, 147,000 km for Saturn, 62,000 km for Uranus, and 59,000 km for Neptune. Jupiter's insubstantial dusty ring, the magnificent ice particles of Saturn's rings, and the dark boulders in the narrow rings encircling Uranus and Neptune all lie within the Roche limit for the relevant planet. The Earth's Roche limit is 15,584 km, and if our Moon ever ventured within this distance from the Earth's center, it would be pulled apart by tidal forces and our planet would have rings. Nevertheless, the Moon is much farther away from the Earth, at a mean distance of approximately 384,400 km.

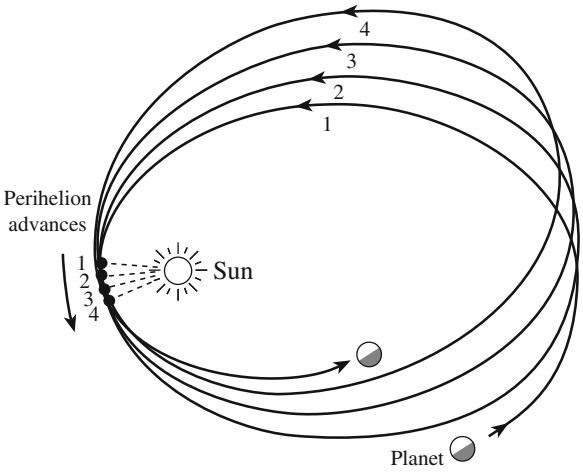


Fig. 3.7 Precession of Mercury's perihelion Instead of always tracing out the same ellipse, the orbit of Mercury pivots around the focus occupied by the Sun. The point of closest approach to the Sun, the perihelion, is slowly rotating ahead of the point predicted by Newton's theory of gravitation. This at first was explained by the gravitational tug of an unknown planet called Vulcan that was supposed to revolve about the Sun inside Mercury's orbit, but we now know that Vulcan does not exist. Albert Einstein (1879–1955) explained Mercury's anomalous motion in 1915 by inventing a new theory of gravity in which the Sun's curvature of nearby space makes the planet move in a slowly revolving ellipse

3.5 What Causes Gravity?

We cannot see the force of gravity and Newton did not know how it was exerted. Albert Einstein (1879–1955) subsequently explained it by supposing that a massive body like a star bends nearby space. This bending of space is the cause of the star's gravity. However, such effects are noticeable only in extreme conditions near a very massive cosmic object like a star, and the differences between Newton's and Einstein's theories of gravity are indistinguishable in everyday life.

One result of the Sun's curvature of nearby space is that planetary orbits are not exactly elliptical. This solved a perplexing problem with the motion of Mercury, the nearest planet to the Sun. 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 (Fig. 3.7). As a result, the point of Mercury's closest approach to the Sun, the perihelion, advances by a small amount – only 43 s of arc per century, beyond the location predicted using Newton's theory.

This anomalous twist in Mercury's motion was discovered in 1854 and recognized as an unexplained problem in 1859 by the French astronomer Urbain Jean Joseph LeVerrier (1811–1877). It was not explained for more than a half-century, when Einstein (1915) proposed that the planet is directed along a path in curved

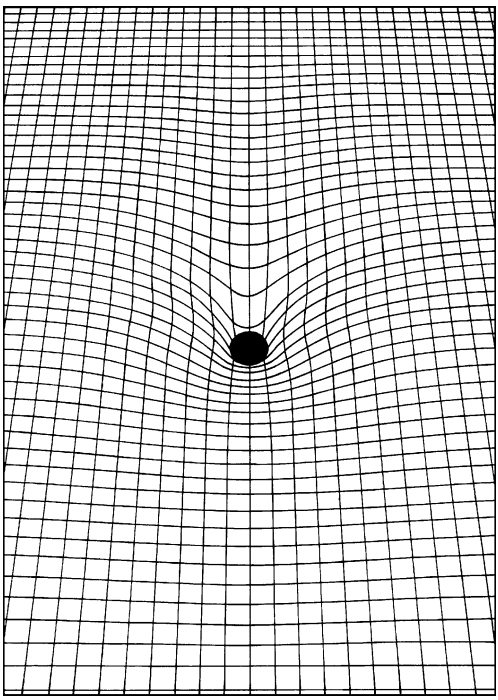


Fig. 3.8 Space curvature A massive object creates a curved indentation on the "flat" space described by Euclidean geometry, which applies in our everyday life on the Earth, where we do not directly encounter astronomical amounts of matter. Notice that the amount of space curvature is greatest in the regions near a cosmic object like a star, whereas farther away, the effect is lessened

space (Fig. 3.8), making the planet overshoot its expected location. Observations of Mercury trace out the invisible curvature. Roseveare (1982) has discussed Mercury's perihelion from Le Verrier to Einstein; also see Nobili and Will (1986).

In the very paper that explained Mercury's anomalous motion, Einstein showed that the curvature of space near the Sun also deflects the path of light from other stars (Fig. 3.9). The otherwise straight trajectory of starlight is bent by the Sun's gravity. The effect can be measured during a solar eclipse when stars pass behind the darkened Sun.

Newton previously had speculated that massive bodies might bend nearby light rays under the assumption that light has mass, and the German astronomer Johann George von Soldner (1726–1833) estimated the amount of light bending produced by the Sun using Newtonian gravity (Soldner 1801). In 1911, Einstein confirmed Soldner's result; however, when he took space curvature into account, the expected deflection was doubled (Einstein 1915).

The successful measurement of this deflection of starlight during the total solar eclipse on 29 May 1919 (Dyson et al. 1920), made Einstein famous, practically overnight. The initial measurements were not exact, and amounted to just a factor of two; nevertheless, the Sun's curvature of nearby space has now been measured with increasingly greater precision for nearly a century, confirming Einstein's prediction to two parts in a hundred thousand, or to the fifth decimal place. His *General Theory of Relativity* (Einstein 1916), which replaces the Sun's gravity

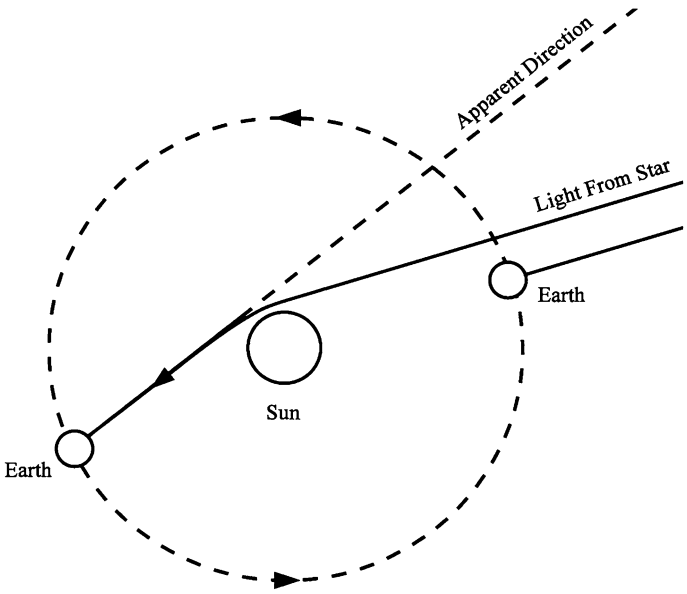


Fig. 3.9 Sun bending starlight As the Earth orbits the Sun, an observer's line of sight to a star or other cosmic object can pass near to the Sun or far from it. The massive Sun curves nearby space, bending the trajectory of starlight passing near to it, and this produces an apparent change in a star's position. The amount of bending and change in stellar position that were predicted by Einstein's *General Theory of Relativity* was first confirmed in 1929 during a total eclipse of the Sun

with geometry, has been verified by so many solar experiments that it now is widely accepted.

Natário (2011) provides a good description of the *General Theory of Relativity* without calculus, whereas Will (1993) has reviewed observational verification of the theory.

The solar curvature of space has been measured with increasingly greater precision for nearly a century, confirming Einstein's prediction. According to Einstein's theory, a light ray passing a minimum distance R_0 from the center of a star of mass M will be deflected by the angle

$$\theta = 2(1+\gamma)\frac{GM}{R_0 c^2} \text{ radians}, \tag{3.44}$$

where $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} is the universal gravitational constant, the speed of light is $c = 2.9979 \times 10^8$ m s^{-1}, and $\gamma = 1.000000000$, or exactly one, in Einstein's theory of gravitation. Newton's theory of gravitation implies $\gamma = 0$. For the Sun, with the mass $M_\odot = 1.989 \times 10^{30}$ kg, the *light bending* is:

$$\theta = 1.75\left(\frac{R_\odot}{R_0}\right)\left(\frac{1+\gamma}{2}\right) \text{ seconds of arc}, \tag{3.45}$$

where the Sun's radius is $R_\odot = 6.955 \times 10^8$ m and we have used 1 radian $= 2.06265 \times 10^5$ ″, and ″ denotes second of arc.

In one test, radio astronomers have combined observations of distant quasars from telescopes located on opposite sides of the world, turning the Earth into a gigantic interferometer that accurately measures the positions of radio sources as our line of sight to them nears the Sun (Fomalont and Sramek 1976). The change in position determines the amount of space curvature, as with the solar eclipse results for stars, but with much greater accuracy and without a total solar eclipse, since the Sun is a relatively weak interferometric radio source. Such Very Long Baseline Interferometry, abbreviated VLBI, has confirmed the predicted deflection of radio waves by the Sun to the 0.0003, or 0.03 %, level (Robertson et al. 1991; Shapiro et al. 2004; Fomalont et al. 2009).

Another test of Einstein's theory of gravitation measures the time required for a radar signal to make a round trip between the Earth and a planet (Shapiro et al. 1971), or the time for a radio signal to travel from a spacecraft home (Shapiro et al. 1971; Reasenberg et al. 1979). 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. Radio links with the *Cassini* spacecraft, for example, indicate that $\gamma = 1.00000 \pm 2.1 \times 10^{-5}$, or precisely unity with an accuracy of 2 parts in 100,000 (Bertotti et al. 2003).

A modern, extended test of Einstein's theory involves the measurement of the periastron advance and light bending of a binary pulsar that has much stronger gravitational fields than found in our solar system, with additional indication of gravitational radiation from observations of its orbital decay (Focus 3.6).

Focus 3.6 Testing relativity with the binary pulsar

The American radio astronomers Russell A. Hulse (1950–) and Joseph H. Taylor, Jr. (1941–) discovered the now famous *binary pulsar* PSR B1913 + 16 in 1974, during a deliberate search for new pulsars using the latest computer technology with the 305 m radio antenna at Arecibo, Puerto Rico (Hulse and Taylor 1975; Hulse 1994). Their mini-computer was programed to scan a range of possible pulsar periodicities, pulse durations, and frequency dispersions, registering a signal whenever a pulsar passed through the telescope beam. After 14 months at Arecibo, and the discovery of 40 pulsars, Hulse, a graduate student at the University of Massachusetts at Amherst, found the enigmatic PSR B1913 + 16, a pulsar in a binary system. PSR denotes pulsar, B designates a binary companion, and 1913 + 16 describes the position of the pulsar in the sky.

The pulsar rotates on its axis 17 times a second, so the pulse repetition period is about 0.059 s. Moreover, the period changes, by about 0.00003 s, and the period change is itself cyclical, increasing and decreasing, rising and

falling every 7.75 h. This meant that the pulsar was in orbital motion at this longer period, with the pulses being compressed together when the pulsar approached the Earth and pulled apart when moving away. A pulsar is also a neutron star, which has a mass of about the mass of the Sun collapsed to a radius of about 10 km, and as it turned out, the radio pulsar PSR B1913 + 16 was in rapid orbit with another neutron star that did not emit detectable radio pulses, perhaps because its radiation beam is not aimed at the Earth.

Hulse completed his degree, and left the field of radio astronomy just a few years later. So, precise timing of the radio pulses from PSR B1913 + 16 were continued by Hulse's advisor, Joe Taylor, and his other graduate students, permitting a determination of the orbital parameters of the system, as well as measurements of the mass of the pulsar and its silent companion. The minimum separation of the two neutron stars at periastron is about 1.1 $R_\odot$ and the maximum separation at apastron is 4.8 $R_\odot$ where the Sun's radius $R_\odot = 6.955 \times 10^8$ m. The periastron shift was enormous, at 4.2266 degrees per year, compared with Mercury's 43 s of arc per century, and this permitted the astronomers to infer a mass of $M = 2.828\ M_\odot$ solar masses for the binary system, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. The individual masses could be determined from another relativistic effect, and they weighed in at 1.441 $M_\odot$ and 1.387 $M_\odot$, as would be expected for two neutron stars (Taylor and Weisberg 1989).

More importantly, after four years of measurements and the analysis of about 5 million pulses, Taylor and his colleagues found that the orbital period was slowly becoming shorter, implying a slow shrinking of the average orbit size. The rate of decrease of the orbital period was 76.5 millionths of a second per year or 7.65×10^{-5} s year^{-1}, indicating that the two stars are drawing closer and closer to each other, approaching at about 2.5 m per year. This rate of orbital decay is just the change expected if their orbital energy is being radiated away in the form of gravitational waves, which had never been seen before (Taylor 1992, 1994; Taylor and Weisberg 1982; Weisberg and Taylor 2005).

Einstein (1916) predicted such *gravitational radiation*, showing that any accelerating mass would emit it – as ripples in the curvature of space-time. Gravity waves travel at the speed of light, as electromagnetic radiation does. But while electromagnetic waves move through space, gravity waves are an undulation of space itself.

Gravity waves are produced whenever a mass moves, but they are exceedingly faint when generated and become diluted as they propagate into the increasing volume of space. They are so weak, and their interaction with matter so feeble, that Einstein himself questioned whether they would ever be detected. The gravitational radiation loss of the orbital energy of PSR B1913 + 16 nevertheless exactly matches the amount predicted by Einstein's theory, providing clear and strong evidence for the existence of

gravitational radiation – for which Hulse and Taylor received the 1993 Nobel Prize in Physics.

Neither Einstein nor anyone else ever predicted that two neutron stars would be found that emit gravitational waves detected by timing the pulsar emission of one of them, and Taylor, Hulse, and their colleagues did not set out to find a binary neutron star, let alone detect gravitational waves. It was another one of those serendipitous discoveries that make astronomy so wonderfully unexpected and surprising.

A double radio pulsar has also been used to test special and general relativity. Designated PSR B0737–3039, the system also has strong gravitational fields, a rapid perihelion precession of 16.90 degrees per year, and an orbital decay attributed to gravitational radiation (Lyne 2004; Kramer et al. 2006; Breton et al. 2008). Unlike the binary pulsar PSR B1913 + 16, this new system contains two pulsars, attributed to two rotating neutron stars that emit radio pulses. They orbit each other at a speed of 300 km s^{-1} and complete one orbit every 2.4 h.

Joss and Rappaport (1984) have reviewed neutron stars in interacting binary systems. Backer and Hellings (1986) and Taylor (1994) have provided reviews of pulsar timing and general relativity. Kramer and Stairs (2008) have summarized knowledge of the double pulsar, and Hughes (2009) has discussed gravitational waves from merging compact binaries.

Chapter 4
Cosmic Motion

4.1 Motion Opposes Gravity

4.1.1 Everything Moves

All that exists, from atoms to planets and stars to galaxies, is always moving. This motion keeps cosmic objects suspended in space.

Galileo Galilei (1564–1642) imagined an ideal world in which there are no external forces acting on an object, and supposed that such an object will keep on moving at constant speed (Galilei 1632). Isaac Newton (1643–1727) extended the idea in his first law of motion, which states that every object continues in its state of rest, or of uniform velocity in a straight line, as long as no net force acts on it (Newton 1687). In other words, a moving object continues in motion with the same speed and in the same direction unless an external force is applied to it.

The most significant outside force in the universe is that of gravity, and it is motion that opposes gravitational attraction. Motion and gravity together shape the universe, giving it form and structure. So everything moves, and the way cosmic objects move is governed by the rules of motion and gravitation.

4.1.2 Escape Speed

The energy of motion is known as kinetic energy, and for a mass m moving at speed V, the kinetic energy is $mV^2/2$, so the faster something moves the more kinetic energy it has. If an object moves fast enough, and its kinetic energy becomes large enough, it can overcome the gravitational forces acting upon it and move out of their sphere of influence.

The minimum speed required to counteract and overcome the gravitational force on an object is known as the escape speed, since the object can then escape into surrounding space. The escape speed, denoted V_{esc}, needed for a small body of

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_4, © Springer-Verlag Berlin Heidelberg 2013

mass, m, to break away from the gravitational pull of a larger mass, M, is obtained by equating the kinetic energy of the small mass to the gravitational potential energy holding it in (Sect. 3.2). That is:

$$\text{Kinetic energy} = \frac{1}{2}mV_{esc}^2 = \frac{GMm}{D} = \text{Gravitational potential energy}, \tag{4.1}$$

or

$$V_{esc} = \left[\frac{2GM}{D}\right]^{\frac{1}{2}} = \sqrt{\frac{2GM}{D}}, \tag{4.2}$$

where D is the distance between the centers of the larger and smaller mass, and the Newtonian constant of gravitation $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

Although the escape speed is often called the escape velocity, the escape speed does not depend on the direction of motion, whereas strictly speaking a velocity includes both the speed and direction. No matter what the direction of travel is, an object moving at the escape speed can break away from another object's gravitational force, provided, of course, that it isn't directed into the surface of the larger mass.

The escape speed is independent of the small mass m, and it is dependent only on the distance D of the small mass and the value of the big mass M. At larger distances the escape speed becomes smaller because the strength of the gravitational force exerted by the big mass is less.

Any object, from an atom to a rocket, must move faster than the escape speed at a planet's surface if it is to move off into surrounding space. The reason why there is no hydrogen in the Earth's atmosphere, for example, is that at large altitudes, up in the ionosphere, the temperature is so high that the light-weight hydrogen atoms move at speeds greater than our planet's escape speed, and evaporate off into space. To obtain the surface escape speed of an object at its radius R, just let $D = R$ in the expression for escape speed.

Example: Escape speed of the Earth, Moon, and Sun

What is the minimum speed needed for a rocket to escape from the gravitational pull of the Earth, V_{escE}, and from the Moon, V_{escM}? We can use the expression $V_{esc} = (2GM/R)^{1/2}$, where the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the M is the mass of the object, R is its radius, and the ½ superscript denotes the square root. For the Earth, the mass $M_E = 5.9736 \times 10^{24}$ kg and the mean radius $R_E = 6.371 \times 10^6$ m, to give $V_{escE} = 1.12 \times 10^4$ m s^{-1}. For the Moon, we have a mass of $M_{moon} = 7.348 \times 10^{22}$ kg and a mean radius of $R_{moon} = 1.737 \times 10^6$ m, to give $V_{escM} = 2.38 \times 10^3$ m s^{-1}. That explains why a lunar lander requires much less rocket propulsion to leave the Moon to rejoin its orbiting command module than either spacecraft needs to leave the Earth. If the command module was orbiting the Moon in synchronous orbit, to remain always above

the same point on the Moon, it would have to have an orbital period equal to the Moon's rotation period of 27.3 Earth days $= 2.36 \times 10^6$ s, and the orbital distance, D, of the command module would be $D = [GM_{moon} P^2/(4\pi^2)]^{1/3} \approx 8.8 \times 10^7$ m, which is a substantial portion of the mean distance between the Earth and the Moon, of 3.844×10^8 m, resulting in poor visibility of the lunar surface from the command module. It has to move much faster around the Moon and closer to it. It is the Sun that dominates the mass of the solar system, with a mass of $M_\odot = 1.989 \times 10^{30}$ kg. The escape velocity at the visible disk of the Sun is $V_{esc\odot} = (2GM_\odot/R_\odot)^{1/2} \approx 6.18 \times 10^5$ m s^{-1}, where the Sun's radius $R_\odot = 6.955 \times 10^8$ m. The most distant comets reside in a remote reservoir, known as the Oort cloud, located at a distance of $D \approx$ 100,000 AU $= 10^5$ AU. The orbital period of such a comet will be $(10^5)^{3/2} \approx 32$ million years, and its orbital speed will be about $V = 2\pi D/P \approx 94$ m s^{-1}, a very slow and leisurely motion by cosmic standards; note 1 AU $= 1.496 \times 10^{11}$ m and 1 year $= 3.156 \times 10^7$ s. This makes sense, for the orbital speed will fall as the inverse square root of the distance. The Earth orbits the Sun at an average orbital speed of about 30 km s^{-1} = 30,000 m s^{-1}.

Using the mass and radius of the Earth, the Sun, and the Earth's Moon, we obtain respective escape speeds of 11.2, 618 and 2.38 km s^{-1}. If you want to send a rocket off into interplanetary space, it has to move faster than the escape speed at the Earth's surface, about 11.2 km s^{-1} $= 1.12 \times 10^4$ m s^{-1}. Owing to its larger mass, the escape velocity of the Sun is about 54 times larger than that of the Earth in spite of the Sun's larger radius. At the visible disk of the Sun, we have $V_{esc\odot} = 6.117 \times 10^5$ m s^{-1}. The escape speed from the surface of the Moon is just 2.38×10^3 m s^{-1}, which explains why the relatively small *Lunar Module* spacecraft could land on the Moon and blast off it with relatively low rocket propulsion, returning to its larger, mother spacecraft, the *Lunar Command Module*, that was orbiting the Moon and was launched from the Earth with considerably greater rocket thrust. The Moon's low escape speed also helps explain why it has no atmosphere to speak of.

The mass, radius, and escape speeds of representative planets and stars are given in Table 4.1.

4.2 Orbital Motion

A planet would continue going the way it started, moving along a straight line, if it were not for the Sun's gravitational force that deflects the planet into a curved solar orbit. Therefore, it is the Sun's gravitational attraction that keeps the planets forever moving along their orbital paths. But why doesn't the enormous solar

Table 4.1 Mass, radius and escape speed of some cosmic objects

Object	Mass (kg)	Radius (m)	Escape speed (km s^{-1})
Ceres, largest asteroid	1.17×10^{21}	3.8×10^{5}	0.64
Earth's moon	7.348×10^{22}	1.737×10^{6}	2.38
Earth	5.9736×10^{24}	6.371×10^{6}	11.2
Jupiter	1.90×10^{27}	7.15×10^{7}	59.5
Sun	1.989×10^{30}	6.955×10^{8}	618
Sirius B, white dwarf star	2×10^{30}	1×10^{7}	5,200
Neutron star	2×10^{30}	1×10^{4}	2×10^{5}

gravity pull all of the planets into the Sun? Motion holds the planets in their orbits, opposing the relentless pull of the Sun's gravity and keeping the planets from falling into the Sun.

Each planet is 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, maintaining an equilibrium between motion and gravitation that keeps the planets in perpetual motion.

For the planetary orbits, or any other orbit of small eccentricity, the length of the orbit is close to a circular one. The mean orbital speed, V_{OP}, of a planet in circular motion about the Sun at a distance D_P, is:

$$V_{OP} = \frac{2\pi D_P}{P_P}, \tag{4.3}$$

where $\pi = 3.14159$, the circumference of a circle with radius D_P is $2\pi D_P$ and P_P is the orbital period. The mean orbital velocity of the Earth around the Sun is, for example, is 29.8 km $s^{-1} = 2.98 \times 10^4$ m s^{-1}, where the mean Earth–Sun distance is 1 AU $= 1.496 \times 10^{11}$ m and the orbital period is 1 year $= 3.1557 \times 10^7$ s.

When there are two objects orbiting a common center of mass, and one of them has a very small mass when compared to the mass of the other one, as is the case of planets orbiting the massive Sun, the orbital speed depends only on the dominant mass and the distance of the orbiting object from it. For the planets, the orbital speed depends only on the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg and the distance of the planet, or:

$$V_{OP} = \left[\frac{GM_\odot}{D_P}\right]^{1/2} = \frac{V_{esc\odot}}{\sqrt{2}}. \tag{4.4}$$

which follows from Kepler's third law assuming a circular orbit or one of small eccentricity. It tells us that the more distant planets move at a slower speed. The orbital speed is independent of the planet's mass, which is why the planetary realm, known as the solar system, is dominated by the Sun.

This equation also indicates that the escape speed is $\sqrt{2}$ times larger than the orbital speed of a body. The $\sqrt{2}$ factor is a very small number, just 1.414, so the

orbital speed is very close to the escape speed at the relevant distance. Of course, a planet couldn't be moving just as fast as the escape speed, or any faster than that, for it would then escape from the solar system, moving off into interstellar space; and if the planet moved any slower than its orbital speed, it would be pulled into the Sun and consumed by it.

We can square both sides of the previous two equations and collect terms to obtain

$$P_P^2 = \frac{4\pi^2}{GM_\odot} D_P^3 \tag{4.5}$$

which is the Newtonian expression for Kepler's third law (Sect. 3.3).

Example: How fast does the Moon move around the Earth?

The Moon orbits our planet at a mean distance from the Earth of $D_M = 3.844 \times 10^8$ m with an orbital period P_M of 27.32 days, where 1 day = 86,400 s. This is the Moon's sidereal orbital period, from fixed star to fixed star. For a circular orbit, the Moon's mean orbital speed about the Earth would be $V_{OM} = 2\pi D_M / P_M = 1.02 \times 10^3$ m s^{-1}. We can compare this orbital speed to the escape speed, V_{escE}, from the Earth's gravity at the Moon's mean distance, $V_{escE} = (2GM_E/D_M)^{1/2} = 1.44 \times 10^3$ m s^{-1}, where the mass of the Earth is $M_E = 5.9736 \times 10^{24}$ kg and the Newtonian gravitational constant is $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. The mean orbital speed of the Moon V_{OM} is just equal to $V_{escE}/\sqrt{2}$ at the Moon's distance, which shows that the Moon is bound to the Earth by its gravitational pull, diminished by the distance to the Moon, and that the Moon is perpetually falling toward the Earth while moving around it.

Calculations of the speed of an orbiting object also apply to communications, military, and weather satellites, which might be launched into geosynchronous orbits with an orbital period equal to the Earth's rotation period.

Example: Geosynchronous orbits

In a geosynchronous orbit, a satellite's orbital period equals the Earth's rotation period, so the satellite stays in the same location above the planet's surface. The distance, D_{GS}, of this kind of satellite above the center of the Earth can be obtained from a rearrangement of Kepler's third law:

$$D_{GS} = \left[\frac{GM_E P_r^2}{4\pi^2}\right]^{1/3}, \tag{4.6}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the mass of the Earth $M_E = 5.9736 \times 10^{24}$ kg, the rotation period of the Earth is $P_r = 24$ h $= 86{,}400$ s, and the constant $\pi = 3.14159$. Substitution into this formula gives $D_{GS} \approx 4.22 \times 10^7$ m. The more exact value of $D_{GS} = 4.2164 \times 10^7$ m is obtained using the Earth's sidereal rotation period of $P_r = 86{,}164$ s. For a geosynchronous satellite orbiting the Earth's equator, the altitude H of the satellite above sea level will be $H = D_{GS} - a_e = 3.5786 \times 10^7$ m, where the equatorial radius of the Earth is $a_e = 6.3781 \times 10^6$ m.

When two orbiting objects have comparable mass, as is the case for some binary stars, then the mean orbital velocity, V_{O1} of an object of mass M_1 orbiting another mass M_2 at a distance a is given by:

$$V_{O1} = \left[\frac{GM_2^2}{(M_1 + M_2)a}\right]^{1/2}. \tag{4.7}$$

Here a is the separation of the two objects. If r_1 and r_2 denote their respective distances from a common center of mass, and we assume circular orbits, then $r_1M_1 = r_2M_2$ with $a = r_1 + r_2 = r_1\,(M_1 + M_2)/M_2$.

Astronomers record the spectral lines of a star, and look for periodic variations in the observed line-of-sight velocity, $V_{OBS1} = V_{O1} \sin i$, where i is the inclination angle between the perpendicular to the orbital plane and the line of sight. The observed period of variations in the detected radial velocity along the line of sight is the orbital period, P, given by Kepler's third law:

$$P^2 = \frac{4\pi^2 a^3}{G(M_1 + M_2)}. \tag{4.8}$$

We can use these equations to obtain an expression for the mass, M_2:

$$M_2^3 \sin^3 i = \frac{PV_{OBS1}^3}{2\pi G}(M_1 + M_2)^2. \tag{4.9}$$

When the mass of object 1 greatly exceed the mass of object 2, as is the case when looking for previously unseen exoplanets orbiting a nearby star, or when $M_1 \gg M_2$,

$$M_2 \sin\, i \approx \left(\frac{P}{2\pi G}\right)^{1/3} V_{OBS1}\, M_1^{2/3}, \tag{4.10}$$

and the mass of the star, M_1, can be inferred from other observations (Sect. 10.1).

4.3 The Moving Stars

4.3.1 Are the Stars Moving?

Each night the stars rise, move slowly across the dark sky, and then disappear from view; but this slow apparent movement of the stars is not due to the motions of the stars themselves. It is caused by the rotating Earth, which spins beneath the celestial sphere. Despite eons of stellar observations in antiquity, there was no evidence that any of these stars were moving.

Yet, if the stars were motionless, their mutual gravitation eventually would pull them together into a single mass. Without motion, there would be nothing to keep the stars apart, and they could not be suspended in space. So there is no star that is completely at rest, and the stars must be moving ever so slightly from their apparent places in the night sky.

Moreover, the speeds of the moving stars are not modest. Observations indicate that the stars are moving at speeds of about 10 km s^{-1} relative to their stellar neighbors. The Sun, for example, is currently traveling at a speed of about 20 km s^{-1}, or 20,000 m s^{-1}, relative to other nearby stars. This is about 1,000 times faster than a car moves on a highway.

Stars also move together at larger speeds in directed motions. Both the Sun and nearby stars, for example, are whirling about the remote center of the Milky Way at a speed of 220 km s^{-1}. If these stars traveled at faster speeds, they would move off into space, even out of the Milky Way; if they were moving at slower speeds, they would be pulled by gravitation into the center of the Milky Way. But because they are so far away, the stars seem to be moving slowly through space, only gradually changing their apparent separation and grouping.

4.3.2 Components of Stellar Velocity

Stars seem to be moving here, there and everywhere, so it is not easy to figure out where they are going. However, a star's motion manifests in two ways, depending on the method used to observe it, and these two components of velocity can be combined to give the direction of motion (Fig. 4.1). The "sideways" velocity component is directed perpendicular or transverse to the line of sight, with a speed designated by $V_\perp$, where the subscript $\perp$ denotes perpendicular. The other component, a radial velocity with a speed denoted by V_r, is the velocity moving toward or away from us along the line of sight to a star. When these two velocity components are known, we can determine the speed and direction of a star in three dimensions. The Pythagorean theorem gives the magnitude of the star's space velocity, its true speed in space, V_S, given by $V_S^2 = V_r^2 + V_\perp^2$.

A star's motion across the line of sight produces an angular change in position, called proper motion, which depends on both the star's distance and the

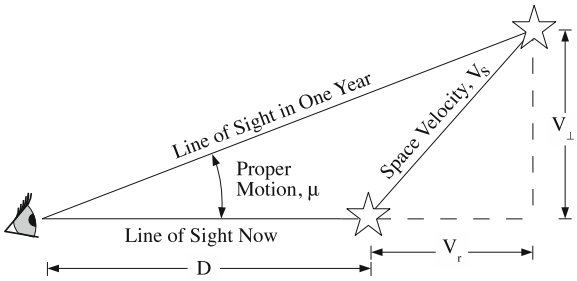


Fig. 4.1 A star moves The space velocity, V_S of a star relative to an observer can be resolved into two mutually perpendicular components: (1) the radial velocity, V_r, directed along the line of sight; and (2) the tangential velocity, $V_\perp$, which is perpendicular or transverse to the line of sight. From the Pythagorean theorem $V_S^2 = V_r^2 + V_\perp^2$. Over a given interval of time, shown here as one year, the star will move through a proper motion angle μ, which depends on $V_\perp$ and the star's distance, D, from the observer. In this figure, the proper motion $\mu = V_\perp/D$ is exaggerated greatly by more than 10,000 for even the closest star. At a distance of only 6 light-years, Barnard's star has the largest known proper motion of 10.3 s of arc per year

perpendicular speed $V_\perp$. The radial velocity is observed through the Doppler effect, which measures how the star's spectral lines appear to shorten or lengthen in wavelength depending on the relative velocity of the star and the observer, and whether the motion is toward or away from the observer. When a star is moving directly away, then there is no perpendicular component to its motion, and if the star is moving directly across the line of sight, then the radial component of the star's motion is reduced to zero.

It is difficult to judge a star's speed if it is headed straight toward or away from us, just as it is difficult to determine how fast a distant car is moving on a highway. However, if a star crosses at right angles to our line of sight, we could see a change in its position. To detect that change, astronomers needed to look at the nearest stars where the angular change in position is greatest.

Given enough time, the displacement of a nearby star's celestial position can be detected. The English astronomer Edmond Halley (1656–1742) first noticed the change when he compared the positions of extremely bright stars, such as Sirius and Arcturus, with those measured by the Greek astronomer Hipparchus around 150 BC and recorded in Ptolemy's reproduction of Hipparchus' catalogue. Halley's comparison indicated that at least three stars had changed position and moved (Halley 1717).

So it took more than 1,800 years before anyone noticed that a star could move. Nowadays, with vastly improved technology and observations from spacecraft, the motions of many tens of thousands of stars are known with great accuracy.

4.3.3 Proper Motion

The stellar motion that Halley detected is an angular change in a star's position over time, due to its velocity transverse or perpendicular to the line of sight. The angular rate of change is known as *proper motion*, which is intrinsic to the star and belongs to it, in contrast to any improper motion that might be caused by the Earth's movement in space.

Proper motion is not a velocity; it is the angular rate at which a star moves across the sky over years or centuries, and it does not by itself determine the speed of motion. To convert a star's proper motion into a velocity or speed, we must know the star's distance, and in Halley's time no one knew the distance of any star other than the Sun.

For a star located at distance D, the proper motion μ is:

$$\mu = V_{\perp}/D \,\text{rad s}^{-1}, \tag{4.11}$$

where 1 rad $= 2.06265 \times 10^5$ $''$ and $''$ denotes seconds of arc. Proper motion is designated by the Greek letter mu, or the symbol μ. The speed perpendicular to the line of sight, $V_{\perp}$, is known as the *transverse velocity*. If $V_{\perp}$ is given in units of km s^{-1} and D is in units of parsecs, we have:

$$\text{Annual Proper Motion} = \mu = 0.211 V_{\perp}/D \,'' \,\text{yr}^{-1}, \tag{4.12}$$

and

$$V_{\perp} = 4.74 \,\mu D \,\text{km s}^{-1}, \tag{4.13}$$

where 1 yr $= 3.156 \times 10^7$ s, and the μ in this case is called the annual proper motion. One parsec is abbreviated 1 pc, and 1 pc $= 3.08568 \times 10^{16}$ m is the typical separation between adjacent stars. The coefficient 0.211 comes from $2.06265 \times 3.156/3.08568$ in the various conversion factors, and $4.74 = 1.0/0.211$.

4.3.4 Radial Velocity

The other component of a star's velocity, the *radial velocity* directed along the line of sight, can be measured using the Doppler shift of a spectral feature in the star's radiation. Such a feature, called a spectral line, has a definite, well-known wavelength(Sect. 6.1).

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. Such a shift is named after the Austrian scientist, mathematician, and schoolteacher Christiaan Doppler (1803–1853) who discovered it more than one and a half centuries ago (Doppler 1842; Andrade 1959). If the motion is toward the observer, the shift is to shorter wavelengths; when the motion is

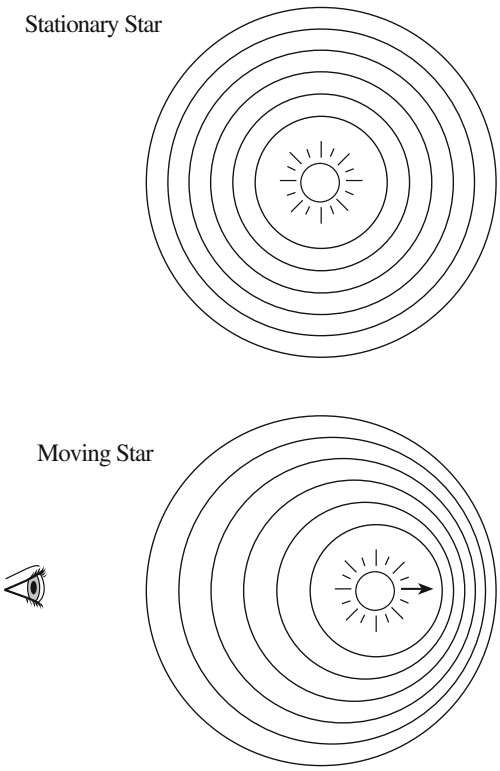


Fig. 4.2 Doppler effect A stationary source of radiation (*top*) emits regularly spaced light waves that get stretched out or scrunched up if the source moves (*bottom*). Here a star moving away (*bottom right*) from the observer (*bottom left*) is shown. The stretching of light waves that occurs when 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 speed of the component of the source's motion along the line of sight. The Doppler effect is named after the Austrian physicist Christiaan Doppler (1803–1853), who first considered it in 1842

away, the wavelength becomes longer (Fig. 4.2). We notice the effect for sound waves when listening to the changing pitch of a passing ambulance siren. The tone of the siren is higher as the ambulance approaches and lower when it moves away.

If the spectral line is emitted at a specific wavelength, $\lambda_{emitted}$, by a source at rest, the wavelength, $\lambda_{observed}$, observed from a moving source is given by the relation:

$$z = \frac{\lambda_{observed} - \lambda_{emitted}}{\lambda_{emitted}} = \frac{V_r}{c} \quad \text{for } V_r \ll c, \tag{4.14}$$

where the quantity z is known as the redshift, V_r is the speed of the source's radial motion along the line of sight away from the observer; since the speed and direction are known, V_r denotes the radial velocity. The speed of light $c = 2.9979 \times 10^8$ m s^{-1}. 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 wavelength. The greater the speed along the line of sight in either direction, the bigger the wavelength shifts.

The notation $V_r \ll c$ in our equation for the redshift, z, means that the formula applies for radial velocities, V_r, much less than the speed of light, c. This is the case for the motions of stars in our Milky Way. Nevertheless, remote collections of billions of stars, the galaxies, are moving at radial velocities that increase with their distance, and the exceptionally remote ones can have radial velocities that approach the speed of light, or $V_r \approx c$. In this case our equation has to be modified (see Sect. 14.3).

4.3.5 Observed Proper Motions of Stars

The star with the largest proper motion races across the sky at about 10.4 s of arc, denoted as $10.4''$, each year. This is *Barnard's star*, named after the American astronomer Edward E. Barnard (1857–1923) who discovered it (Barnard 1916). In our lifetime this star will move by roughly half the angular diameter of the Moon; however, because it is a dim, faint star a telescope is required to see it. Barnard's star is 1/27th of the brightness of the faintest star that can be seen with the unaided eye. It is a relatively nearby star, located at a distance of just 5.98 light-years, or 1.834 pc, and its large proper motion is attributed to both a high transverse speed and the closeness of the star.

Barnard's star moves across the line of sight at a speed of $V_\perp = 90.4$ km s^{-1}. When combined with its radial velocity of $V_r = -110.6$ km s^{-1}, with the negative sign indicating that the star is approaching us, a space velocity of $V_S = 142.7$ km s^{-1} relative to the Sun is obtained, from $V_S^2 = V_r^2 + V_\perp^2$. At its radial velocity, Barnard's star will move one light-year closer to us in about 2,100 years, using 1 light-year $= 9.461 \times 10^{15}$ m and 1 year $= 3.156 \times 10^7$ s to convert between units.

Example: How fast does Barnard's star move?

Barnard's star has an annual proper motion of $\mu = 10.4''$ year^{-1}. Its distance, D, inferred from its parallax (see Sect. 10.1) is $D = 5.98$ light-years $= 1.834$ pc, where 1 pc $= 3.26$ light-years. The star's transverse velocity, perpendicular to the line of sight, is $V_\perp = 4.74\ \mu D =$ 90.4 km s^{-1}. The star's redshift is $z = -3.689 \times 10^{-4} = -0.0003689$, so

its radial velocity is $V_r = z \times c = -110.6$ km s^{-1}, where the speed of light $c = 2.9979 \times 10^5$ km s^{-1}. The space velocity, V_S, of Barnard's star is inferred from $V_S^2 = V_\perp^2 + V_r^2$, or $V_S = 142.8$ km s^{-1}.

The closest star, Proxima Centauri, is just 4.24 light-years away, and at a distance of 5.98 light-years Barnard's star is nearly that close. However, Proxima Centauri is also moving closer, with a radial velocity of -21.7 km s^{-1}, so it will keep its status as the closest star for a very long time to come. The proper motion of Proxima Centauri is 3.85″ year^{-1}.

The star with the second largest proper motion, at 8.7″ yr^{-1}, is Kapteyn's star, named for the Dutch astronomer Jacobus C. Kapteyn (1851–1922), who first catalogued it (Kapteyn 1898). It has a distance of 12.8 light-years or 3.92 pc, so it is moving across the line of sight at a speed of $V_\perp = 162$ km s^{-1}. Kapteyn's star has a radial velocity of $V_r = 245.5$ km s^{-1}, giving it a true space velocity relative to the Sun of $V_S = 293.6$ km s^{-1}. This intriguing star moves around the center of the Milky Way in the opposite direction to the other nearby stars. It may have originated outside the Milky Way disk and is now hurtling through it.

Most proper motions are exceedingly small and usually measured in seconds of arc per century, or milliarcseconds per year, which means the same thing. Due to atmospheric blurring the angular resolution of the best telescope at the best location on the Earth is only about 0.2″, and we would have to wait more than 20 years to measure a proper motion of this size. However, the effect is cumulative; therefore successive generations of astronomers can measure proper motion. After 20 centuries, the proper motion of many stars might be 20″, which explains why Halley was able to detect the effect using ancient observations.

It is much easier to measure proper motion from space, outside the Earth's atmosphere. Instruments aboard the *HIPPARCOS* satellite have pinpointed the positions and established the proper motions of more than 100 thousand stars with an astonishing precision of 0.001″. The stellar distances are inferred from parallax measurements, and that explains the spacecraft's name, an acronym for *High Precision PARallax Collecting Satellite*. The perpendicular velocities can be determined from the proper motions and distances.

Astronomers specify the proper motion μ_α in right ascension α and the proper motion μ_δ in declination δ. The magnitude of the total proper motion, μ, is given by the vector addition of its components $\mu^2 = \mu_\delta^2 + \mu_\alpha^2 \cos^2 \delta$, where the *cos* δ factor accounts for the projection of μ_α on the celestial sphere. The components of proper motion and the radial velocities of stars with exceptionally high proper motion are listed in Table 4.2, where the proper motions are in units of milliarcseconds per year, or 10^{-3} ″ year^{-1}, and abbreviated mas year^{-1}, and the + or − sign of the radial velocity indicates motion away or toward the observer, respectively.

Table 4.2 Stars with the highest proper motion[a]

Star	$\mu_\alpha \cos\delta$ (mas year^{-1})	μ_δ (mas year^{-1})	Parallax (mas)	Radial velocity (km s^{-1})
Barnard's star	−798.71	1,0337.77	549.30	−110.6
Kapteyn's star	6,500.34	−5,723.17	255.12	+245.5
Groombridge 1830	4,003.69	5,814.64	109.22	−98.0
Lacaille 9352	6,766.3	1,327.99	303.89	+9.7
Gliese 1 (GJ 1)	5,633.95	−2,336.69	229.32	+23.6

[a] The designation mas is short for milliarcseconds or $0.001 = 10^{-3}$ s of arc

4.3.6 Motions in Star Clusters

Gravitation can constrain the paths of stars that are congregated within star clusters (Table 4.3). As many as 1 million stars, for example, are crowded together in a typical globular star cluster. The cluster is tightly bound by gravity, which gives it a distended spherical shape and relatively high stellar density toward the center (Figs. 4.3, 4.4). The name of this category of star cluster is derived from the Latin *globules*, for "a small sphere". Another type of stellar grouping, known as an open star cluster, includes up to a few thousand stars that were formed at the same time, but are only bound loosely to one another by mutual gravitational attractions (Fig. 4.5). Unlike globular star clusters, which can be held together by its stars' mutual gravitational pull for tens of billions of years, an open star cluster will disperse within a few million years.

The stars in a globular cluster are moving around like a swarm of bees, or like hot, subatomic particles inside a star. The stellar motions oppose the combined gravitational attraction of all of the stars, preventing them from gathering together and collapsing to the center of the star cluster.

In a short elegant discussion, the great British astronomer Arthur Stanley Eddington (1882–1944) demonstrated that the internal kinetic energy of a star cluster is half its gravitational potential energy (Eddington 1916). He also pointed out that this result could have been obtained at once from what is known as the virial theorem, a formula whose previous use had been almost entirely restricted to gases. In Eddington's application, stars replace the atoms and molecules of a gas.

The *virial theorem* describes the stability of a finite, self-gravitating collection of particles, either atoms or stars, which is bound by gravitational forces. It states that the total kinetic energy averaged over time is just equal to half the total

Table 4.3 Physical properties of star clusters

Open star cluster	Globular star cluster
N_S = total number of stars = 100 to 1,000	N_S = total number of stars = 10^4 to 10^6
R_C = radius = 1 to 10 pc $\approx$ (3 to 31) $\times$ 10^{16} m	R_C = radius = 10 to 100 pc $\approx$ (3 to 31) $\times$ 10^{17} m
Age = 10^7 to 10^9 year	Age = (10 to 14) $\times$ 10^9 year = 10 to 14 Gyear

Fig. 4.3 Globular star cluster NGC 6934 Several hundred thousand stars swarm around the center of the globular star cluster NGC 6934, which lies at a distance of about 50,000 light-years from the Earth. These ancient stars are estimated to be about 10 billion years old. This sharp image, obtained from the *Hubble Space Telescope*, is about 3.5 min of arc and 50 light-years across. (Courtesy of NASA/ESA.)

gravitational potential energy. For N_S stars of individual mass M_S the star cluster has a total mass of $M_C = N_S M_S$, and it will be gravitationally bound together in a stable configuration if:

$$\frac{1}{2} M_S < V_S >^2 = \frac{G M_C M_S}{2 R_C}, \tag{4.15}$$

and

$$< V_S > = \left[\frac{G N_S M_S}{R_C} \right]^{1/2} = \frac{V_{esc}}{\sqrt{2}} \tag{4.16}$$

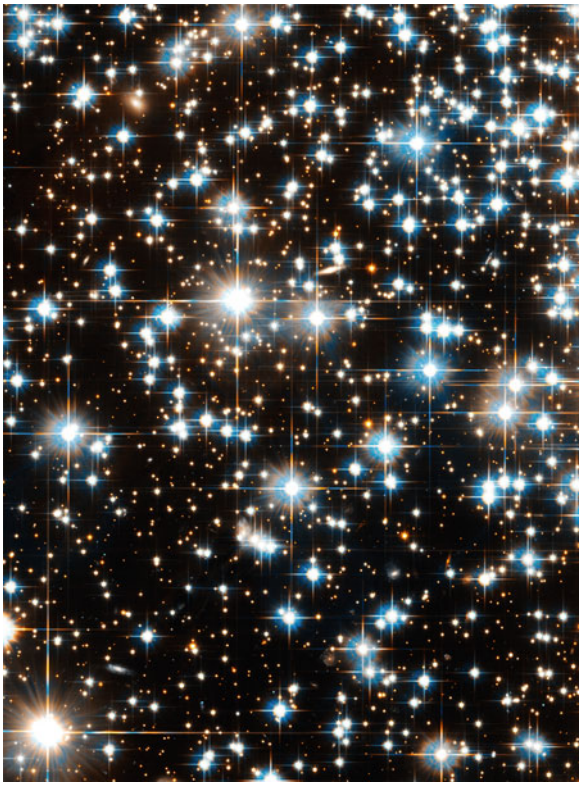


Fig. 4.4 Faint stars in a globular cluster This five-day exposure from an instrument aboard the *Hubble Space Telescope* includes the faintest detectable stars in the globular star cluster NGC 6397, which is located about 8,500 light-years away from the Earth. Some of these objects are white dwarf stars, the collapsed, burned-out relics of former stars like the Sun. White dwarfs cool down at a predictable rate, which can be used to measure the age of this globular cluster, estimated to be about 12 billion years. The crossed lines radiating from the bright stars are diffraction spikes caused by the struts that support the telescope mirror. (Courtesy of NASA/ESA/Harvey Richer, University of British Columbia.)

where V_{esc} denotes the escape velocity of the cluster, R_C is the radius of the star cluster, V_S is a star's velocity and the brackets $< >$ denote a time average with a time-averaged stellar speed of $<V_S>$, and the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

Example: How fast do stars move in a bound star cluster?

The number of stars, N_S, in a globular star cluster can be about a million, or $N_S = 10^6$, each with a mass, M_S, about equal to that of the Sun $M_S = M_\odot = 1.989 \times 10^{30}$ kg. They are apparently bound together in a sphere with a radius of $R_C = 10$ pc $= 3.086 \times 10^{17}$ m. According to the virial theorem, the kinetic energy of the stars, moving at an average star velocity $<V_S>$, must balance just half the gravitational pull of all the stars on any one star, or that $M_S <V_S>^2/2 = GN_SM_S^2/(2R_C)$, where the Newtonian gravitational constant $G = 6.674 \times 10^{11}$ m^3 kg^{-1} s^{-2}. Substituting the numbers into this equation we obtain $<V_S> = 2.07 \times 10^4$ m s^{-1} = 20.7 km s^{-1}.

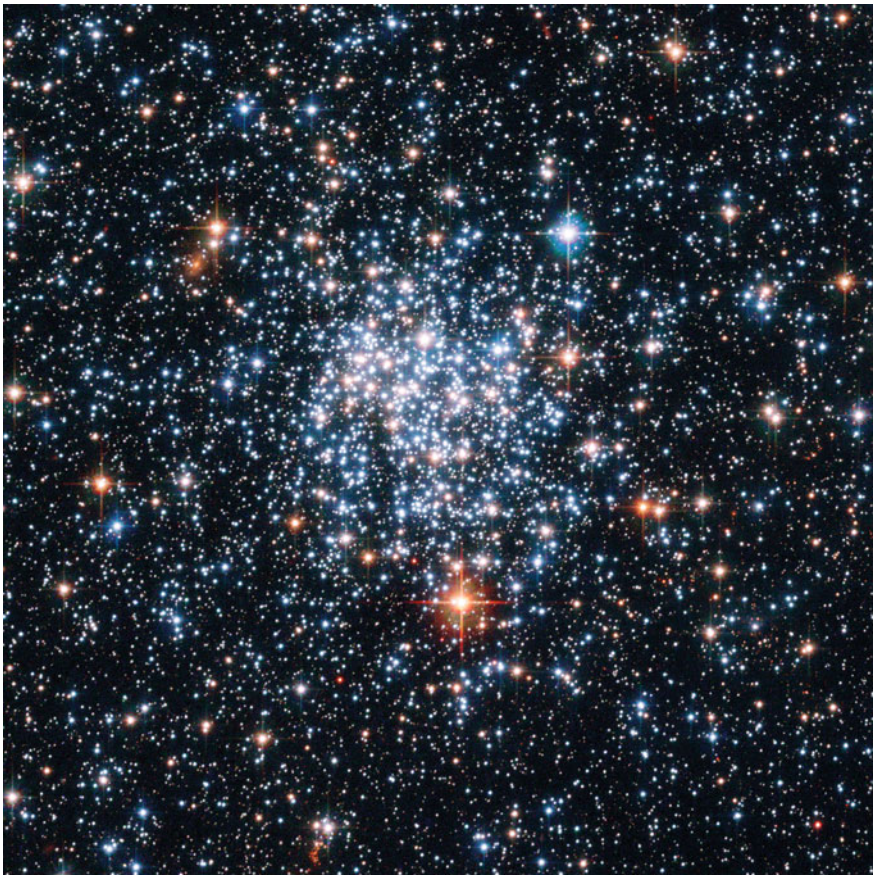


Fig. 4.5 Open star cluster NGC 265 A brilliant cluster of bright blue stars is located in the Small Magellanic Cloud, about 200,000 light-years away and about 65 light-years across. This *Hubble Space Telescope* image subtends an angle of about 70 s of arc. (Courtesy of NASA/ESA/E.Olszewski University of Arizona.)

If the stars move on average at a slower speed than $<V_S>$, they will be pulled gravitationally into each other and the cluster will collapse. If the stars move at an average speed that is faster than $<V_S>$, they eventually will disperse because the cluster cannot hold together. This is what is happening to open star clusters, and to star associations that are bound even more loosely. In fact, some stars are moving out of certain stellar associations at unexpectedly rapid speeds.

4.3.7 Runaway Stars

Some stars race through space with an abnormally high velocity relative to the surrounding interstellar medium. These high-speed stars are known as *runaway*

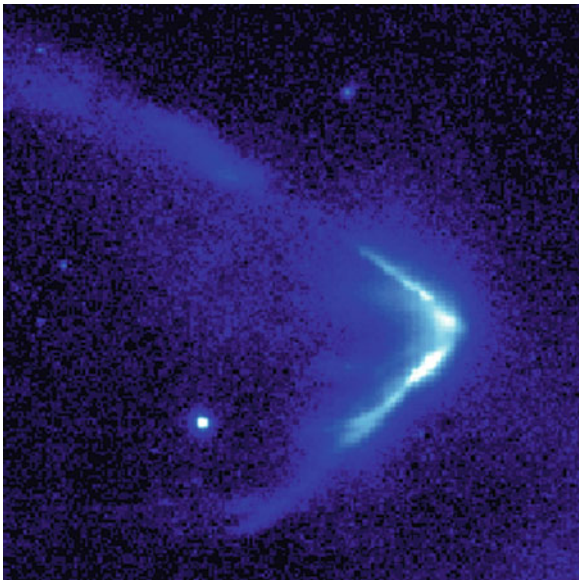


Fig. 4.6 Runaway star A high-speed star slams into dense interstellar gas, creating a bow shock wave that may be a million kilometers wide. The star is thought to be relatively young, only millions of years old. Moving at a speed of about 100 km s^{-1}, it has journeyed 160 light-years since its birth, most likely in a loosely bound stellar association. (Courtesy of NASA/ESA/R. Sahai/JPL.)

stars, because they are moving away from their place of origin. They are former members of very loose star clusters, known as stellar associations, containing bright, and relatively hot, massive and young stars, that are designated as O and B stars.

As first noticed by the Armenian astronomer and statesman Viktor Ambartsumian (1908–1996), these O and B stars are expanding away from one another and from a common origin at speeds of about 10 km s^{-1} (Ambartsumian 1949). The associations are now dispersing and disintegrating, but still moving together in a roughly spherical shape due to their relatively young age. They are not expected to stay together for longer than a few tens of millions of years.

Runaway stars are moving with faster speeds than other stars in the OB associations but with proper motions that often point away from the stellar association to which they once belonged. These runaways are most likely escaped members of former binary star systems that once belonged in the association, until one of the two stars exploded. As described by the Dutch astronomer Adriaan Blaauw (1914–2010), runaway stars are very massive stars whose high space velocities are comparable to the orbital velocities expected for massive binary-star systems (Blaauw and Moran 1954; Blaauw 1961, 1964). Because massive stars burn their thermonuclear fuel faster, and have a shorter lifetime than normal stars, one member of such a binary system will quickly exhaust its thermonuclear reserves and explode as a supernova, thereby releasing the other member as a high-velocity star. The evolution and explosive fate of such massive stars is considered in Sect. 13.5.

The *Hubble Space Telescope* has captured striking images of runaway stars plowing through regions of dense interstellar gas and creating brilliant bow-shock

structures and trailing tails of glowing gas (Fig. 4.6). These features are formed when the stars' powerful stellar winds slam into the surrounding gas. The shocks indicate that the runaway stars are traveling at speeds between 50 and 100 km s^{-1} relative to the dense gas through which they are moving. This is five or ten times faster than the expansion speeds of the stellar associations or the average speeds of stellar motions with respect to nearby stars or the local interstellar medium.

4.4 Cosmic Rotation

In addition to moving through space, an astronomical object also rotates or spins about its axis. This rotation often can be traced back to an origin from a more distended object of slower spin, but sometimes it is related to a glancing collision in the past.

The period of rotation is the time it takes to complete one revolution, or the time for the planet or star to spin into the same orientation in space. For the planets and the Sun, this intrinsic rotation period is known as the sidereal rotation period, from fixed star to fixed star; it has been corrected for any observational effects such as the Earth's orbital motion around the Sun.

4.4.1 Unexpected Planetary Rotation

For solid rocky planets, the rotation period is everywhere the same on the planet's surface. The Earth, for example, rotates once very 24 h or 86,400 s at all latitudes, or at every angular distance north or south of the equator. As a result, all points of the globe take the same amount of time to complete one rotation and a day lasts 24 h everywhere on the planet. If the rotation period differed at different latitudes, the solid planet would break apart.

You might think that it's easy to determine the rotation period of a planet. All you need to do is watch how long it takes for a prominent surface feature to spin around behind the planet and reappear. But this was not the case for Mercury, which is so close to the Sun that most people have never even seen the planet, let alone resolved anything on its surface. And the situation was even worse for cloud-covered Venus, whose surface can never be seen.

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 orbital period. Just as the Earth's 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-m (18-inch) telescope, and he concluded that the same side of the planet did, indeed, always face the Sun

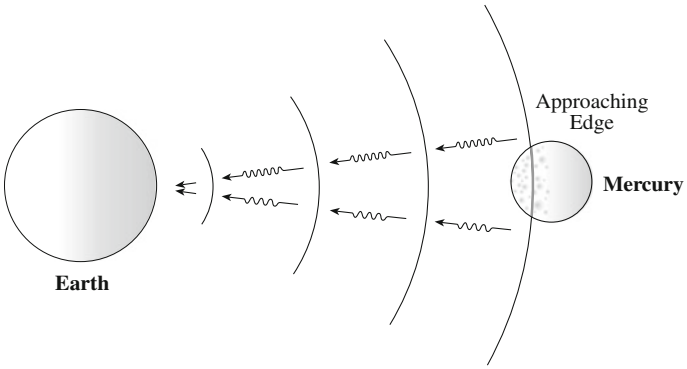


Fig. 4.7 Radar probes of Mercury A radio signal spreads out as a spherical wave, and Mercury intercepts a small fraction of them. As the wave sweeps by the planet, it is reflected in spherical wavelets whose wavelengths are Doppler-shifted by the rotational motion of Mercury's surface. The waves from the receding side are red-shifted toward 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

(Schiaparelli 1889; Defrancesco 1988). For three-quarters of a century, telescopic observers agreed with his conclusion. All of these astronomers were dead wrong!

In 1967, Mercury's true rotational period was determined with radio signals that rebounded from the planet (Dyce et al. 1967). The world's largest radio telescope, located in Arecibo, Puerto Rico, was used to transmit 2 million watts of pulsed radio power at the planet, and to receive the faint echo. This technique is known as radio detection and ranging, abbreviated radar, and it is also used to locate and guide airplanes near airports.

Each pulse was finely tuned, within a narrow range of wavelengths around 0.497 m, and emitted for only about a millisecond. Upon hitting the planet, its rotation de-tuned the pulse, slightly spreading it over a wider range of wavelengths (Fig. 4.7). 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, which arrived back at Arecibo shortly before the next radar pulse was sent. The rotational velocity and period were calculated from the broadened, wavelength-shape of the return echo, using the well-known expression for the Doppler effect.

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 narrow spectral line or a finely tuned radio pulse at wavelength λ by an amount $\Delta\lambda$ given by the expression:

$$\frac{\Delta\lambda}{\lambda} = \frac{V_{rot}}{c}. \tag{4.17}$$

The period, P, of rotation for an object of radius R is $P = 2\pi R/V_{rot}$, and V_{rot} is the rotation velocity.

The radar result for Mercury came as an unexpected surprise. Its 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, which follows from $3 \times 58.646 = 2 \times 87.969$, and it is technically known as spin–orbit coupling. In comparison, the Earth's Moon has a 1:1 spin–orbit resonance in which its rotation period is equal to its orbital period.

The Italian scientist Giuseppe Colombo (1930–1984) provided an explanation for this result in terms of the Sun's varying tidal forces as Mercury revolves about its elongated orbit (Colombo and Shapiro 1966; Goldreich and Peale 1966). 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 3:2 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 1:1 resonance with a rotation period equal to its orbital period.

No human eye has ever gazed on the surface of Venus, which is forever hidden by a thick overcast of impenetrable clouds, but radio waves can penetrate this obscuring veil and touch the landscape hidden beneath. By bouncing pulses of radio radiation off the surface of Venus, the radar astronomers also discovered in 1967 that this planet spins in the backward direction, opposite to that of its orbital motion. That is, unlike the other terrestrial planets, Venus does not rotate in the direction in which it orbits the Sun.

The radar observations also showed that Venus spins with a period longer than any other planet, at 243.018 Earth days. This rotation period is even longer than the planet's 224.7 Earth-day period of revolution around the Sun, so the day on Venus is longer than its year. Tides raised by the Sun in the planet's thick atmosphere may explain why Venus turns very slowly and in the wrong way, but it might have alternatively been knocked into a backwards rotation by a collision with a planet-sized object in the early epochs of the solar system, when such collisions were more common.

Unlike the rocky terrestrial planets, the gaseous giant planets do not rotate at a uniform rate, and this results in a non-spherical shape. The outward force of rotation opposes the inward gravitational force, and this reduces the pull of gravity in the direction of spin. As a result, the giant planets rotate faster in their equatorial middle, where there is a perceptible bulge, and slower at the flattened poles. So they have an oblate shape that is elongated along the equator (Table 4.4). The same thing even happens to the Earth, but by a relatively small amount since it is solid instead of gaseous inside.

The apparent outward force that draws a rotating body away from the center of gravitational acceleration is known as *centrifugal force*, from the *Latin centrum* meaning "center" and *fugere*, meaning "to flee". It tends to push the equatorial

Table 4.4 Oblateness of the giant planets and the Earth[a]

Planet	Equatorial radius, R_e (km)	Polar radius, R_p (km)	Oblateness $(R_e–R_p)/R_e$
Earth	6,378.140	6,356.755	0.003353
Jupiter	71,492	66,854	0.0649
Saturn	60,268	54,364	0.0980
Uranus	25,559	24,973	0.0229
Neptune	24,766	24,342	0.0171

[a] The radii are given in units of kilometers, abbreviated km. The radii of the giant planets are those at the level where the atmospheric pressure is equal to one bar, the pressure of air at sea level on Earth

regions out. The ratio of the centrifugal acceleration at the equator to the gravitational acceleration at the equator is

$$m = \frac{\omega^2 R^3}{GM} = \frac{4\pi^2 R^3}{P^2 GM} \tag{4.18}$$

for a planet rotating at an angular velocity $\omega = 2\pi/P$, rotation period P, radius R and mass M, where the universal constant of gravitation $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

The rapid rotation of a planet or star might push the equatorial regions out so far that it rips the object apart, and this provides an upper limit to the possible rotation speed and a lower limit to the rotation period (Focus 4.1). If the rotation is too fast, and the equatorial push is too much, there is nothing left to rotate.

Focus 4.1 How fast can a planet or star rotate?

The rotation of a planet or star forces its equatorial regions out, and if the speed of rotation is too fast the object will fall apart. This will happen if the equatorial rotation velocity of the star, V_{rot}, exceeds the escape velocity, V_{esc}, which for an object with mass, M, and radius, R, occurs when:

$$V_{rot} = \frac{2\pi R}{P} \geq V_{esc} = \left(\frac{2GM}{R}\right)^{1/2}, \tag{4.19}$$

where the symbol $\geq$ denotes greater than or equal, the constant $\pi \approx 3.14159$ and the Newtonian gravitation constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1}s^{-2}. Collecting terms in this equation, we see that the break up happens for rotation periods P of

$$P \leq \left(\frac{2\pi^2 R^3}{GM}\right)^{1/2}. \tag{4.20}$$

where the symbol $\leq$ denotes less than or equal. We would obtain the same condition, within a factor of the square root of two, or $\sqrt{2}$, if we set the ratio of the centrifugal acceleration at the equator equal to the gravitational

acceleration at the equator. For a planet or star of uniform mass density $\rho = M/(4\pi R^3/3)$, the mass per unit volume, the upper limit to the period of rotation is:

$$P \leq \left(\frac{8\pi}{3G\rho}\right)^{1/2} = \sqrt{\frac{8\pi}{3G\rho}}. \tag{4.21}$$

For a rocky planet with a mass density of 3,000 kg m^{-3}, like that of the Earth's crust, the fastest possible rotation period is P ≈ 6,470 s ≈ 1.8 h. Most asteroids, for example, do not rotate faster than once every 2.2 h.

Stars with a mean mass density like that of the Sun, at 1,410 kg m^{-3}, would be expected to rotate with periods longer than about 2.6 h, but stars do not have a uniform mass density. Moreover, collapsed stars, like neutron stars and pulsars have very high mass densities approaching nuclear densities of 5×10^{17} kg m^{-3}, which is what happens when you press a solar mass into a star about 10^4 m in radius. Our equation then shows that the fastest pulsar probably has a period of about 0.5 ms, or 0.5×10^{-3} s, and thus rotates about 2,000 times a second. Such stars also have exceptionally high escape velocities owing to their compact size.

The instability of uniformly rotating spherical masses was first described by the Scottish mathematician Colin Maclaurin (1698–1746), and the general result ever since then (Maclaurin 1742) is that a rotating sphere becomes unstable when the angular rotational velocity $\omega = 2\pi/P$ rises above $(G\rho)^{1/2}$ (Tassoul 1978). The detailed theory for rotating fluid masses, as well as gaseous ones, has a long, rich history that can be found in Todhunter (1962) and Chandrasekhar (1969).

Despite its great size, Jupiter rotates so fast that day and night each last about 5 h and its full day is less than one-half Earth day. The precise rotation period of 9.9249 h is found by tracking radio bursts that are linked to the planet's spinning magnetic field, which emerges from deep within the planet. Saturn rotates with a day of only 10.6562 h, which is also inferred from the observed periodic modulation in Saturn's radio emission, generated in its spinning magnetic fields. The visible clouds at different latitudes on both giant planets rotate at different speeds and even in different directions. The rotation periods of the some planets and stars, including the Sun, are given in Table 4.5.

4.4.2 The Sun's Differential Rotation

Observations of sunspots have long indicated that the visible solar disk rotates differently at different latitudes, with a faster rate at the equator than at higher

Table 4.5 Rotation periods and rotation velocities of some planets and stars

Object	Rotation period	Radius[a] (m)	Rotation velocity[a] (m s^{-1})
Earth	0.99727 Earth days[b]	6.378×10^6	4.651×10^2
Earth's Moon	27.322 Earth days[c]	1.738×10^6	4.627
Mercury	58.6462 Earth days	2.440×10^6	3.026
Venus	−243.018 Earth days	6.052×10^6	1.81
Jupiter	9.9249 h	7.149×10^7	1.26×10^4
Saturn	10.6562 h	6.027×10^7	9.87×10^3
Sun (equator)	25.67 Earth days	6.955×10^8	1.97×10^3
Vega	12.5 h	1.933×10^9	2.7×10^5
White dwarf star[d]	186.5 s	6.378×10^6	2.1×10^5
Crab pulsar	0.033 s	10^4	1.9×10^6

[a] The equatorial radius is given when the object has a known oblate shape, and in this situation the equatorial rotation velocity is provided

[b] One Earth day is defined as the time for our planet to revolve once with respect to the Sun, and such a solar day is 24 h or 86,400 s long. The Earth's rotation period with respect to stars, or sidereal time, runs about 4 min slower than the solar day. The sidereal day lasts 23 h 56 min 04 s or 8.6164×10^4 s

[c] The sidereal rotation period of the Earth's Moon, from fixed star to fixed star, is 27.322 Earth days. The time from new Moon to new Moon, known as the synodic month, is 29.53 Earth days

[d] The radius of a white dwarf star is assumed to be equal to that of the Earth and its rotation period inferred from the rotation period of the Sun and conservation of angular momentum, so the period scales as the inverse square of the radius

Table 4.6 Differential rotation of the Sun[a]

Solar latitude (degrees)	Rotation period (days)	Rotation speed (km h^{-1})	Rotation speed (m s^{-1})	Angular velocity (nHz)
0 (Equator)	25.67	7,097	1,970	451
15	25.88	6,807	1,891	447
30	26.64	5,922	1,645	434
45	28.26	4,544	1,262	410
60	30.76	2,961	823	376
75	33.40	1,416	393	347

[a] Data from the MDI instrument aboard the *SOHO* spacecraft

latitudes (Carrington 1863; Newton and Nunn 1951; Snodgrass 1983). This is known as differential rotation, since the rate of rotation differs at different latitudes. It indicates that the Sun is not solid, for it would be torn apart by differential rotation if it was solid; instead, most of the Sun is a gaseous plasma. Gilman (1974) has provided a review of the rotation of the Sun.

The rotation speed of the visible solar disk, the photosphere, can be inferred from the Doppler effect of an absorption line originating there, and the results confirm the differential rotation suggested by sunspot observations (Table 4.6). The Sun spins about its axis with a period of about 25 days at the equator, which corresponds to a rotation speed of about 2 km s^{-1} or roughly 7,200 km per hour. But since the Earth is orbiting the Sun in the same direction that the Sun rotates,

the solar rotation period observed from the Earth is about 27 days. The shorter rotation period is the star's true rotation period, and is technically known as the sidereal rotation period, from fixed star to fixed star; the longer period required for a fixed feature to rotate back to the same position as viewed from Earth, is called the synodic rotation period.

The English amateur astronomer Richard Christopher Carrington (1826–1875) determined the solar rotation rate from low latitude sunspots in the 1850s, and defined a fixed solar coordinate system that rotates once every 25.38 days (Carrington 1863), which would correspond to the sidereal rotation period near the solar equator. To compare locations on the Sun over a period of time, the mean observed, or synodic, solar rotation period has been arbitrarily taken to be 27.2753 days. Each rotation of the Sun is then given a unique number called the Carrington rotation number, with rotation 1 beginning on 9 November 1853.

The synodic equatorial rotation period, as observed from Earth, is 26.75 ± 0.05 days, with a differential synodic rotation given by (Sheeley et al. 1992):

$$\omega(\theta) = 13.46 - 2.7\cos^2\theta + 1.2\cos^4\theta - 3.2\cos^6(\theta), \tag{4.22}$$

where ω is the angular velocity in degrees per day, and θ is the co-latitude measured from the poles instead of the equator. An angular velocity of 13.46 degrees per day corresponds to 451 nHz.

Helioseismologists have more recently used 5 min oscillations of the photosphere, produced by internal sound waves, to investigate the internal structure of the Sun (see Sect. 8.5), and the rotation rate inside the Sun has been measured by a change in the periods of the sound waves. Waves propagating in the direction of rotation are carried along by the moving gas, and move faster than they would in a non-rotating Sun. A bird or a jet airplane similarly moves faster when traveling with the wind and takes a shorter time to complete a trip. The resonating sound-wave crests moving with the rotation therefore appear, to a fixed observer, to have shorter periods. Waves propagating against the rotation are slowed down, with longer periods. These opposite effects make the observed solar oscillation periods divide, and such rotational splitting depends on both the depth and the latitude of the sound waves moving within the Sun.

The solar oscillations have a period of about 5 min, so the rotational splitting is roughly 5 min divided by 25 days, or about one part in seven thousand since 1 day = 1,440 min. The photosphere oscillations have to be measured ten or a hundred times more accurately than this to determine subtle variations in the Sun's rotation, or as accurately as one part in a million.

The solar oscillation data indicate that differential rotation, in which the equator spins faster than the poles, is preserved throughout the outer third of the Sun, known as the convective zone (Fig. 4.8). Within this zone, there is little variation of rotation with depth, and the inside of the Sun does not rotate any faster than the outside at the same latitude. At greater depths, the interior rotation no longer mimics that of sunspots, and differential rotation disappears. The internal accord

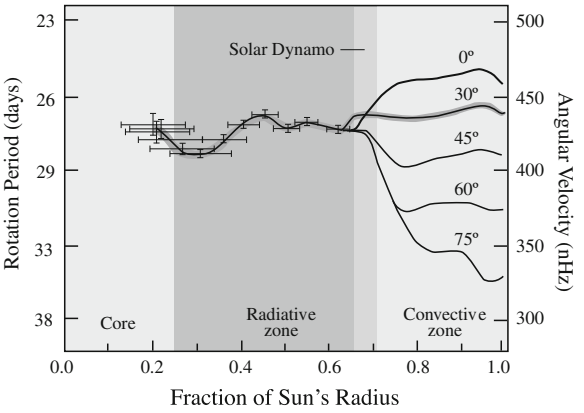


Fig. 4.8 Internal rotation of the Sun The rotation rate inside the Sun, determined by helioseismology using instruments aboard the *SOHO* spacecraft. The outer parts of the Sun exhibit differential rotation, with material at high solar latitudes rotating more slowly than material at equatorial latitudes. This differential rotation persists to the bottom of the convective zone at 28.7 percent of the way down to the center of the Sun. The rotation period in days is given at the left axis, and the corresponding angular velocity scale is on the right axis in units of nanoHertz (nHz), where 1 nHz = 10^9 Hz, or 1 billionth, of a cycle per second. A rotation rate of 320 nHz corresponds to a period of about 36 days (*solar poles*) and a rate of 460 nHz to a period of about 25 days (*solar equator*). The rotation in the outer parts of the Sun is given at latitudes of 0 (*solar equator*), 30, 45, 60, and 75 degrees. Just below the convective zone, the rotational speed changes markedly, and shearing motions along this interface may be the dynamo source of the Sun's magnetism. There is uniform rotation in the radiative zone, from the base of the convective zone at 0.713 to about 0.25 solar radii. The sound waves do not reach the central part of the energy-generating core. (Courtesy of Alexander G. Kosovichev/convective zone/Sebastien Couvidat, Rafael Garcia, and Sylvaine Turck-Chièze/radiative zone. *SOHO* is a project of international cooperation between ESA and NASA.)

breaks apart just below the base of the convective zone, where the rotation speed becomes uniform from pole to pole. Lower down, within the radiative zone, the rotation rate remains independent of latitude, acting as if the Sun were a solid body. Although gaseous, the radiative interior of the Sun rotates at a nearly uniform rate intermediate between the equatorial and high-latitude rates in the overlying solar material. Lang (2009) provides detailed references to pioneering spacecraft studies of the internal rotation of the Sun; also see the review by Thompson et al. (2003).

Thus, the Sun's internal rotation velocity changes sharply at about one third of the way to its center, where the outer parts of the radiative interior, which rotate at one speed, meet the overlying convective zone, which spins faster in its equatorial middle. The transition between these two different regimes takes place in a narrow region of strong rotational shear that most likely plays an important role in the generation of the large-scale solar magnetic field. The differential rotation of other stars may also play a significant role in the generation of their magnetic fields.

4.4.3 Stellar Rotation and Age

The rotation of stars other than the Sun is inferred from the Doppler broadening of their spectral lines. However, astronomers did not realize at first that some stars rotate fast enough for such measurements to be meaningful (Abney 1877). The Doppler effect of atoms moving in the hot stellar atmosphere once was thought to be substantially greater than that of stellar rotation. The Sun, for example, has an equatorial rotational speed of about 2 km s^{-1}, but the thermal speed of hydrogen atoms in the Sun's visible disk, at a temperature of 5,780 K is about 12 km s^{-1}. The Doppler broadening of the hot, moving atoms would therefore be greater than that attributed to rotation, and therefore make the rotational motion very difficult to detect. The only reason that solar rotation could be measured using spectral lines, rather than sunspots, was that the Sun is resolved and one can map out the speeds at various places along the visible disk, but that is not possible for almost all other stars that remain unresolved with even the best telescope.

The turning point came in a seminal paper by two Russian astronomers, Grigory Ambramovich Shajn (1892–1956) and Otto Struve (1896–1963), who showed that relatively young stars rotate faster than older ones and therefore exhibit exceptionally broad spectral lines. They concluded that some stars have equatorial rotational velocities ranging up to 100 km s^{-1} (Shajn and Struve 1929; Struve 1930).

The observed component of radial velocity, V_r, depends on the inclination, i, of the star's pole to the line of sight, with $V_r = V_e \sin i$, where V_e is the rotational velocity at the equator. The inclination is often unknown, so the measurements give a minimum value for the star's rotation velocity, sometimes referred to as the projected rotational velocity.

Stars with rapid rotation are massive, hot and young (Slettebak 1949, 1954, 1955). They include the bright stars Achernar, Alpha Arae, Pleione, and Vega (Peterson et al. 2006), with respective equatorial rotational velocities of up to 300, 470, 329, and 274 km s^{-1}. These stars are rotating so rapidly that their equators bulge outward, giving them a flattened shape. Achernar, for example, is thought to have an equatorial diameter that is about 50 % greater than the distance between its poles. The stars with rapid rotation exhibit intense x-ray emission, due to hot coronae and presumably intense magnetic fields (Pallavicini et al. 1981).

Less massive, cooler and older stars like the Sun rotate with much slower speeds of 10 km s^{-1} or less, and do not exhibit a pronounced equatorial bulge. These stars most likely were formed with fast rotations like the more massive stars, but have slowed down as they aged. Stellar magnetic fields coupled to the surrounding interstellar material act as magnetic brakes over long time intervals. The Sun, for example, probably originated 4.6 billion years ago with a rotation velocity of about 100 km s^{-1}; its magnetism helps to explain why it is now rotating at about 2 km s^{-1} (Sects. 12.1 and 12.2).

Chapter 5
Moving Particles

5.1 Elementary Constituents of Matter

What is matter made of? To find out, we might try breaking any material object into increasingly smaller pieces until we reach a stage when the smallest piece cannot be broken apart. The last step in this imaginary, successive division of matter suggests the existence of unseen *atoms*, a Greek word meaning "indivisible," or something that cannot be divided further.

In the fifth century BC, for example, the ancient Greek philosopher Democritus (ca 470–ca 380 B.C.), and his mentor Leucippus proposed that all matter is composed of combinations of a small number of separate atoms coming together in different ways. They supposed that all substances are composed of four types of elemental atoms: of air, earth, fire, and water. Mud, for example, could be made from earth and water, and fire could turn water into vapor.

Then about two millennia ago, the Roman poet Titus Lucretius Carus (ca 99 BC–ca 55 BC), or Lucretius for short, wrote a wonderful epic poem, in Latin *De Rerum Natura* or in English *On the Nature of the Universe*, which described these indestructible atoms that are so exceedingly small that they are invisible and infinitely vast in number (Lucretius, 55 BC). To Lucretius, atoms were the building blocks of all that exists. This fundamental idea persists to this day. Everything that we see, from a friend to a tree to the Sun and the stars consist of innumerable atoms, all moving randomly about, colliding, gathering together, and breaking apart again. Atoms are immortal, the ingredients of all that existed in the past and the seeds of everything that will exist in the future.

All ordinary matter is composed of elemental *atoms*, and there is a limited number – just 94 naturally occurring atoms, detected directly on the Earth or in astronomical spectra. These atoms are known also as *chemical elements*, because they cannot be decomposed by chemical means. They are very small and exceedingly numerous. A simple drop of water contains about 100,000 billion, billion, or 10^{23} atoms – close to the number of stars in the universe. An additional

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_5, © Springer-Verlag Berlin Heidelberg 2013

24 atoms have been produced as the result of artificial nuclear reactions within particle accelerators rather than by natural processes.

Atoms combine to form molecules, and there are many more kinds of molecules than there are atoms. The vast numbers of molecules differ only in the kind and relative number of the atoms of which they are constructed. A molecule may be a combination of single chemical atoms, such as the oxygen molecule, O_2, that we breathe, which consist of two oxygen atoms each designated by O. A molecule also may contain different elements, as in water, designated H_2O, which is composed of two atoms of hydrogen, H, and one of oxygen, O. The Earth's transparent atmosphere consists mainly of molecules of oxygen, O_2 (21 %), and nitrogen, N_2 (78 %), with trace amounts of carbon dioxide, CO_2, and water vapor. Methane, CH_4, the natural gas used in a stove, consists of one atom of carbon, C, and four of hydrogen, H. Organic molecules contain more complex combinations of carbon, hydrogen, and other atoms.

But the atom is not indivisible after all. Elemental atoms can be broken into smaller subatomic pieces. The first subatomic particle was found through investigations of electricity. The English physicist Michael Faraday (1791–1867) discovered that the electrical charge carried by different atoms is always an integer multiple of a basic amount, an atomicity of electrical charge – the electron (Faraday 1839, 1844). The concept of such an indivisible quantity of charge was proposed to explain the chemical properties of atoms.

It was in 1894 that the Irish physicist George Johnstone Stoney (1826–1911) coined the word *electron* to describe the fundamental unit of electricity (Stoney 1881, 1894), and electricity is indeed transferred by the flow of electrons. Interactions between electrons hold the atoms in a molecule together in a chemical bond. Similar but much weaker interactions among electrons hold many molecules together.

The English physicist Joseph John Thomson (1856–1940) and his colleagues identified the electron as a particle and determined its charge-to-mass ratio (Thomson 1897a, b). Thomson was studying cathode rays, which carry electrons between electrodes in a tube of gas; he showed that the electrons are deflected when either an external magnetic field or an electric field is applied, which meant that they are electrically charged. Using these curved trajectories, Thomson showed that electrons are very light, roughly 1/1,000th of the mass of the least massive atom, hydrogen. It would take 30 billion, billion, billion, or 3×10^{28}, electrons to make a total mass of just 1 oz, or 28 g.

Thomson received the 1906 Nobel Prize in Physics for his investigations of the conduction of electricity by gases. A few years later, the American physicist Robert A. Millikan (1868–1953) determined the elementary charge of the electron by measuring the electrical force on charged droplets of oil suspended against gravity between two metal electrodes (Millikan 1910, 1913). He was awarded the 1923 Nobel Prize in Physics for this and related work.

The electron, which carries a negative electrical charge and has no known components or substructure, is believed to be a truly elementary particle. The elementary charge of the electron, which is denoted by e, has a value of

$e = 1.602 \times 10^{-19}$ C. The negative electric charge of the electron has a value of $-e$, but no one knows why it has the mass and charge it does. When the electrical current in a house is turned on to light a lamp, about 1 million trillion, or 10^{18}, electrons flow through the wires every second.

In the early 20th century, the New Zealand-born British physicist Ernest Rutherford (1871–1937) and his colleagues showed that radioactivity is produced by the disintegration of atoms, and they discovered that radioactive material emits energetic subatomic particles. When bombarding gold leaf with beams of these particles, they found to their astonishment that about 1 in 20,000 particles bounced right back from where it originated, whereas all of the others passed through the gold. This meant that atoms are largely empty space and that most of the mass of an atom is concentrated in a *nucleus* that is 100,000 times smaller than an atom (Rutherford 1911, 1914). The nucleus of an atom contains less than 10^{-15} of the atomic volume, but it includes almost all the atom's mass.

Within a decade, Rutherford was able to show that the nucleus of different atoms contains various amounts of the nucleus of the simplest atom, hydrogen. He named this nuclear building block a *proton*, from the Greek for "first", since it was the first nuclear particle to be discovered.

A proton is positively charged, with a charge of $+e$, equal in amount to that of an electron but opposite in charge. It is positive. Now, particles with an opposite sign to their electric charge attract one another. So electrons and protons are always attracted to each other. Negatively charged electrons surround positively charged protons in an atom, and the total positive charge of the protons is equal to the total negative charge of the electrons. An atom has no net electrical charge and it is electrically isolated from external space.

Particles with the same electrical charge are driven apart by an electrical repulsion. Rutherford therefore postulated the existence of an uncharged nuclear particle, later called the *neutron*, to help hold protons together in the atomic nucleus and prevent the protons from dispersing as they repelled each other. After an eleven-year search, the English physicist James Chadwick (1891–1974) discovered the neutron (Chadwick 1932a, b).

Protons and neutrons collectively are named *nucleons*, because they are the two constituents of the atomic nucleus. They consist of yet smaller components, known as *quarks*. So, the proton and the neutron are not truly elementary particles. Nevertheless, the nuclear fusion reactions that make stars shine can be understood by assuming that all atomic nuclei are composed of protons and neutrons.

These nucleons are bound together in an atomic nucleus by an exceptionally strong force, the nuclear force or nucleon–nucleon force, which allows them to cling tightly to one another and build the dense, compact atomic nucleus. Although powerful, this attractive force has a short range, operating over very limited distances. The strong force decreases to insignificance at distances greater than about 1 million billionths, or 10^{-15}, of a meter, and closes the nucleus in at an atom's center. The nucleons cannot be pushed any closer together, and this sets the physical size of the nucleus.

Therefore, an atom is largely empty space, like the room in which we are sitting. A tiny, heavy, positively charged nucleus lies at the heart of an atom, surrounded by a cloud of relatively minute, negatively charged electrons that define most of an atom's size and govern its chemical behavior.

The radius of an atom is approximately 10^{-10} m, and it consists of a swarm of electrons orbiting a nucleus, whose radius is about 10^{-15} m for the nucleus of the hydrogen atom and about 10^{-14} m for the nuclei of the heaviest atoms, such as uranium. The fermi unit of length, where 1 fermi = 1 fm = 10^{-15} m, is used in nuclear physics. The diameter of the proton is 1.75 fm, and its radius is 0.875 fm.

A proton and a neutron have about the same mass, which is nearly 2,000 times that of an electron. To be exact, the mass of the proton and the mass of the neutron are respectively 1,836 and 1,839 times the mass of an electron.

The mass of an atomic nucleus is always less than the sum of the masses of its protons and neutrons because they have expended energy to bind themselves together. It is this binding energy that is released in nuclear reactions.

What holds solid objects together? Why doesn't a chair fall apart when the wind blows on it? All durable material objects that surround us consist of atoms and, given the emptiness of an atom, we might wonder why we cannot easily crush them into smaller entities. The answer is that when we push any two pieces of material together the forces of electrical repulsion between the atomic electrons in their adjacent surfaces resist the pressure.

Although the kilogram is a useful unit of mass for describing large objects, the mass of an atom and the mass of its nucleus, which contains nearly all of the atom's mass, are conveniently measured in atomic mass units, abbreviated a.m.u. or u, where

$$\mathrm{u} = 1\,\mathrm{a.m.u.} = 1.660539 \times 10^{-27}\,\mathrm{kg}, \tag{5.1}$$

which is equal to one twelfth of the mass of the carbon-12 atom, denoted ^{12}C. The lightest atom is hydrogen with an atomic weight of 1.007825 u; the heaviest stable atom is lead 208 with an atomic weight of 207.97665 u. Unstable radioactive atoms like uranium 235 are a bit heavier.

Nuclear astrophysicists like to use energy, E, to express mass, m, through the $E = mc^2$ relation, where c is the speed of light, and the convenient unit of nuclear energy is the MeV, where 1 MeV = 1.602176×10^{-13} J, which is also equal to 1,000 keV; note that the conversion factor 1.602176 is the elementary charge of the electron.

Characteristics of these atomic and nuclear particles are given in Table 5.1.

The simplest and lightest atom consists of a single electron circling around a nucleus composed of a single proton without any neutrons. This is an atom of hydrogen. Most of the universe, and the majority of the stars, is composed mainly of hydrogen. The nucleus of helium, the next most abundant atom in the cosmos, contains two neutrons and two protons; so it naturally has two electrons that balance the electrical charge of the protons (Fig. 5.1).

So an atom is composed of a dense and massive nucleus containing protons and neutrons, and surrounded by electrons. As in the example of helium, electrically

Table 5.1 Physical properties of electrons, protons, neutrons, and atoms[a]

Electron

m_e = mass of electron = 5.4858×10^{-4} u = 9.10938×10^{-31} kg = 0.5109989 MeV/c^2

e = elementary charge = 1.6022×10^{-19} C; electron charge = $-e$

r_e = classical electron radius = 2.8179×10^{-15} m

σ_T = Thomson cross section $\frac{8\pi}{3} r_e^2 = 6.65246 \times 10^{-29}$ m^2

Atomic nucleus (*protons*, *neutrons*, *alpha particles*)

m_P = mass of proton = 1.007276466 u = $1.6726218 \times 10^{-27}$ kg = 938.27203 MeV/c^2

e = charge of proton = $+1.602 \times 10^{-19}$ C

m_n = mass of neutron = 1.008664916 u = $1.6749274 \times 10^{-27}$ kg = 939.56536 MeV/c^2

m_α = mass of alpha particle = mass of helium nucleus = 4.001506179 u = $6.6446567 \times 10^{-27}$ kg = 3727.37924 MeV/c^2

Z = total number of protons in nucleus = atomic number

A = total number of protons and neutrons in nucleus = atomic mass number

R = nuclear radius = $r_0 A^{1/3}$ for an atom with mass number A and $r_0 = 1.25 \times 10^{-15}$ m

Atom

a_0 = Bohr radius = 5.2918×10^{-11} m

u = atomic mass unit = $1.66053886 \times 10^{-27}$ kg = 931.494061 MeV/c^2

m_H = mass of hydrogen atom = 1.007825 u = 1.6739326 kg

m_{He} = mass of helium atom = 4.002602 u = 6.646476×10^{-27} kg

Δm_{He} = mass defect of helium atom = $m_{He} - 2m_P - 2m_n - 2m_e$ = 0.030378 u

[a] The mass, m, values are given in atomic mass units u = 1.660539×10^{-27} kg, in kilograms or kg, and as the rest mass energy divided by the square of the speed of light $c = 2.9979 \times 10^8$ m s^{-1}; this energy is given in units of MeV = $1.60217646 \times 10^{-14}$ J

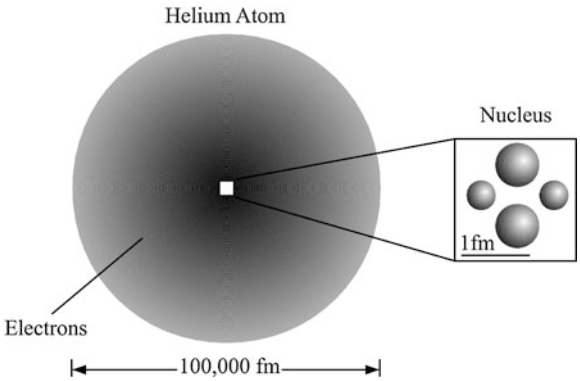


Fig. 5.1 The helium atom An atom of helium contains two electrons that swarm about the atom's nuclear center in a cloud of largely empty space. The *shading* shows that the electrons can be anywhere but are most likely to be found near the center of the atom. The magnified nucleus of the helium atom consists of two protons and two neutrons bound together by a strong nuclear force. The nucleus and each of its four particles are spherically symmetrical. The size of the helium nucleus is about 1 fermi, or 1 fm, which is equivalent to 10^{-15} m. The atom is about 100,000 times bigger than the nucleus, with an atom size of about 10^5 fm, or 10^{-10} m. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

neutral atoms have as many electrons as there are protons and are therefore without net charge. An ionized atom has fewer electrons than protons, and therefore has a positive charge. The atomic number, designated by the symbol *Z,* is equal to the number of protons in an atom's nucleus. Hydrogen has an atomic number of 1, helium 2, carbon 6, nitrogen 7, oxygen 8, lead 82, and uranium 92. The atomic mass number, designated *A*, is equal to the total number of protons and neutrons in the nucleus. The radius, *R*, of the nucleus of an atom of atomic mass number *A* is $R = r_0 A^{1/3}$, where $r_0 = 1.25 \times 10^{-15}$ m = 1.25 fermi = 1.25 fm, so a heavier atom has a bigger, and more massive nucleus.

The atoms of a particular element all have the same number of protons in their nucleus, but the number of neutrons can vary, giving rise to different isotopes of the same element. Elements of atomic number 83–94 are composed entirely of radioactive isotopes that are unstable and are known to decay into other elements.

5.2 Heat, Temperature, and Speed

5.2.1 Where Does Heat Come From?

Heat is a form of energy caused by the motion of tiny unseen particles, such as the molecules in a gas, which are in a state of ceaseless motion. These particles move randomly in all directions and do not contribute to the overall motion of the gas in which they reside. The faster they move, the hotter the body.

The energy of motion is called kinetic energy, after the Greek word *kinesis* meaning "motion" – the word cinema has the same root, referring to motion pictures. In a star, individual gas particles are changing direction constantly in an irregular zigzag trajectory, and the heat can be enormous. However, all of the particles in a star move together in the same overall direction, and they are responsible for the bulk kinetic energy of star motion (Fig. 5.2).

All gas molecules are always moving, and the hotter they become the faster they move and the greater their kinetic energy. The lowest possible temperature is absolute zero, or zero on the Kelvin scale denoted K (Kelvin 1848). At absolute zero, molecules cease to move and are completely at rest; they have no kinetic energy. At this temperature, the constituent particles stick together and behave as a frozen solid, resembling ice.

Raise the temperature above absolute zero and molecules move about and collide, turning ice into liquid water. Further increase the temperature and molecules move so fast that they overcome the cohesive forces that bind them together. Gas is formed, capable of nearly unlimited expansion in all directions as when water evaporates.

So, by adding more heat we can transform a solid into a liquid and then a gas, and we can reverse the process by removing heat and lowering the temperature. All our familiar objects exist in one of these three fundamental states – the solid,

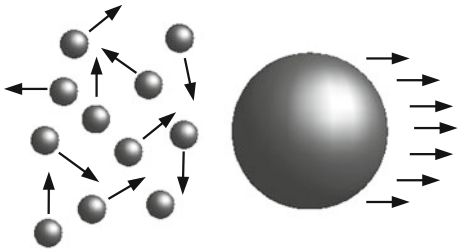


Fig. 5.2 Random and ordered motion Particles within a hot gas (*left*) move here and there in random directions that continually change as the result of collisions between particles. This supports the gas against inward gravitational forces. A planet or star (*right*) moves along a well-defined, ordered trajectory determined by external gravitational forces on it. When a large number of stars has gathered together and is confined within a star cluster, the stars also move in random directions, supporting their combined gravitational pull. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

liquid and gaseous ones. At higher temperatures within cosmic objects, there is a fourth state of matter, known as plasma, in which the atoms are torn into their subatomic ingredients. The range of temperatures found in the universe is illustrated in Table 5.2.

The German physician Jules Robert Mayer (1814–1878) reasoned that heat is a form of energy (Mayer 1842) – generally called "force" in his time and related to the motivating force of fire (Carnot 1824). Mayer found that heat energy can change form, and this had a crucial role in the discovery of the *conservation of energy*. Heat energy can be produced by or transferred into another type of energy, but the energy never disappears. The total energy is conserved (Mayer 1842; Helmholtz 1847).

The English physicist James Prescott Joule (1818–1889) soon provided experimental verification of the conservation of energy for particular cases, in a

Table 5.2 Range of cosmic temperatures

Location	Temperature (K)
Absolute zero	0
Cosmic microwave background radiation	3
Water's freezing (triple) point	273
Water's boiling point	373
Incandescent light bulb	2,500
Visible solar disk	5,780
Center of Sun	1.5×10^7
Atom bomb	3.5×10^8
CERN particle accelerator (proton–proton collision)	10^{13}
Big bang (at 10^{-44} s)	10^{32}

lecture titled "On Matter, Living Force, and Heat" (Joule 1847, 1850). Today, we identify Joule's "living force" with kinetic energy. The joule unit of energy, abbreviated J, is appropriately named after him.

The German physician and physicist Hermann von Helmholtz (1821–1894) provided the connection between the conservation of energy and the kinetic energy of motion, suggesting that the Sun could gain heat from its gravitational energy by contracting (Helmholtz 1847, 1856, 1908). This mechanism would keep the Sun shining for tens of million of years, but it was eventually realized that the Earth and Sun are 4.6 billion years old, and another source of energy, due to nuclear reactions, was required to conserve energy in the Sun for such long times.

5.2.2 Thermal Velocity

If we bring a hot body into contact with a colder one, the fast-moving particles in the hot body collide at the boundary with the slower-moving particles of the colder body. This transfers to them a part of the kinetic energy. The fast-moving particles gradually slow down and the slow ones accelerate, averaging out the temperature differences until a state of thermal equilibrium is obtained, and then a single temperature characterizes the situation.

The portion of an object's internal energy that is responsible for its temperature is called thermal energy, and for an equilibrium temperature, denoted by T, the average *thermal energy* given by:

$$\text{Thermal energy} = \frac{3}{2}\,kT, \tag{5.2}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. This constant is named after the Austrian physicist Ludwig Boltzmann (1844–1906), whose doctoral thesis pioneered the kinetic theory of gases (Boltzmann 1868).

A particle is said to move at the *thermal velocity* when its kinetic energy is equal to its thermal energy. For a molecule or particle of mass m and velocity V the kinetic energy of motion is:

$$\text{Kinetic energy} = \frac{1}{2}mV^2. \tag{5.3}$$

When this is set equal to the thermal energy, we obtain the thermal velocity, $V_{thermal}$, of a particle of mass, m, at temperature, T:

$$V_{thermal} = \sqrt{\frac{3kT}{m}} = \left[\frac{3kT}{m}\right]^{1/2}. \tag{5.4}$$

The magnitude of the thermal velocity, which is the thermal speed, increases with the temperature of the gas and decreases with increasing particle mass. Hotter particles move faster and more massive ones move slower.

Example: **Thermal velocity in our atmosphere and in the Sun's visible disk**

At sea level the temperature T is about 288 K, and our air is predominantly composed of nitrogen molecules of mass $m = 2 \times 14 \times$ u, where the atomic mass unit u $= 1.66054 \times 10^{-27}$ kg. Substituting these numbers into our expression for thermal velocity, $V_{thermal} = (3kT/m)^{1/2}$, where the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1}, we obtain a thermal speed, the magnitude of the thermal velocity, of about 507 m s^{-1}. For the visible solar disk, the temperature T is 5,780 K and the abundant hydrogen atoms of mass $m_H = 1.67 \times 10^{-27}$ kg, have a thermal speed of about 12,000 m s^{-1}. They move faster because the hydrogen atoms on the Sun's disk are both hotter and less massive than the nitrogen molecules in our air.

Protons and electrons are perpetually steaming away from the Sun in the solar wind. How hot would they have to be to escape from the Sun's gravitational pull? As we saw in Sect. 4.1, the escape speed of the Sun at its visible disk is $V = V_{esc\odot} = (2GM_\odot/R_\odot)^{1/2} = 6.18 \times 10^5$ m s^{-1}. If that speed was equal to the thermal velocity $V_{thermal} = (3kT/m)^{1/2}$, then the temperature $T = mV^2/(3k)$, where the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1} and m is the mass of the particle. For a proton of mass $m_P = 1.672 \times 10^{-27}$ kg, the temperature is $T_P = 1.54 \times 10^7$ K, and for the electron of mass $m_e = 9.109 \times 10^{-31}$ kg, the temperature is $T_e = 8.4 \times 10^3$ K. The outer solar atmosphere, the corona, which lies just above the visible solar disk, has a million-degree temperature, so it is hot enough for the electrons to escape from the Sun, and nearly hot enough for the protons to do so. Since the escape velocity falls off with increasing distance from the Sun, this also becomes possible.

For the Earth, we have an escape velocity of $V = V_{escE} = (2GM_E/R_E)^{1/2} = 1.12 \times 10^4$ m s^{-1}. Hydrogen atoms do not remain in the Earth's atmosphere, and since their mass is approximately the same as that of their nuclear proton, the temperature required for this escape is $T = mV^2/(3k) \approx 5 \times 10^3$ K. The mean surface temperature of the Earth is about 281 K, not hot enough for hydrogen to escape immediately from the surface. However, the temperature of the Earth's atmosphere increases at higher levels to more than 10^4 K, in its ionosphere, permitting hydrogen to leak off the planet by thermal evaporation, which has been detected from satellite observations as a geocorona.

The equilibrium temperature is sometimes called the *kinetic temperature*, denoted by T_K, and it is given by:

$$\text{Kinetic temperature} = T_K = \frac{mV^2}{3k}. \tag{5.5}$$

5.2.3 Collisions

The numerous particles in a gas are always colliding with each other, and another German scientist, Rudolf Clausius (1822–1888), introduced the concept of the mean free path between collisions. He asked how far, on average, can a molecule move before it comes under the sphere of action of another molecule. When the number of particles per unit volume increases, for example, the mean distance between collisions will decrease.

The average distance covered by a moving particle between successive collisions is called the *mean free path*, denoted l. It is given by (Clausius 1858):

$$l = \frac{1}{(\sigma N)}, \tag{5.6}$$

where N is the number density of gas particles and the cross sectional area $\sigma = \pi r_0^2$ for collision radius r_0. For an atom we might take $r_0 = a_0 = 5.298 \times 10^{-11}$ m, the Bohr radius to obtain $l \approx 10^{20}\, N^{-1}$ m. A molecule would be somewhat larger, depending on the number of atoms it contains.

For a gas at temperature T and pressure P,

$$l = \frac{kT}{\pi r_0^2 P}, \tag{5.7}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. In our atmosphere at sea level, with a temperature $T = 288$ K and a pressure $P = 10^5$ Pa, the mean free path between collisions of nitrogen molecules with radius $r_0 \approx 2 \times 10^{-10}$ m, is about $l \approx 3 \times 10^{-7}$ m, and each air molecule is subjected to many billions of collisions every second.

The approximate mean time, τ, between collisions is given by the ratio of l divided by the thermal velocity, or:

$$\tau \approx l\left(\frac{m}{3kT}\right)^{1/2}, \tag{5.8}$$

where l is the mean free path, m is the particle mass, and k is the Boltzmann constant.

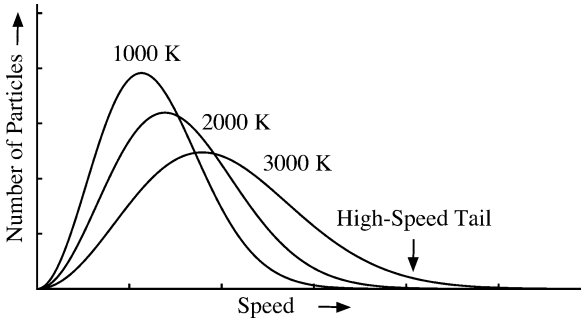


Fig. 5.3 Maxwell distribution of particle speeds The speeds of particles with the same mass and three different temperatures. The peak of this distribution shifts to higher speeds at higher temperatures. There is a small fraction of particles having high speed, residing in the high-speed tail of the distribution, and this fraction increases with temperature. The fraction of particles with low speed becomes smaller at higher temperatures but does not vanish. The peak also shifts to higher speeds at lower mass when the temperature is unchanged. The Scottish scientist James Clerk Maxwell (1831–1879) derived this distribution in 1873. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

5.2.4 The Distribution of Speeds

Also in the mid-nineteenth century, the Scottish physicist James Clerk Maxwell (1831–1879) introduced a statistical approach to the kinetic theory of gases, which recognizes that every gas particle has a different speed and that each collision between particles changes the speeds of those particles. He proposed that the numerous collisions between the large numbers of molecules in a gas produce a statistical distribution of speeds, in which all of the speeds might occur with a different and known probability. This is called the *Maxwell speed distribution*.

In thermal equilibrium, the average value of the kinetic energy of particles in a gas might be distributed equally among all of the particles, but this equality is only statistically true. Most particles move with the same average speed, but not all of them. Some are faster than average and others slower. Gas particles can gain or lose speed by collisions with one another, so they do not all move at the same average speed. In any given instant, the speed and kinetic energy of most of the particles are close to the average value, but there is always a small percentage that moves faster or slower than the average.

The Maxwell speed distribution, illustrated in Fig. 5.3, gives the fraction of gas molecules, or other particles, moving at a particular speed (the magnitude of the velocity vector) at any given temperature and mass (Maxwell 1860). The most probable speed is close to the average in thermal equilibrium; therefore, the most likely speed for any given type of particle increases with its temperature. However, there is a range of speeds, both higher and lower than the average value, and this range also increases with the temperature. In other words, the Maxwell distribution

function becomes broader and its peak shifts to higher speeds when the temperature rises.

Although this distribution function appears to be symmetrical, it has a high-speed tail. Particles in the high-speed tail have greater kinetic energy than other particles in the distribution, and they have an important role in the nuclear-fusion reactions that make the Sun and other stars shine.

The Maxwell speed distribution applies to all types of particles in thermal equilibrium, as long as there are many of them. In addition to atoms or molecules, it can be used to describe the speeds of numerous subatomic particles inside the Sun, which have been freed from their atomic bonds at very high temperatures. The distribution also describes the speeds of millions of stars that are collected together in star clusters, in which the stellar motions are analogous to those of gas particles.

The number of particles, $N(V)dV$, with speeds between V and $V + dV$ is given by:

$$N(V)dV = N_{tot}f(V)dV \tag{5.9}$$

where N_{tot} is the total number of particles of a given mass in the system. The Maxwell speed probability density distribution, $f(V)$, is given by (Maxwell 1860):

$$f(V) = 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2} V^2 \exp\left[\frac{-mV^2}{2kT}\right], \tag{5.10}$$

or equivalently

$$f(V) = \left[\frac{2}{\pi}\left(\frac{m}{kT}\right)^3\right]^{1/2} V^2 \exp\left[\frac{-mV^2}{2kT}\right], \tag{5.11}$$

where T is the temperature, m is the mass of the molecule, or particle, under consideration, and the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. This function has three parts, a constant term, a V^2 term, and an exponential term.

The most probable speed, V_P, is obtained by setting the differential $df(V)/dV$ equal to zero and solving for V, which yields

$$V_P = \sqrt{\frac{2kT}{m}}. \tag{5.12}$$

The *most probable speed* is the speed associated with the highpoint in the Maxwell distribution, and it is the speed most likely to be possessed by any particle of mass m at temperature T. For the diatomic nitrogen molecule at sea level temperature of 288 K, we have a most probable speed of $V_P = 414$ m s^{-1}, since the mass $m = 2 \times 14 \times$ u for atomic mass unit u $= 1.66054 \times 10^{-27}$ kg.

Only a small fraction of the particles have the most probable speed, since there is also a range of speeds, both higher and lower than the most probable one, and this range also increases with the temperature. In other words, the Maxwell distribution function becomes broader and its peak shifts to higher speeds when the temperature rises.

A measure of the range of possible speeds, or the spread of the Maxwell distribution, is the full width to half maximum, ΔV, of the exponential term, which is a Gaussian function of the form $\exp[-V^2/(2\sigma^2)]$ of standard deviation σ. Such a function has $\Delta V = 2.355\ \sigma$, and for the Maxwell distribution we have:

$$\Delta V \approx 2.355(kT/m)^{1/2} \approx 1.66 V_p. \tag{5.13}$$

This is only approximate since there is an enhancement at high velocities owing to the V^2 term in the Maxwell distribution, but the main point is that both the most probable velocity and the spread in the Maxwell distribution increase with temperature.

The most probable speed is very close to the average speed in thermal equilibrium, since $V_P = (2/3)^{1/2}\ V_{thermal} = 0.816\ V_{thermal}$. The mean speed $< V >$ is the mathematical average, denoted by $< >$, of the speed distribution. It is given by:

$$< V > = \int_0^\infty V f(V) dV \tag{5.14}$$

$$< V > = \left(\frac{8kT}{\pi m}\right)^{1/2} = \frac{2}{\sqrt{\pi}} V_P. \tag{5.15}$$

The root mean square speed, V_{rms}, is the square root of the average squared speed

$$V_{rms} = \left(\int_0^\infty V^2 f(V) dV\right)^{\frac{1}{2}} \tag{5.16}$$

or

$$V_{rms} = \left(\frac{3kT}{m}\right)^{1/2} = < V^2 >^{1/2} = \left(\frac{3}{2}\right)^{1/2} V_P. \tag{5.17}$$

The Maxwell speed distribution can be expressed as an energy distribution for the kinetic energy $E = mV^2/2$. The number $N(E)$ of gas particles with a kinetic energy between E and $E + dE$ is given by:

$$N(E)dE = \frac{2N_{tot}}{\pi^{\frac{1}{2}}(kT)^{\frac{3}{2}}} E^{\frac{1}{2}} \exp\left(-\frac{E}{kT}\right) dE = N_{tot} f(E) dE. \tag{5.18}$$

The total internal energy, U, of a total of N_{tot} particles in thermal equilibrium is given by (Maxwell 1860; Boltzmann 1868):

$$U = \int_0^\infty E N(E) dE = \frac{3}{2} N_{tot} kT. \tag{5.19}$$

Table 5.3 Atmospheres of Venus, Mars, and Earth[a]

	Venus	Mars	Earth
Constituent			
Carbon dioxide, CO_2	96	95	0.038
Nitrogen, N_2	3.5	2.7	78
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 (bar)	92	0.007–0.010	1.0 (at sea level)
Surface temperature (K)	735	183–268	184–330
		Mean 210	Mean 287.2

[a] Percentage composition of atmospheric constituent

Table 5.4 Atmospheres of the giant planets and the Sun[a]

Constituent	Sun	Jupiter	Saturn	Uranus	Neptune
Hydrogen, H_2	84	86.4	97	83	79
Helium, He (atom)	16	13.6	3	15	18
Water, H_2O	0.15				
Methane, CH_4	0.07	0.21	0.2	2	3
Ammonia, NH_3	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. Blanks indicate unobserved compounds. (Courtesy of Andrew P. Ingersoll.)

5.3 Molecules in Planetary Atmospheres

The molecular ingredients of an atmosphere can be determined by observing the unique spectral signatures of different molecules. The atmospheres of the eight major planets in our solar system are mainly composed of molecules of the cosmically abundant atoms – hydrogen, H, carbon, C, oxygen, O, and nitrogen, N, but in various percentages given in Tables 5.3 and 5.4. The main ingredient of the atmospheres of Venus and Mars is carbon dioxide, abbreviated CO_2. Our breathable air contains 21 % oxygen molecules, denoted O_2, and 78 % nitrogen molecules, abbreviated N_2. The atmospheres of the giant planets are mainly composed of molecular hydrogen, denoted H_2.

As suggested by George Johnstone Stoney (1826–1911), 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 (Stoney 1898, 1900). An atmosphere is held near a planet by its gravity, but since gas has a natural tendency to expand into space, only planets with a sufficiently strong gravitational pull can retain an atmosphere of a given composition. The ability of a planet to retain an atmosphere also depends on its temperature, determined by both the

planet's distance from the Sun and the atmospheric greenhouse effect. Lighter, hotter molecules will move faster than heavier, colder ones, and the fast ones will be more likely to escape the planet's gravitational grasp.

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 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.

As shown by John S. Lewis (1941–), the composition of the planets and their atmospheres is intimately connected with their distance from the Sun (Lewis 1974, 2004).

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, or

$$V_{thermal} = \sqrt{\frac{3kT}{m}} = \left[\frac{3kT}{m}\right]^{1/2}. \tag{5.20}$$

where the Boltzmann constant $k = 1.3806 \times 10^{-23}$ J K^{-1}, 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 speeds. Anything will cease to move when it reaches absolute zero on the kelvin scale of temperature.

The thermal velocity can be compared to the planet's escape velocity, V_{esc}, given by

$$V_{esc} = \sqrt{\frac{2GM}{D}} = \left[\frac{2GM}{D}\right]^{1/2}, \tag{5.21}$$

where M is the planet's mass, the universal gravitational constant is $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and D is the distance between the center of the planet and the gas particle. The escape velocities at the surfaces or cloud tops of the planets range between 4 and 60 km s^{-1}.

A planet tends to retain molecules that are moving at velocities less than the planet's escape velocity, and a molecule's velocity increases with temperature. At a given temperature, a molecule's velocity increases with decreasing molecular mass, so lighter molecules move at faster speeds and are more likely to escape a given planet or satellite than heavier ones. The high-speed tail in the Maxwell–Boltzmann distribution 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.

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 every natural planetary satellite in the solar system except Titan.

For the Earth, Mars and Venus, the thermal velocity of all molecules is smaller than the escape velocity, but the lightest, fastest molecules can still 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 molecular escape from the exobase is known as *Jeans escape*, after the British scientist James Jeans (1877–1946) who introduced it (Jeans 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 ocean water do not have enough thermal energy to escape from the liquid, but some of them acquire enough as the result of collisions near the ocean's surface.

At and below the exobase in the atmosphere, collisions between particles drive the speed distribution into a Maxwellian one, 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-speed tail of the Maxwell distribution can leave or exit the planet, hence the name exobase.

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 km above the surface, and calculations indicate that about a billion, billion, billion, or 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. So hydrogen gas is very rare in the Earth's atmosphere, present at about 1 part per million by volume. Nevertheless, it is still the third most abundant element in the Earth's surface, in the form of chemical compounds such as hydrocarbons and water.

Notice that lighter particles are lost by thermal evaporation at a much faster rate than heavier ones. Even over 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} kg, 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} kg. Lammer (2008) has reviewed the atmospheric escape and evolution of the terrestrial planets and satellites.

5.4 Gas Pressure

5.4.1 What Keeps Our Atmosphere Up?

Why doesn't the sky fall down, as Chicken Little once said was happening? After all, the Earth's atmosphere is pulled down by the planet's relentless gravity. The answer is that the atmosphere is warmed by the Sun, so its molecules are in continuous motion and collide with one another, producing a gas pressure that prevents them from falling to the ground. So, the atmosphere holds itself up.

The British scientist Robert Boyle (1627–1691) likened air to a "heap of little bodies, lying one upon another," which acted like springs that resist compression. He discovered that the total pressure exerted by all the little springs is inversely proportional to the volume of space in which they are confined (Boyle 1660). Other things being equal, the product of pressure and volume is conserved, so when a gas is compressed into a smaller volume the pressure rises.

When you compress a gas, the molecules move about more rapidly and collide more often. They are resisting being crowded together and pushed into a confining place. If you remove the confinement, the gas will expand out into surrounding space.

Radiation, wind, or a magnetic field also can produce a pressure, known as radiation pressure, wind pressure or magnetic pressure. For example, radiation pressure of sunlight (Debye 1909) can be used to propel a spacecraft by using a solar sail, and both solar radiation pressure and solar wind pressure push different types of comet tails away from the Sun. The solar radiation pressure can also cause a dust grain to slowly spiral into the Sun; a drag due to the component of radiation pressure tangential to the grain's motion. This drag is known as the *Poynting–Robertson effect* (Poynting 1904; Robertson 1937).

The SI unit of pressure is the pascal, abbreviated Pa, which is a force per unit area with 1 Pa = 1 N m^{-2}, where the newton, abbreviated N, is the unit of force. Atmospheric pressure on the Earth and other planets often is measured in the bar unit of pressure, where 1 bar = 100,000 Pa = 10^5 Pa. The Earth's standard atmospheric pressure at sea level is defined as 1.01325 bar.

The radiation pressure of sunlight on a perfectly reflecting surface at the Earth's distance from the Sun is about 10^{-5} Pa. The pressure of the Sun's winds just outside our planet is a startling 10^{18} Pa, but it drops to about 10^{-13} Pa when the winds spread out to their boundary with interstellar space. The pressure of intense magnetic fields just above the visible solar disk is about 30 Pa.

Some representative pressures, displaying their enormous range in the universe, are given in Table 5.5.

Table 5.5 Range of cosmic pressures

Location	Pressure[a] (Pa)
Interstellar space	10^{-13}
Sunlight radiation pressure at Earth orbit	10^{-5}
Beneath foot of a tarantula	1
Visible solar disk[b]	10
Atmospheric pressure on Mars	10^{3}
Earth's atmosphere at sea level[c]	10^{5}
Inside a champagne bottle	5×10^{5}
Surface pressure of atmosphere on Venus	9×10^{6}
Inside a fully charged scuba tank	10^{7}
Center of the Earth	4×10^{11}
Center of Jupiter	7×10^{12}
Center of the Sun	2×10^{16}

[a] Unless otherwise stated in the location, the pressure is the gas pressure
[b] The gas pressure at the visible solar disk, known as the photosphere, is about the same as the vacuum pressure inside an incandescent light bulb
[c] The standard atmosphere of Earth has a sea-level pressure of 1.01325 bar = 101,325 Pa

5.4.2 The Ideal Gas Law

Under most conditions, an atmosphere behaves like an ideal gas, in which the randomly–moving molecules have no volume, like a point, and do not interact with each other, except by collisions. The ideal gas law for the gas pressure, P_g, is given by:

$$P_g V = N_{tot} kT, \tag{5.22}$$

where V is the volume, N_{tot} is the total number of molecules, the Boltzmann constant $k = 1.3806 \times 10^{-23}$ J K^{-1}, and T is the temperature. Such a relation between pressure, volume and temperature is known as an equation of state.

Using the particle number density $N = N_{tot}/V$, the ideal gas law can also be written

$$P_g = NkT. \tag{5.23}$$

Example: **Gas pressure and number of molecules at sea level in the Earth's atmosphere**
The standard atmospheric pressure on Earth at sea level is equal to 1.01325×10^5 Pa. The gas temperature is 288 K, so the number density of atmosphere molecules at sea level is $N = N_{tot}/V = P/kT = 2.5 \times 10^{25}$ m^{-3}.

The ideal gas law, also known as the perfect gas law, is a simplified equation of state that is a good approximation to many gases under many different conditions.

It describes the pressure within the Earth's atmosphere, and it is also used to specify the gas pressure exerted by subatomic particles in the hot interiors of stars. The ideal gas law was first stated by the French engineer Émile Clapeyron (1799–1864), and derived from kinetic theory by the German physicists August Krönig (1822–1879) and Rudolf Clauisus (1822–1888) (Clapeyron 1834, 1856; Clausius 1850, 1857, 1870).

An alternative expression for the ideal gas law is:

$$P_g = \frac{\rho kT}{\bar{m}} = \frac{\rho kT}{\mu m_H} \tag{5.24}$$

for a gas of mass density ρ and average mass per particle given by:

$$\bar{m} = \rho / N = \mu m_H \tag{5.25}$$

for mean molecular weight μ or the mean particle mass in units of the mass of the hydrogen atom of $m_H = 1.00794$ u $\approx 1.67 \times 10^{-27}$ kg, which is roughly equal to the atomic mass unit u $= 1.6605 \times 10^{-27}$ kg and good enough for order of magnitude estimates.

For a diatomic molecule composed of atoms of atomic mass number A, the mean molecular weight is $\mu = 2A$, which is 28 for molecular nitrogen N_2, where the mass number of the nitrogen atom ^{14}N is 14.

Stellar interiors consist of ionized gas. For a fully ionized hydrogen gas, there will be an equal number of protons and electrons, so the mean mass will be:

$$\bar{m} = \frac{m_e + m_P}{2} = \frac{m_H}{2}, \tag{5.26}$$

where the mass of the electron, m_e, is negligible when compared to the mass of the proton, m_P, which is equal to that of the hydrogen atom $m_H = 1.67 \times 10^{-27}$ kg. For a fully ionized helium gas $\bar{m} = 4m_H/3$.

When the ionized gas is composed of three different kinds of particles, each with its own mass m_i and number density N_i denoted by subscript i, the mean particle mass is

$$\bar{m} = \frac{N_1 m_1 + N_2 m_2 + N_3 m_3}{N_1 + N_2 + N_3} = \frac{\rho}{N}. \tag{5.27}$$

For an ionized gas containing hydrogen, H, helium, He, and an element of mass number A,

$$\frac{\bar{m}}{m_H} = \frac{\rho}{N m_H} = \frac{2}{1 + 3X + 0.5Y} \tag{5.28}$$

where X, Y, and Z represent the concentration by mass of hydrogen, helium and heavier elements and $X + Y + Z = 1$. Their number densities are given by:

$$N_H = \frac{X\rho}{m_H}, N_{He} = \frac{Y\rho}{4m_H}, N_A = \frac{Z_A \rho}{A m_H}, \tag{5.29}$$

where the atomic number $Z_A \approx A/2$ is the mass abundance of an element of atomic mass number A. As an example, the abundances observed in the disk of the Sun are $X = 0.71$, $Y = 0.27$ and $Z = 0.02$; in the solar core nuclear fusion reactions have converted about half the hydrogen into helium and $X = 0.34$, $Y = 0.64$ and $Z = 0.02$.

Any hot gas exerts gas pressure, and the gas pressure will vary with distance from whatever is heating the gas. The Earth's atmosphere, for example, is heated from below, at the warm ground, and from above, by the Sun's radiation. Unlike our atmosphere, the Sun is heated from inside, at the center of its hot dense core.

5.4.3 The Earth's Sun-Layered Atmosphere

Our thin atmosphere is pulled close to the Earth by its gravity and suspended above the ground by molecular motion. And because air molecules are mainly far apart, our atmosphere is mostly empty space, and it always can be 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. Yet, even at the bottom of the atmosphere the density is only about 1/1000th of that of liquid water; an entire liter of this air weighs only 1 g.

At greater heights there is less air pushing down from above, so the compression is less and the pressure and density of air gradually fall off into the vacuum of space. There is a simple formula that expresses the drop in the atmosphere pressure at increasing distance, or radius from a planet center. The gas pressure, $P_g(r)$, at radius, r, is given by the *barometric equation*, also known as the *barometric law*:

$$P_g(r) = P_g(R) \exp\left[-\frac{(r-R)}{H}\right] \tag{5.30}$$

where $P_g(R)$ is the surface pressure at radius R, the height above the ground is $r - R$, and H is the *atmosphere scale height* given by:

$$H = \frac{kT}{\bar{m}g} \tag{5.31}$$

and

$$H \approx \frac{kTR^2}{\bar{m}GM} \tag{5.32}$$

where k is the Boltzmann constant, $\bar{m} \approx \mu m_H$ is the mean molecular mass for mean molecular weight μ, where $m_H = 1.67 \times 10^{-27}$ kg is the mass of the hydrogen atom, and the g is the local acceleration of gravity given by $g = GM(r)/r^2 \approx GM/R^2$ for a planet of mass M, at least as long as $r - R$ is much less than R.

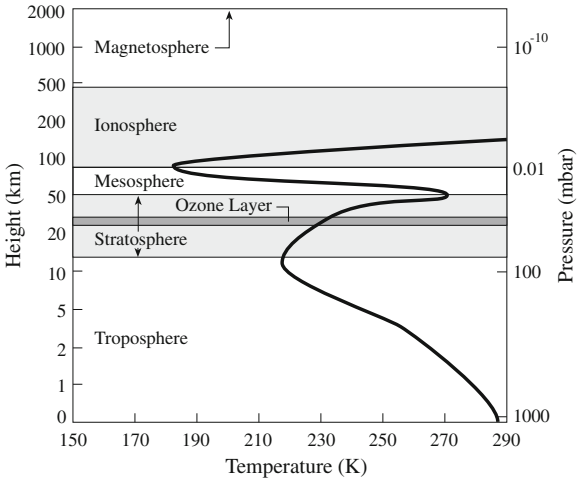


Fig. 5.4 Earth's layered atmosphere The atmospheric pressure (*right scale*) decreases with altitude (*left scale*). This is because fewer particles are able to overcome the Earth's gravitational pull and reach higher altitudes. The temperature (*bottom scale*) also decreases steadily with height in the ground-hugging troposphere, but the temperature increases in two higher regions that are heated by the Sun: the *stratosphere*, with its critical *ozone layer*, and the *ionosphere*. The stratosphere is heated mainly by ultraviolet radiation from the Sun, and the ionosphere is created and modulated by the Sun's x-rays and extreme ultraviolet radiation, which breaks apart atmospheric molecules, and strips electrons from atoms to produce ions. The process of ionization by the Sun's invisible rays releases heat to warm the ionosphere, so the temperature rises with altitude. In the ionosphere, at about 100–500 km above the ground, temperatures skyrocket to higher than anywhere else in the atmosphere. At higher altitudes, the atmosphere thins out into the exosphere, or the "exit to the outside sphere." The temperature is so hot out there and the particles move so fast that some atoms and molecules slowly evaporate

The temperature of the atmosphere also decreases at increasing height near the ground; however, it is not a simple fall-off when we keep going upward. It falls and rises in two full cycles as we move off into space (Fig. 5.4). That's because the Earth's atmosphere is heated from both the warm ground below and from above by ultraviolet and x-ray radiation from the Sun.

The temperature decreases steadily with height in the lowest regions of our atmosphere, since it is heated from the warmer ground below. When warm currents move up from the Earth's surface, they expand in the lower pressure and become cooler. The average air temperature drops below the freezing point of water of 273 K at only a kilometer or two above the Earth's surface, and bottoms out at a height of about 12 km above sea level. This is the greatest height achieved by the air currents. All our weather occurs below this altitude, controlled by visible sunlight.

Global atmospheric circulation, driven by differential solar heating of the equatorial and polar surfaces, creates complex, wheeling patterns of weather in this region, leading to the designation troposphere, from the Greek *tropo* for "turning."

The average extent of the ground-hugging troposphere varies with latitude, from about 16 km above the warm equator to roughly 8 km over the cold poles.

The vertical extent of the troposphere was detected near the end of the nineteenth century when Leon Philippe Teisserenc de Bort (1855–1913), a French meteorologist, launched hundreds of unmanned balloons that carried thermometers and barometers to altitudes as great as 15 km (de Bort 1902). At this height, the temperature no longer decreased with altitude, and seemed to remain nearly constant. If the temperature was unchanging, the ingredients of the atmosphere above the troposphere might settle down into layers, or strata, depending on their weight, so de Bort named this region the stratosphere.

Contrary to everyone's expectations, the temperature increases at greater heights within the stratosphere, rising to nearly ground-level temperatures at about 50 km above the Earth's surface; but we still use the name stratosphere to designate the layer of the atmosphere that lies immediately above the troposphere. The Sun's invisible ultraviolet radiation is largely absorbed in the stratosphere, where the radiation warms the gas and helps make ozone.

Above the stratosphere we come to the mesosphere, from the Greek *meso* for "intermediate." The temperature declines rapidly with increasing height in the mesosphere, from temperatures of about 265 K at 50 km altitude to far below freezing at about 85 km, where the temperature reaches the lowest levels in the entire terrestrial atmosphere.

The mesosphere has been known as the "ignorosphere" because it is too high to be reached by airplanes and too low to be studied by most spacecraft. The air at this height is too thin to support research balloons or aircraft, but thick enough for atmospheric friction, or air drag, to cause satellites to decay quickly from orbit. Sounding rockets pass through this region too rapidly to permit detailed study.

Instruments aboard rockets, as well as radio signals from the ground, have been used to examine the higher levels of the Earth's atmosphere, where the temperatures rise to above those on the ground owing to extreme ultraviolet and x-ray radiation from the Sun. This radiation is energetic enough to break the atmospheric molecules apart, and to strip electrons off their component atoms, producing ions. They reside in the ionosphere (Focus 5.1), where the temperature can rise to about 1200 K.

Focus 5.1 The Earth's ionosphere

On December 12, 1901, the Italian electrical engineer Guglielmo Marconi (1874–1937) startled the world by transmitting a radio signal in Morse code across the Atlantic, from England to Newfoundland. Marconi became an international hero and established the American Marconi Company, which later evolved into the Radio Corporation of America, abbreviated RCA. In 1909 Marconi and the German inventor Karl F. Braun (1850–1918) received the Nobel Prize in Physics for "their contribution to the development of wireless telegraphy."

Because radio waves travel in straight lines, and cannot pass through the solid Earth, no one expected that Marconi could send a radio signal halfway around the world. Radio waves get around the Earth's curvature by reflection from an electrically charged layer, now called the ionosphere, extending into space from roughly 70 to 500 km above the Earth's surface. The atoms in the ionosphere are highly ionized, and many of their electrons have therefore been set free from atomic bonds. As independently shown by Arthur E. Kennelly (1861–1939), then at the Harvard School of Engineering, and Oliver Heaviside (1850–1925) in England, these electrons give the ionosphere a high electrical conductivity, which enables them to turn the radio signals back toward the ground, reflecting them like a metal mirror and not allowing the radio signals to pass through (Kennelly 1902).

The rapid expansion of radio broadcasting in the 1920s, as well as the concurrent development of pulsed radio signals, helped specify the structure and daily variation of the ionosphere. Edward Appleton (1892–1965) and his students measured the height of the reflecting layer by determining the elapsed time between transmitting a radio pulse and receiving its echo from the ionosphere; like all electromagnetic radiation, the radio waves travel at the speed of light. They showed that there are at least three such reflecting layers, now labeled D, E and F, at respective altitudes of 70, 100 and 200–300 km (Appleton 1932; Appleton and Barnett 1925).

In 1947, Sir Appleton was awarded the Nobel Prize in Physics for his investigations of the physics of the upper atmosphere, especially for his discovery of the so-called Appleton layer.

The mystery of exactly what produces and controls the ionosphere was not solved until after World War II, when captured German V-2 rockets were brought to the United States. These and subsequent rockets, built by American engineers, were used by the Naval Research Laboratory to loft detectors above the atmosphere, showing that the Sun emits very energetic radiation at invisible x-ray and extreme ultraviolet wavelengths (Byram et al. 1956). When this radiation reaches the upper atmosphere it breaks the nitrogen and oxygen molecules into their constituent atoms and ionizes them, producing free electrons and atomic ions. The ionosphere above your head therefore develops as the Sun rises and decays as the Sun sets; it lingers on during the night but is not energized then.

The process of ionization by the Sun's invisible rays releases heat to warm the ionosphere, so the temperature increases with altitude in it. Within the ionosphere, the temperatures rise to higher values than anywhere else in the entire atmosphere. Indeed, some scientists prefer to call this region the thermosphere, or "hot" sphere. The thermosphere overlaps with the E and F regions of the ionosphere, beginning at about 90 km and extending upward to about 500 km.

At higher altitudes, the atmosphere thins out into the *exosphere*, or the "exit and outside sphere." In the exosphere the gas density is so low that an atom can

completely orbit the Earth without colliding with another atom. The temperature is so hot out there, and the atoms move so fast, that some atoms can escape the Earth's gravitational pull and travel into outer space. It is therefore the thermosphere at the top of the ionosphere that caps our Sun-layered atmosphere and provides the Earth's threshold into space.

5.4.4 Pressure, Temperature, and Density Inside the Sun

Our Sun is a giant sphere of extremely hot gas, rarefied on the outside and compacted on the inside. Unlike the Earth, it has no solid surface. The high compressibility of the solar gas brings about a rapid increase in density as we go from its visible disk to its center, and as the result of this crowding, the gas particles collide more frequently with higher speeds than elsewhere in the Sun. The compacted gas particles also push more vigorously outward, producing strong gas pressure that keeps the Sun from collapsing.

Within the Sun's dense, central core, the density has increased to 151,300 kg m^{-3}, which greatly exceeds that of any solid or liquid bodies in our everyday environment. The central density is more than ten times greater than the density of solid lead, at 11,340 kg m^{-3}, but still behaving like a gas.

The center of the Sun is just slightly more than 100 times denser than the Sun taken as a whole, at a mean value of 1409 kg m^{-3}. That's the number you get when dividing the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg by its volume, $4\pi R_\odot^3/3$, for a solar radius of $R_\odot = 6.955 \times 10^8$ m. The mean density of the Sun is near that of water, at 1,000 kg m^{-3}, but it is an average density and the Sun is much too hot to be solid or liquid anywhere inside.

The central temperature can be estimated by assuming that a proton at the center of the Sun is hot enough and moving fast enough to counteract the gravitational compression it experiences from the rest of the star. When you do the arithmetic, this balanced condition occurs at a central temperature of 15.6 million K (Sect. 8.2). That is how hot the center of the Sun has to be to avoid collapsing under its own weight, and something has to keep it that hot. The heat is supplied by nuclear reactions at the core of the Sun (Chap. 8).

At this temperature and the central density, the central pressure needed to resist the weight of the overlying gas is 2×10^{16} Pa, or 200,000 million times the pressure of our atmosphere at sea level.

Example: **Gas pressure at the center of the Sun**
The mass density at the center of the Sun is $\rho = 1.51 \times 10^5$ kg m^{-3}, and the mass is provided by protons of mass $m_P = 1.67 \times 10^{-27}$ kg. So the number density of protons at the center of the Sun is $N = \rho/m_P = 0.90 \times 10^{32}$ m^{-3}. The central solar temperature is $T_C = 1.56 \times 10^7$ K. So the gas pressure at the center of the Sun is $P = NkT = 1.9 \times 10^{16}$ Pa, using the

Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. This is about 200 billion, or 2×10^{11}, times as great as the atmospheric pressure at sea level on Earth, at about 10^5 Pa. That's because there is both a much higher temperature and a much higher particle number density in the core of the Sun.

In contrast to both the central and mean densities, the outer layers of the Sun are quite rarefied. This is because there is less overlying material to support at greater distances from the center, so there is a drop in pressure, density and temperature. The compression is less, so the gas gets thinner and cooler (Fig. 5.5). Halfway from the center of the Sun to the visible disk, the density is the same as that of water, and about nine tenths of the distance from the center to the Sun's apparent edge, we find material as tenuous as the transparent air that we breathe on Earth.

At the visible solar disk, the rarefied gas is about one thousand times less dense than our atmosphere at sea level. Out there, in the more rarefied outer parts of the Sun, the temperature has fallen to 5,780 K. Examination of this outer, cooler solar atmosphere tells us about the elemental constituents of the Sun. But in most of the Sun, from its core to its upper atmosphere, there are no atoms; at these hot temperatures the atoms have been torn by collisions into their subatomic ingredients, mainly protons and electrons, to make an ionized gas known as plasma.

5.5 Plasma

5.5.1 Ionized Gas

Plasma has been called the *fourth state of matter*, to distinguish it from the solid, liquid and gas states (Fig. 5.6). The *plasma* is a hot, completely ionized gas, consisting of ions and electrons that have been pulled free of atoms. The high temperatures result in all the atoms losing their normal complement of electrons to leave positively charged ions behind, and it is too hot for the free electrons and ions to join together and form permanent atoms. Because the total negative electrical charge of the free electrons is equal to the total positive charge of the ions, plasma is electrically neutral over a sufficiently large volume.

The ionosphere is plasma. The interiors of most stars are plasma consisting mainly of electrons and protons, but this plasma still behaves like a gas that is described by thermal equilibrium and the Maxwellian speed distribution. Most of the matter in the universe is in the plasma state.

Ions are atoms that have lost one or more electrons, and the ionization energy is the amount of energy needed to remove electrons from an atom. This energy is also known as the ionization potential, and in atomic physics, the ionization

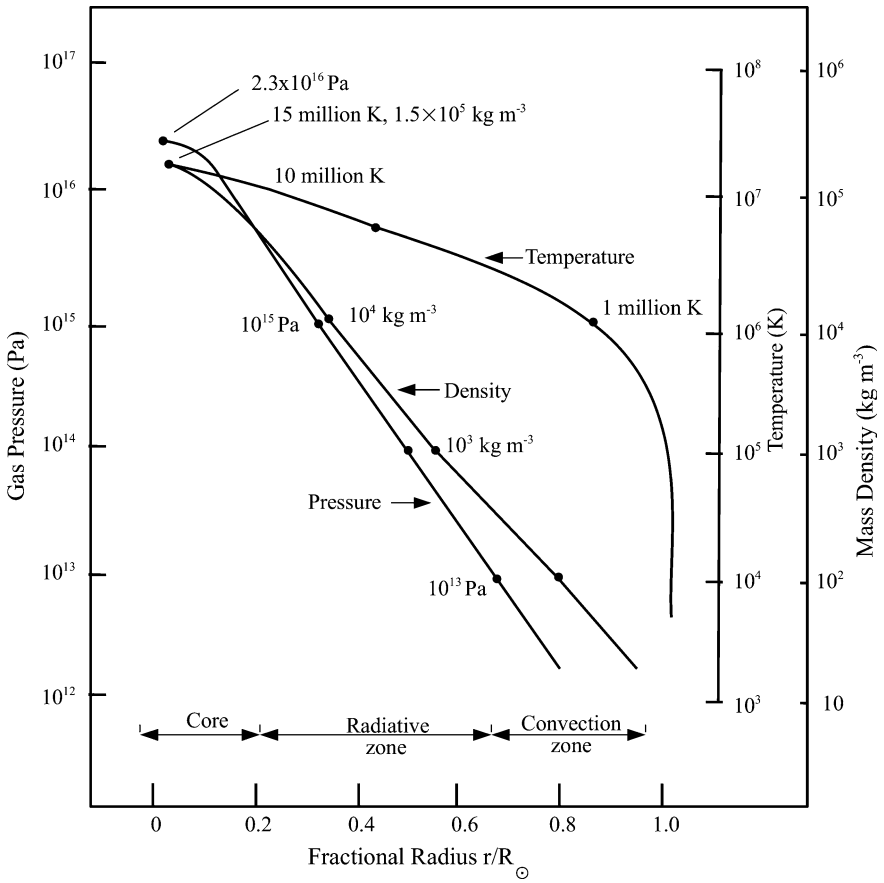


Fig. 5.5 Internal compression of the Sun The variation of pressure, temperature, and mass density with fractional radial distance from the Sun's center (*left*) to its visible disk (*right*). At the center of the Sun, the temperature is 15.6 million K and the mass density is 151,300 kg m^{-3}; the central pressure is 2.33 × 10^{16} Pa, or 233 billion times that of the Earth's atmosphere at sea level (one bar is equivalent to 100,000 Pa). Nuclear reactions occur only in the central core to about 25 % of the Sun's radius. The energy produced in the core is transported by radiation to 71 % of the star's radius, where the temperature has dropped to about 2 million K and the density has fallen to about 200 kg m^{-3}. The energy then is transported by convection out to the Sun's visible disk, known as the photosphere, where the temperature is 5,780 K, and the pressure and density have dropped off the scales of the graph. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

energy, or potential, is measured using the unit of electron volt, abbreviated eV. By definition, the electron volt is the amount of kinetic energy gained by a single unbound electron when it is accelerated through an electric potential difference of 1 V. It is the energy an electron gains when it passes across the terminals of a 1 V battery.

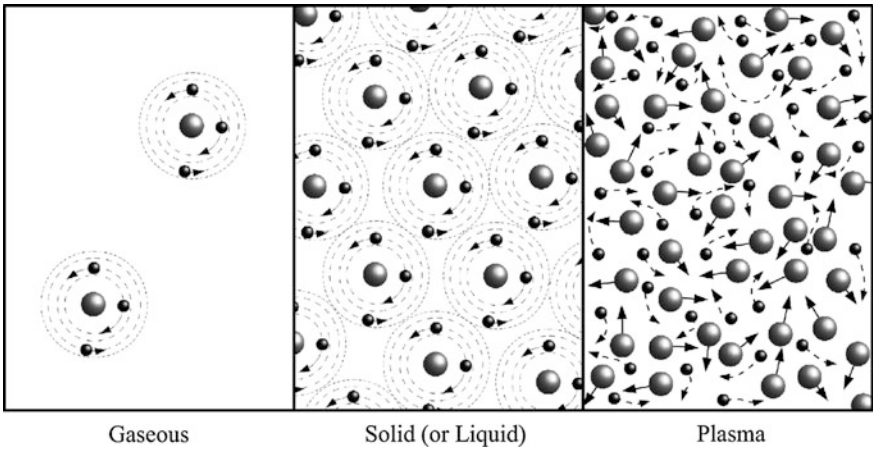


Fig. 5.6 States of matter The locations of atoms, *large dashed circles*; their component electrons, *small filled circles*; and central nucleus, *large filled circles*, for the gaseous (*left*), liquid or solid (*center*), and plasma (*right*) states of matter. In the gaseous state, the atoms are widely separated and free to move about. The atoms are practically touching one another in the solid and liquid states. At sufficiently high temperature and pressure, the atoms cease to exist and the plasma state is created. The atoms are torn into their constituents by frequent collisions at high temperatures. Plasma consists of bare nuclei and unattached electrons moving about in random directions within the former empty space of atoms. In the plasma state, matter regains the compressibility of the gaseous state and plasma behaves like a gas. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

The eV unit of energy is also used to describe x-rays, which were first produced by connecting a high-voltage power supply across the ends of an evacuated glass tube. The x-rays are in the 1–100 keV range of energy, where 1 keV = 1,000 eV.

The conversion from the eV unit to the joule unit of energy is

$$1\,\text{eV} = 1.602176 \times 10^{-19}\,\text{J} \approx 1.602 \times 10^{-19}\,\text{J}, \tag{5.33}$$

which is numerically equal to the elementary charge, in coulombs, of the electron.

The amount of energy required to remove the least tightly bound electron from a neutral atom is called the first ionization potential, denoted by the Roman numeral I. The additional energy needed to remove the next least tightly bound electron is the second ionization potential. More generally, the nth ionization potential, or the nth ionization energy, is the energy required to strip off the nth electron after the first $n - 1$ electrons have been removed.

For the one and only electron of the hydrogen atom, the first ionization potential is 13.5984 eV. The first, second and third ionization potentials of atomic oxygen are 13.6181, 35.117 and 54.934 eV. The temperature, T, required to ionize hydrogen, ripping off its sole electron and leaving a proton behind, is obtained by dividing its first ionization potential by the thermal energy kT, for the Boltzmann constant $k = 1.38065 \times 10^{-23}\,\text{J K}^{-1}$, giving a temperature $T \approx 1.58 \times 10^5$ K.

At this temperature, or any higher one, hydrogen atoms will be completely torn apart, into their subatomic ingredients, the electrons and protons, which happens inside most stars.

5.5.2 Plasma Oscillations and the Plasma Frequency

If all the electrons in plasma were displaced by a small amount with respect to the ions, the force of electrical attraction between the electrons and ions would pull the electrons back, but when pushed the displacement can continue. The back and forth motion is a natural oscillating one, with electrons moving at the *plasma frequency*, designated ν_P, given by (Tonks and Langmuir 1929):

$$\nu_P = \left[\frac{e^2 N_e}{4\pi^2 \varepsilon_0 m_e}\right]^{1/2} = 8.98\, N_e^{1/2}\ \text{Hz}, \tag{5.34}$$

where the electron density N_e is in units of m^{-3}, the electron charge $e = 1.602 \times 10^{-19}$ C, the electron mass $m_e = 9.1094 \times 10^{-31}$ kg, the permittivity of free space is $\varepsilon_0 = 8.8542 \times 10^{-12}\ \text{F m}^{-1}$, and $\pi \approx 3.14159$.

The ionosphere reflects radio waves at the plasma frequency ν_P or at the plasma wavelength, denoted λ_P. Since the radio radiation travels at the speed of light, c, we have $\lambda_P \times \nu_P = c = 2.9979 \times 10^8\ \text{m s}^{-1}$.

Example: **The temperature, origin, and plasma frequency of the Earth's ionosphere**

The F layer of the ionosphere, located about 200 km above our heads, contains oxygen atoms that are missing two electrons and a number density of free electrons of $N_e = 10^{12}\ \text{m}^{-3}$. The temperature, T, required to create these ions can be estimated by equating the thermal energy $3kT/2$ to the third ionization potential for oxygen atoms of 54.934 eV, using the Boltzmann constant $k = 1.38065 \times 10^{-23}\ \text{J K}^{-1}$ and $1\ \text{eV} = 1.6022 \times 10^{-19}$ J. The result is $T = 4.229 \times 10^5$ K. If the photon energy $h\nu$ of the incident solar radiation is equal to this thermal energy, where the Planck constant $h = 6.626 \times 10^{-34}$ J s, then the wavelength of the radiation is $\lambda = c/\nu = 2hc/(3kT) \approx 2.3 \times 10^{-8}$ m, or an x-ray wavelength, so x-rays from the Sun can produce the ionization. The plasma wavelength λ_P corresponding to the plasma frequency ν_P in this layer of the ionosphere is $\lambda_P = c/\nu_P \approx 33$ m, where the speed of light $c = 2.9979 \times 10^8\ \text{m s}^{-1}$, and $\nu_P = 8.98\ N_e^{1/2} = 8.98 \times 10^6$ Hz. Radio transmissions from the Earth at this long wavelength are reflected back down to the ground, and cannot get through the ionosphere. They can nevertheless be reflected at an angle, enabling long-distance radio communications.

We can infer the height and electron density of a given layer in the ionosphere by sending radio signals up into the atmosphere at successively longer wavelengths or shorter frequencies. The ionosphere will not mirror radio waves, and send a signal back, unless their wavelength is longer than the plasma wavelength or their frequency is less than the plasma frequency. This provides a measure of the electron density in the ionosphere, the shorter the reflection wavelength or the higher the reflection frequency, the larger the electron density. Radio waves with wavelengths that are less than the plasma wavelength can pass right through the ionosphere, because, roughly speaking, they are short enough to pass among the electrons. These shorter wavelengths are used in communications with satellites or other spacecraft in outer space beyond the ionosphere. The longer wavelengths are used in radio communications around the Earth through reflection off the ionosphere.

Example: ***Sputnik***
The F layer in the ionosphere has an electron density of about $N_e \approx 10^{12}\ \mathrm{m}^{-3}$; therefore, it has a plasma frequency of $\nu_p \approx 8.98 \times 10^6$ Hz ≈ 8.98 MHz with a corresponding wavelength of $\lambda_P = c/\nu_\mathrm{P} \approx 33$ m. Radio or microwave signals with wavelength shorter than this value can see through the ionosphere and communicate with satellites in and above it. At longer wavelengths, radio signals sent up into the ionosphere are reflected back down to the ground and this is how the ionosphere's electron densities are measured. The orbit of the first artificial satellite, *Sputnik*, which was launched on October 4, 1957, had a semi-major axis of 6,955 km, placing it about 584 km above the mean radius of the Earth; the mean radius is 6,371 km. *Sputnik* therefore orbited within the outer ionosphere; the atmospheric friction led to rapid orbital decay and the demise of the satellite in 3 months. Amateur radio operators monitored the beep of its radio signals throughout the world. At a signal frequencies of 20.0 and 40.0 MHz, or wavelengths of about 15 and 7.5 m, the radio signals from *Sputnik* were just short enough in wavelength to pass through the ionosphere.

5.5.3 Atoms are Torn Apart into Plasma Within the Sun

Whole atoms are only found in the outer visible layers of the Sun, where the temperature is a relatively cool 5,780 K. Raise the temperature by just a factor of three, to about 17,000 K, which happens just beneath the solar disk we see with our eyes, and the Sun's hydrogen atoms are stripped bare, losing their identity.

The hot atoms move rapidly here and there, colliding with each other at high speeds, and the violent force of these collisions is enough to fragment the atoms into their subatomic constituents. And since the Sun is mostly hydrogen, its

interior consists mainly of protons, the nuclei of hydrogen atoms, and free electrons that have been torn off the atoms by innumerable collisions and set free to move throughout the Sun.

What is left is plasma, a seething mass of electrically charged particles, the electrons and protons. The electrical charge of the protons balances and cancels that of the electrons, which have been removed from atoms to also release the protons, so the plasma has no net electric charge. The Sun is just one huge mass of incandescent plasma, compressed on the inside and more tenuous further out.

Plasma can be packed more tightly than complete atoms. This is because the electrons in an atom are located at relatively remote distances from the atomic nuclei, so atoms are largely empty space and once the electrons are removed the protons can be compressed together more than atoms can.

5.6 Sound Waves and Magnetic Waves

5.6.1 Sound Waves

Sound is transmitted in waves that are produced by perturbations in an otherwise undisturbed gas or liquid. These waves can be described as a propagating change in the mass density. For a fluid medium, which can be either a gas or a liquid, we assume an initial equilibrium in which the fluid is at rest, with initial velocity $V_0 = 0$ and a constant mass density ρ_0 and pressure P_0, where the subscript 0 denotes the initial undisturbed condition. We then assume a perturbation ρ_1 in the mass density ρ that becomes $\rho = \rho_0 + \rho_1$; the perturbation velocity is denoted as V_1, and the perturbation pressure designated P_1 with a subscript 1 for the perturbed condition. The equations of hydrodynamics have a plane wave solution for the perturbed density in the x direction given by:

$$\rho_1 \propto \exp\left[i\left(\frac{2\pi x}{\lambda} - \omega t\right)\right] \tag{5.35}$$

where the frequency, ω, is related to the wavelength, λ, by:

$$\omega^2 = \left(\frac{2\pi}{\lambda}\right)^2 \left(\frac{\partial P}{\partial \rho}\right). \tag{5.36}$$

The pressure, P_1, and velocity, V_1, also satisfy the wave equation. The French mathematician and astronomer Pierre-Simon Laplace (1749–1827) used the ideal gas law to describe the pressure, by $P = NkT$ and $P/\rho = kT/\bar{m}$ to obtain the speed of sound, c_S (Laplace 1816):

$$c_S = \left(\frac{\omega\lambda}{2\pi}\right) = \left(\frac{\partial P}{\partial \rho}\right)^{\frac{1}{2}} = \left(\frac{\gamma P_o}{\rho_o}\right)^{\frac{1}{2}} = \left(\frac{\gamma k T_o}{\bar{m}}\right)^{\frac{1}{2}}, \tag{5.37}$$

where γ is the adiabatic index. For a monatomic gas, the γ is $5/3 = 1.667$ and for a diatomic gas $\gamma = 7/5 = 1.400$. The Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. The mean molecular mass is $\bar{m} = \mu \times u$ where the mean molecular weight is μ and the atomic mass unit u $= 1.66054 \times 10^{-27}$ kg. Isaac Newton (1642–1727) considered the speed of sound in an isothermal calculation in his *Principia* (*Book II*, *Proposition* 49), instead of an adiabatic one, obtaining essentially the same result with a $\gamma = 1.000$ for an isothermal perturbation.

The speed of sound in an ideal gas is proportional to the square root of the temperature, but it is nearly independent of pressure or mass density for a given gas. At a constant temperature, the ideal gas pressure has no effect on the speed of sound because the pressure and the density, which is also proportional to pressure, have equal but opposite effects on the speed of sound, and the two contributions cancel out exactly.

Our atmosphere consists mainly of diatomic molecules N_2 at 78 % and oxygen O_2 at 21 %. For diatomic molecules $\gamma = 1.400$, and the mean molecular weight $\mu = 2A$ for a diatomic molecule composed of atoms of atomic mass number A. Nitrogen and oxygen have A = 14 and A = 16, respectively; therefore, the mean molecular weight of our atmosphere is $2 \times 14 \times 0.78 + 2 \times 16 \times 0.21 = 28.56$. Substituting these numbers with the other known constants into the equation we find that sound moves though the air at a speed of about 20 $T^{1/2}$ m s^{-1}, or to be precise:

$$c_{air} = 20.0457[T]^{1/2}\,\mathrm{m\,s^{-1}}, \tag{5.38}$$

where T is temperature on the kelvin scale, and $T = 273.15 + T_c$ if you are using a temperature T_C in °C. At sea level, the temperature is 288 K or 15 °C and the sound speed is about 340 m s^{-1}. The speed of sound decreases with altitude, due to lower temperatures found there, but even at the cruising altitudes of most aircraft the temperature is less than about 216 K corresponding to a speed of sound of less than 294 km s^{-1}.

The speed of motion divided by the speed of sound is called the Mach number, and anything that moves at a speed greater than Mach 1 is said to be traveling at supersonic speed. Most modern fighter aircraft are supersonic. Such aircraft have broken the sound barrier and can produce a sonic boom.

Example: **Sound waves in the Earth's atmosphere**
Assuming that the Earth's atmosphere is mainly composed of diatomic nitrogen molecules, with an adiabatic index of $\gamma = 7/5$ and a molecular mass of 2×14 u, where the atomic mass unit u $= 1.660539 \times 10^{-27}$ kg, the sound speed at ground level, where the temperature is $T = 288$ K, is about $c_s = [\gamma kT/(28\mathrm{u})]^{1/2} \approx 346$ m s^{-1}, where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. The first shout from a drowning man, located 1 km out in the ocean, would be heard at the beach in just 2.9 s.

We detect sound by small changes in the sound pressure against our vibrating eardrum, above and below the normal atmospheric pressure. The threshold of hearing for most individuals is a sound pressure of 2×10^{-5} Pa. Sound pressure is inversely proportional to distance from the source of the sound, unlike sound intensity that falls off with the inverse square of the distance.

At sea level, sound waves move about 4.3 times faster in water than in air. The larger mass density of water slows the sound waves in water relative to air, and this nearly makes up for the compressibility differences of the two media. Submarines use pulses of sound waves to detect the direction and distance of ships or the depth of the ocean floor. The acronym SONAR is used for such SOund Navigation And Ranging.

Sound waves are generated by turbulence (Lighthill 1952, 1954; Proudman 1952), and it is such turbulent motions that give rise to the roar of a jet airplane engine. The convective rise and fall of gas in the outer layers of the Sun also create sound waves (Biermann 1948; Schwarzschild 1948; Schatzman 1949) that are produced by the turbulent motion (Goldreich and Kumar 1990). Most of the solar sound waves are trapped inside the Sun, and they are used to investigate its internal properties (Sect. 8.5).

Within the Sun we can assume, to a first approximation, that it consists of a fully ionized monatomic gas with an adiabatic index of $\gamma = 5/3 = 1.667$ and the mass is dominated by the protons, with a mass $m_P = 1.67 \times 10^{-27}$ kg. Similar assumptions apply to the hot expanding solar atmosphere, resulting in the supersonic solar wind.

Example: **Sound speed in the solar wind**
A perpetual wind of protons and electrons is blowing out from the Sun in all directions through interplanetary space. At its origin near the Sun, the temperature is 10^6 K, and the highly conducting wind stays nearly that hot all the way to the Earth and beyond. Using this temperature with $\gamma = 1.667$, a proton mass of $m_P = 1.6726 \times 10^{-27}$ kg, and the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, we find that the sound speed in the solar wind is $c_S \approx (\gamma kT/m_P)^{1/2} \approx 10^5$ m s^{-1} or 100 km s^{-1}. The solar wind has a fast component moving at a speed of about 750 km s^{-1} and a slow one moving at about half that speed, so the solar wind is always supersonic, with a speed exceeding the speed of sound.

5.6.2 Magnetic Waves

In addition to ponderous material particles, like atoms, molecules, electrons, and ions, there are also magnetic fields that permeate the universe. The trajectories of charged particles, the electrons and the ions, are guided by these magnetic fields,

which act as a wall to them. The charges can spiral around the magnetic fields but cannot cross them.

The theory that deals with the interaction of a hot gas, or plasma, and a magnetic field is called magnetohydrodynamics, or MHD for short. As the ponderous name suggests, the equations are a combination of those of electromagnetism and fluid mechanics. The Swedish electrical engineer Hannes Alfvén (1908–1995) pioneered the study of MHD, receiving the 1970 Nobel Prize in Physics for this work.

Alfvén proposed the possible existence of oscillations produced by magnetic tensions, which are now known as *Alfvén waves*. These waves propagate in the direction of the magnetic field with the Alfvén velocity V_A given by (Alfvén 1942a, b):

$$V_A = \frac{B}{\sqrt{\mu_0 \rho}}, \tag{5.39}$$

where B is the magnetic field strength in tesla, ρ is the mass density in units of kg m^{-3}, and the magnetic permeability $\mu_0 = 1.2566 \times 10^{-6}$ N A^{-2}.

Neglecting the contribution of electrons to the mass density and assuming that there is a single ion species we obtain:

$$V_A = \frac{B}{\sqrt{\mu_0 N_i m_i}}, \tag{5.40}$$

where N_i is the number density of the ions in m^{-3} with mass m_i in kg, for protons $m_i = 1.6726 \times 10^{-27}$ kg. The ion mass density provides the inertia for the oscillations and the magnetic field provides the restoring force. The ion motion and the magnetic field perturbations are in the same direction, both transverse to the direction of propagation.

The kinetic energy density of the ions is given by:

$$\frac{1}{2} m_i V_A^2 N_i = \frac{B^2}{2\mu_0}, \tag{5.41}$$

which is equal to the magnetic field energy density $B^2/(2\mu_0)$.

Cosmic magnetic fields are always being jostled, twisted and stirred around, and tension acts to resist the motions and pull the disturbed magnetism back. This generates Alfvén waves that propagate along magnetic fields, somewhat like a vibrating string. These waves do not form shocks, and once generated they can propagate large distances, directing their energy along open magnetic fields. Alfvén suggested that these waves could contribute to the heating of the outer solar atmosphere (Alfvén 1947).

Instruments aboard *Mariner 5* detected magnetic fluctuations attributed to large-amplitude Alfvén waves in the interplanetary medium during the spacecraft's voyage to Venus in 1967 (Belcher et al. 1969). The pressure of these waves may push the solar wind to a higher speed than it would otherwise have (Cranmer

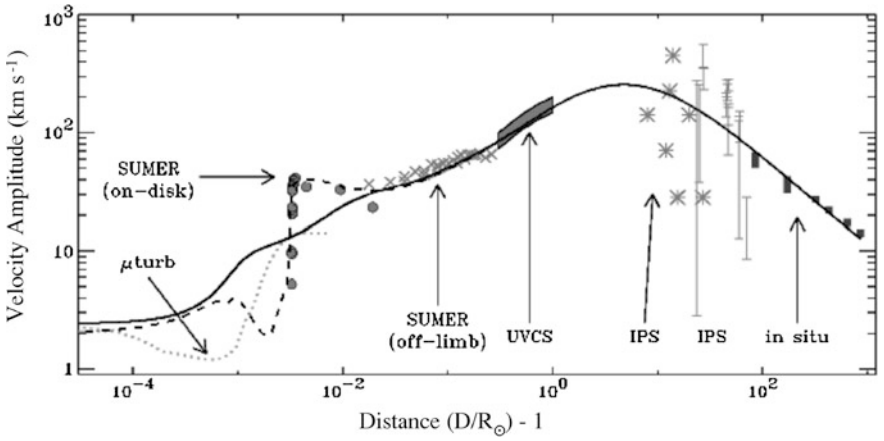


Fig. 5.7 Alfvén waves The velocity amplitudes, or speeds, of Alfvén waves, expressed as transverse velocities of the oscillating magnetic field lines, versus distance, D, above the visible disk of the Sun or photosphere, given in units of the solar radius $R_\odot$. The *solid curve* fits the data observed from spacecraft whose instruments are specified by their acronym. The left-most two sets of data (*dotted* and *dashed curves*) represent radial motions and may not correspond directly to the transversely oscillating Alfvén waves (adapted from a figure provided by Steven R. Cranmer)

and van Ballegooijen 2005). The magnetic waves have been measured for more than 40 years throughout the plane of our solar system, both near to and far from the Sun (Fig. 5.7). Magnetometers aboard the *Ulysses* spacecraft have detected the effects of Alfvén waves above the Sun's polar regions, and observations from the *Hinode* spacecraft have found their signatures near the visible solar disk, with perhaps enough energy to power the Sun's winds (De Pontieu et al. 2007).

Example: **Alfvén waves in the interplanetary medium**
Spacecraft measure an interplanetary magnetic field strength just outside the Earth of $B = 2.5 \times 10^{-9}$ T. The observed proton density in the solar wind at the Earth's orbit is $N_P = 5 \times 10^6$ m^{-3} for protons of mass $m_P = 1.67 \times 10^{-27}$ kg. The associated Alfvén velocity is $V_A = B/(\mu_0 N_P m_P)^{1/2} = 2.44 \times 10^4$ m s^{-1} = 24 km s^{-1}, with $\mu_0 = 1.2566 \times 10^{-6}$ N A^{-2}. The magnetic fields are tied to the Sun at one end, and stretch out into interplanetary space at the other, with a strength that is inversely proportional to the distance. The proton density is proportional to the inverse square of the distance, but the Alfvén velocity goes as the inverse square root of this density. So the two distance variations cancel, and we expect comparable Alfvén velocities throughout interplanetary space, which have in fact been observed from 10 to 100 km s^{-1} (Fig. 5.7).

Chapter 6
Detecting Atoms in Stars

6.1 What is the Sun Made Out Of?

When sunlight is spread out into its different colors or wavelengths, it is cut by several dark gaps. They were first noticed by the English astronomer William Hyde Wollaston (1766–1828) in 1802 (Wollaston 1802), and then investigated in greater detail by German astronomer Joseph von Fraunhofer (1787–1826). Fraunhofer had detected and catalogued more than 300 gaps, assigning Roman letters to the most prominent (Fraunhofer 1817).

By directing the incoming sunlight through a slit and then dispersing it with a prism, Fraunhofer was able to overcome the blurring of colors from different parts of the Sun's disk, discovering numerous dark features in this spectral display. When coarse wavelength resolution is used, adjacent bright emission obscures the dark places, which are no longer found.

The dark gaps of missing colors found in a display of the Sun's radiation intensity as a function of wavelength, or in its spectrum, are now called *lines* because they each look like a line in the spectral display. They are designated further as *absorption lines* because they are produced when atoms in a cool, tenuous gas absorb the radiation of hot, dense underlying material. They are also known as Fraunhofer absorption lines, in recognition of his work. Such lines also can appear in emission when a gas is heated, and they are known as *emission lines*.

Both absorption and emission lines (Fig. 6.1) identify the ingredients of the cosmos. For stars, astronomers mainly use absorption lines to determine the composition of their relatively cool, outer atmospheres. In contrast, emission lines reveal the ingredients of the hotter, rarefied emission nebulae.

The Sun is so bright that its light can be spread out into small wavelength intervals with enough intensity to be detected, thereby displaying numerous dark absorption lines (Fig. 6.2). One instrument used to make and record such a spectrum is called a spectroheliograph, a composite word consisting of *spectro* for "spectrum," *helio* for the "Sun," and *graph* for "record" (Fig. 6.3). It uses the

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_6, © Springer-Verlag Berlin Heidelberg 2013

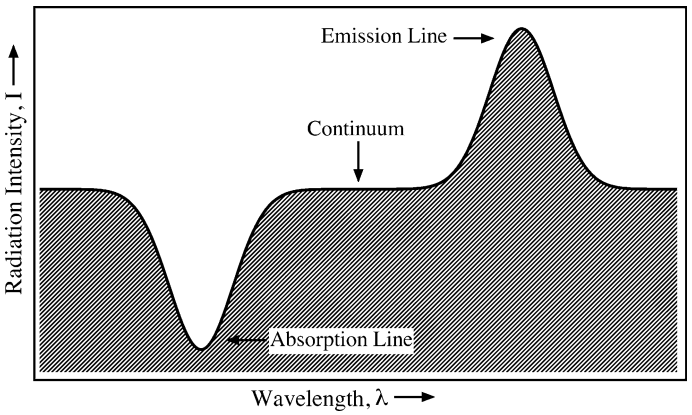


Fig. 6.1 Absorption and emission lines The spectrum of a star or other cosmic object displays the intensity of its radiation as a function of wavelength, denoted by the Greek symbol lambda, λ. The object's continuum radiation can be detected at all wavelengths. Atoms can produce absorption at a specific wavelength. This feature is called an *absorption line* because it looks like a line in the spectrum. When atoms are excited at high temperatures, they can radiate an *emission line*. The motion of the absorbing or emitting atoms broadens these lines (see Fig. 6.6). The line wavelength indicates which atom is responsible for the absorption or emission, and the intensity of a stellar line is related to both the number of atoms and the physical conditions in the star's atmosphere. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

grooves of a diffraction grating to reflect sunlight into different locations according to color or wavelength, similar to what a rainbow and compact disk do.

The Sun's absorption lines provided the first clues to the composition of the stars. In the mid-19th century, the German physicist Gustav Kirchhoff (1824–1887) and his chemist colleague Robert Bunsen (1811–1899) showed that every chemical element when burned and vaporized into a gas emits brightly colored lines. And the unique wavelengths of these lines coincide with those of the dark absorption lines in the Sun's spectrum.

When Kirchhoff and Bunsen vaporized an individual element in a flame, the hot vapor produced a distinctive pattern of sharply defined, bright lines. Moreover, when the light produced by a hot radiating object, such as a tungsten lamp, was passed through the cooler vaporized gas, dark lines were produced at exactly the same locations. Kirchhoff generalized this into a law stating that the powers of emission and absorption of a body at any particular wavelength are the same at any given temperature. He also concluded that the visible solar disk was hot and incandescent, producing a continuum spectrum (the sort without lines), which became crossed by the dark Fraunhofer lines when passing through cooler overlying gas.

By comparing the Sun's absorption lines with the emission lines of elements vaporized in the laboratory, Kirchhoff identified in the solar atmosphere several elements known on the Earth. The lines designated by Fraunhofer with the letters H and K were associated with calcium, and iron was assigned the letter E.

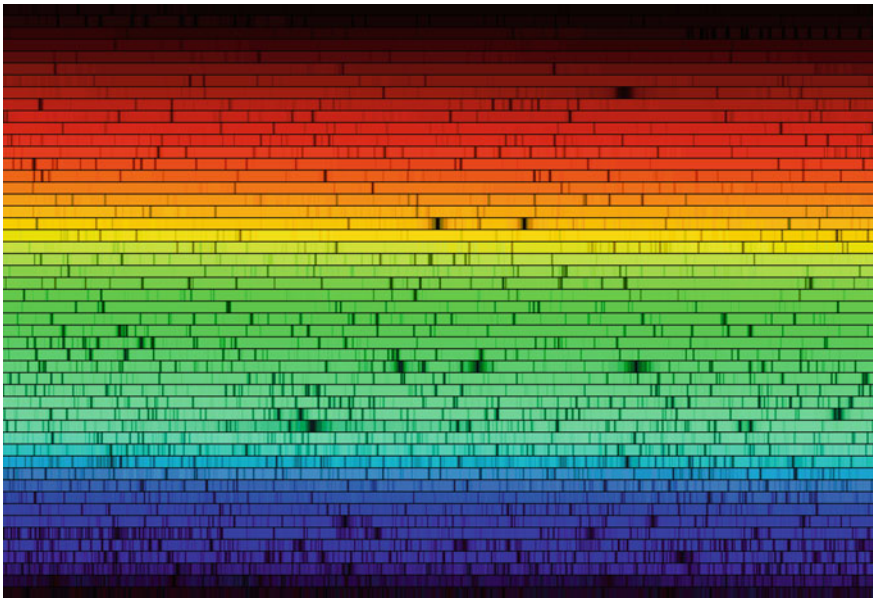


Fig. 6.2 Solar spectrum The visible portion of the Sun's radiation has been dispersed into its spectral components, displaying radiation intensity as a function of wavelength. When we pass from long wavelengths to shorter ones (*left–right*, *top−bottom*), the spectrum ranges from *red* thorough *orange*, *yellow*, *green*, *blue* and *violet*. Dark gaps in the spectrum, called Fraunhofer absorption lines, are due to absorption by atoms 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. This high-resolution version of the spectrum of our Sun was created from a digital atlas observed with the Fourier Transform Spectrometer at the McMahon-Pierce Solar Facility at the National Solar Observatory on Kitt Peak, near Tucson Arizona. (Courtesy N. A. Sharp, NOAO/NSO/Kitt Peak FTS/AURA/NSF.)

The letter b was attributed to magnesium, and the close pair of dark lines in the yellow, specified by the letter D, was attributed to sodium; they produce the distinctive yellow color of sodium vapor streetlights used at the time. This suggested that the Sun – and presumably all stars – are composed of terrestrial elements (Kirchhoff and Bunsen 1860, 1861; Kirchhoff 1861a, b).

The Swedish spectroscopist Anders Jonas Ångström (1814–1874) subsequently identified hydrogen in the solar spectrum; it is associated with Fraunhofer's letters C and F. Ångström published a comprehensive atlas of more than 1,000 absorption lines in the Sun's spectrum, identifying them with hydrogen, sodium, calcium, barium, strontium, magnesium, copper, iron, chromium, nickel, cobalt, zinc, and gold (Ångström 1868). His scale of wavelengths for measuring the spectral lines is still used by some astronomers; it is now named the Ångström, abbreviated Å, with $1\ \text{Å} = 10^{-10}\ \text{m} = 0.1\ \text{nm}$.

Unfortunately, the Fraunhofer lines designated A and B are not related to the composition of the Sun. They only appear in spectra gathered beneath the Earth's

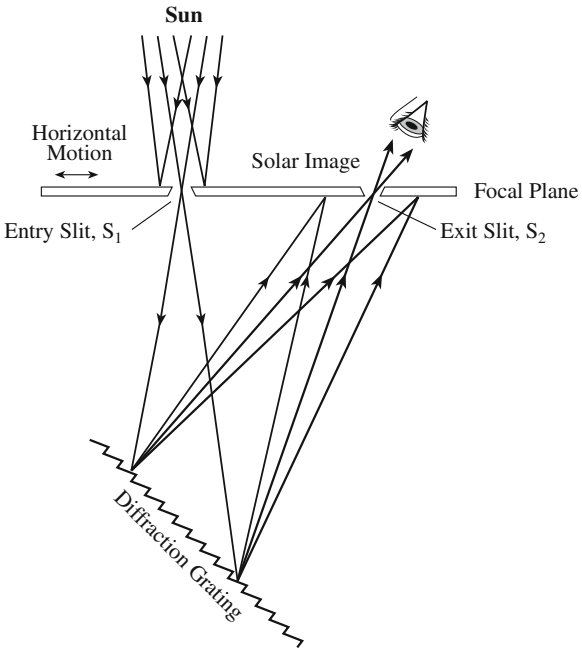


Fig. 6.3 Spectroheliograph A small section of the Sun's image at the focal plane of a telescope is selected with a narrow entry slit, S_1, and this light passes to a diffraction grating to produce a spectrum. A second slit, S_2, at the focal plane selects a specific wavelength from the spectrum. If the plate containing the two slits is moved horizontally, then the entrance slit passes adjacent strips of the solar image. The light leaving the moving exit slit then builds up an image of the Sun at a specific wavelength

atmosphere, and are caused by oxygen in the terrestrial atmosphere. The oxygen molecules absorb sunlight at the wavelengths of the A and B Fraunhofer lines, creating the dark lines that are superposed on the Sun's spectrum.

Still, there are some very strong absorption lines that are due to the Sun, and they extract large amounts of energy from sunlight. They are produced by hydrogen, sodium, magnesium, calcium, and iron (Table 6.1), but iron accounts for more lines than any other element. Because abundant heavy iron accounts for the Earth's high average mass density and because most of the other solar lines corresponded to elements known on the Earth, it was initially thought that the Sun is made of the same material as the Earth; but this turned out to be only partly true. Many of the visible spectral lines are associated with hydrogen, a relatively rare element on Earth (Abundant hydrogen is found bound up in the Earth's water, but hydrogen gas escapes from the Earth and is not found in noticeable quantities in its atmosphere.). Hydrogen is nevertheless the most abundant element in the universe.

The Earth is primarily made of heavy elements that are relatively uncommon in the Sun and the rest of the universe. Hydrogen is about 1 million times more

Table 6.1 Prominent absorption lines and elements detected in sunlight[a]

Wavelength (nm)	Fraunhofer letter	Element symbol and name
393.368	K	Ionized calcium, Ca II
396.849	H	Ionized calcium, Ca II
410.175	h	Hydrogen, H_{δ}, Balmer delta transition
422.674	g	Calcium, Ca I
431.0	G	CH molecule
434.048		Hydrogen, H_{γ}, Balmer gamma transition
438.356	d	Iron, Fe I
486.134	F	Hydrogen, H_{β}, Balmer beta transition
516.733	b_4	Magnesium, Mg I
517.270	b_2	Magnesium, Mg I
518.362	b_1	Magnesium, Mg I
526.955	E	Iron, Fe I
588.997[b]	D_2	Sodium, Na I
589.594	D_1	Sodium, Na I
656.281	C	Hydrogen, Hα, Balmer alpha transition
686.719	B	Molecular oxygen, O_2, in the Earth's atmosphere
759.370	A	Molecular oxygen, O_2, in the Earth's atmosphere

[a] The photosphere is the visible solar disk. The wavelengths are in nanometer units, where 1 nanometer = 1 nm = 10^{-9} m. Astronomers have often used the Ångström unit of wavelength, where 1 Ångström = 1 Å = 0.1 nm. Joseph von Fraunhofer used letters to designate the spectral lines before they were chemically identified, and the subscripts denote components that were not resolved by Fraunhofer. A Roman numeral I after an element symbol denotes an electrically neutral, or unionized, atom, with no electrons missing, whereas the Roman numeral II denotes a singly ionized atom with one electron missing. The lines A and B are produced by molecular oxygen in the terrestrial atmosphere

[b] Fraunhofer's D line includes the two sodium lines, designated D_1 and D_2, and the helium line at 587.6 nm, designated D_3

abundant than iron in the Sun, but iron is a main constituent of the Earth, which cannot even retain hydrogen gas in its atmosphere for any significant length of time. Asplund et al. (2009) have reviewed observations of the chemical composition of the Sun.

Helium, the second-most abundant element in the Sun, is so rare on the Earth that it was first discovered in the Sun. The French astronomer Pierre Jules César Janssen (1824–1927) observed an unidentified yellow emission line, with a wavelength of 587.49 nm, in the solar spectrum during the solar eclipse on August 18, 1868, which he observed from India (Janssen 1868). The emission originated in the chromosphere, a thin, slightly hotter layer of gas that lies just above the visible solar disk, or photosphere, which became visible when the Moon blocked the bright glare of the photosphere.

On October 20, 1868, the British astronomer Sir Joseph Norman Lockyer (1836–1920) found the same yellow line in the solar spectrum, but he didn't need a solar eclipse (Lockyer 1869, 1887). It was probably not until the following year that Lockyer became convinced that the yellow line he saw could not be identified with any known element on Earth, and named the element "helium" after the

Table 6.2 The twenty most abundant elements in the Sun

Element	Symbol	Atomic number, Z	Abundance[a] (logarithmic)	Discovery on Earth
Hydrogen	H	1	12.00	1766
Helium	He	2	[10.93 ± 0.01]	1895[b]
Carbon	C	6	8.43 ± 0.05	(ancient)
Nitrogen	N	7	7.83 ± 0.05	1772
Oxygen	O	8	8.69 ± 0.05	1774
Neon	Ne	10	[7.93 ± 0.10]	1898
Sodium	Na	11	6.24 ± 0.04	1807
Magnesium	Mg	12	7.60 ± 0.04	1755
Aluminum	Al	13	6.45 ± 0.03	1827
Silicon	Si	14	7.51 ± 0.03	1823
Phosphorus	P	15	5.41 ± 0.03	1669
Sulfur	S	16	7.12 ± 0.03	(ancient)
Chlorine	Cl	17	5.50 ± 0.30	1774
Argon	Ar	18	[6.40 ± 0.13]	1894
Potassium	K	19	5.03 ± 0.09	1807
Calcium	Ca	20	6.34 ± 0.04	1808
Chromium	Cr	24	5.64 ± 0.04	1797
Manganese	Mn	25	5.43 ± 0.04	1774
Iron	Fe	26	7.50 ± 0.04	(ancient)
Nickel	Ni	28	6.22 ± 0.04	1751

[a] Logarithm of the abundance in the solar photosphere, normalized to hydrogen = 12.00, or an abundance of 1.00×10^{12}. Indirect solar estimates are marked with []. The data are from Asplund et al. (2009)

[b] Helium was discovered on the Sun in 1868, but it was not found on Earth until 1895

Greek Sun god, *Helios*, who daily traveled across the sky in a chariot of fire drawn by four swift horses.

Helium was not found in the Earth until 27 years after its discovery in the Sun. Then the Scottish chemist Sir William Ramsay (1852–1916) discovered its spectral signature in a gaseous emission given off by the uranium mineral, cleveite, when it was heated (Ramsay 1895). Ramsay received the Nobel Prize in Chemistry in 1904 for his discovery of noble, or inert, gaseous elements in the air; they include helium, neon, argon, krypton, xenon and radon. These so-called noble gases do not combine with most other chemical elements, which is one reason it took so long to discover helium on Earth.

Today, helium is used on the Earth in a variety of ways, including inflating party balloons and, in its liquid state, to keep sensitive electronic equipment cold. Although plentiful in the Sun, helium is almost nonexistent on the Earth. It is so terrestrially rare that we are in danger of running out of helium during this century. There have been reports that Japanese scientists plan to mine helium from the Moon's surface, where it has been implanted by winds from the Sun.

Altogether, 92.1 % of the atoms of the Sun are hydrogen atoms, 7.8 % are helium atoms, and all of the heavier elements comprise only 0.1 %. In contrast, the main ingredients of the Earth are the heavier elements such as silicon and iron,

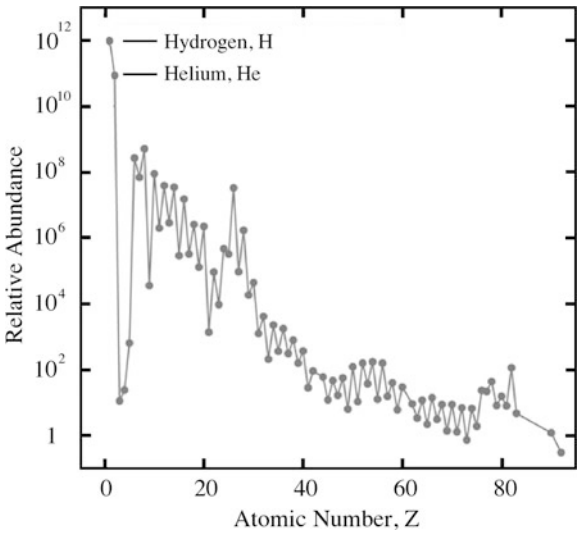


Fig. 6.4 Elemental abundance in solar photosphere The relative abundance of the elements in the Sun's visible disk, the photosphere, plotted as a function of atomic number. The atomic number, denoted by Z, is the number of protons in an atom's nucleus, or roughly half the atomic weight. Heavy elements, with high atomic numbers, are less abundant than light ones, with low atomic numbers; the most abundant element in the Sun is the lightest element, hydrogen. Helium is the second most abundant solar element. The abundance data are plotted in a logarithmic scale normalized to 1 million million, or 1.0×10^{12}, for hydrogen. (Adapted from Asplund et al. 2009.)

which explains the Earth's high mass density – about four times that of the Sun, which is only about as dense as water.

By mass, hydrogen accounts for 71.54 % of the Sun; the helium amounts to 27.03 % by mass, because the helium atom is about four times more massive than the hydrogen atom. All of the heavier solar elements amount to only 1.42 % by mass.

The abundance of the 20 most abundant elements in the Sun are listed in Table 6.2, normalized to a hydrogen abundance of 1 million million, or 10^{12}. There is a systematic decrease in the abundance of solar elements with increasing atomic number (Fig. 6.4), but with a noticeable gap of unexpectedly low abundance for the light elements between helium and carbon.

6.2 Quantization of Atomic Systems

Most of the mass of an atom is concentrated in its relatively small nucleus, which is surrounded by electrons (Sect. 5.1). The nucleus has a positive charge due to the protons in it and is about 100,000 times smaller than the atom. The negatively charged electrons keep the atom distended, enlarging its shape. As a result, an atom is mostly empty space.

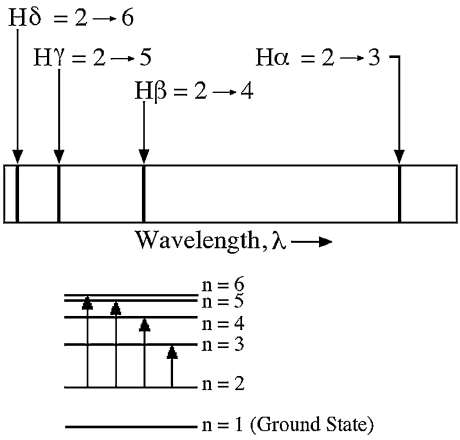


Fig. 6.5 Balmer lines of hydrogen The spectrum of the Sun's optically visible light exhibits four strong absorption lines that are attributed to hydrogen, whose line wavelengths are spaced closer together at shorter wavelengths (*top*). These lines are designated Hα at a *red* wavelength of 656.3 nm, Hβ at a wavelength of 486.1 nm, Hγ at the *blue* 434.1 nm and Hδ at the *violet* 410.2 nm, where 1 nanometer = 1 nm = 10^{-9} m. These spectral features originate when an electron in a hydrogen atom moves from a low to a high electron orbit, the orbital energy of which is a function of the integer n (*bottom*). They were named *Balmer lines* after the Swiss mathematics teacher Johann Balmer (1825–1898), who first derived an equation that describes their wavelengths in terms of integers. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

According to Rutherford's model of the atom, the electrons revolve around the central nucleus, somewhat like the planets that endlessly whirl around the Sun. Unlike a planet, an electron is electrically charged, and a revolving charge emits electromagnetic radiation. That is how radio signals are broadcast, by moving electrons through wires to generate radiation.

This means that something was wrong with Rutherford's model. An electron revolving in an atom-sized orbit radiates light waves, and as a result of this emission, it will lose its kinetic energy of motion. As a result, an electron cannot be perpetually moving around the nucleus. An atomic electron should lose its orbital motion and spiral into the atom's nucleus in less than 1 s. So, the idea of electron orbits did not seem to work.

An unexpected feature of the Sun's hydrogen lines, which are produced by an orbital change of the hydrogen atom's single electron, provided a clue to this enigma. That is, the wavelengths of spectral lines from hydrogen indicate that electrons have to follow certain rules if they want to belong to an atom. Not just any behavior is allowed and only certain orbits are permitted.

Adjacent hydrogen lines in the spectrum of the Sun or any other cosmic object systematically crowd together at shorter wavelengths (Fig. 6.5). The Swiss mathematics teacher Johann Balmer (1825–1898) found a simple equation that describes their regular spacing. The four lines in the Sun's visible hydrogen

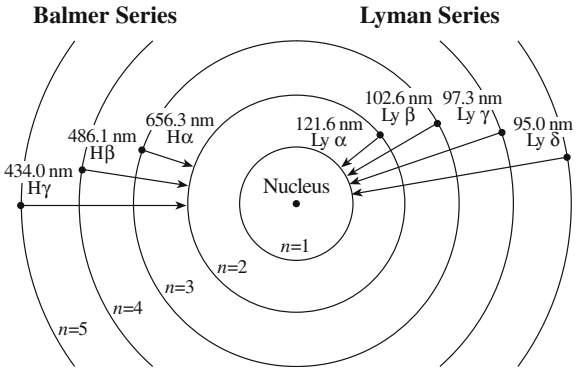


Fig. 6.6 Bohr atom In this model, proposed in 1913 by the Danish physicist Niels Bohr (1885–1962), a hydrogen atom's one electron revolves around the hydrogen nucleus, a single proton, in well-defined orbits described by the integer $n = 1, 2, 3, 4, 5, \ldots$. An electron absorbs or emits radiation when it transitions between these allowed orbits. The electron can jump upward, to orbits with larger n, by absorption of a radiation 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 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 because the size of their radius increases with the square of the integer n

spectrum have wavelengths of 410, 434, 486, and 656 nm, and they are still known as the Balmer lines. Balmer's equation that specified their wavelengths, λ, is (Balmer 1885):

$$\lambda = K\left(\frac{m^2}{m^2 - n^2}\right) = K\left(\frac{m^2}{m^2 - 2^2}\right), \tag{6.1}$$

where the constant $K = 3.6456 \times 10^{-7}$ $m = 364.56$ nm, the integer $n = 2$ and the integers $m = 3, 4, 5, \ldots$. The most intense Balmer line is now known as the *Balmer hydrogen alpha line*, designated Hα, at a red wavelength at 656.3 nm, where 1 nanometer = 1 nm = 10^{-9} m = 10 Å = 10 Ångström.

A few years later, the Swedish physicist Johannes Rydberg (1854–1919) generalized the Balmer equation for all transitions of hydrogen, in the Rydberg formula (Rydberg 1890):

$$\frac{1}{\lambda} = R_\infty\left(\frac{1}{m^2} - \frac{1}{n^2}\right) \tag{6.2}$$

where m and n are integers, the Rydberg constant for hydrogen $R_\infty = 10{,}973{,}731.57\ \text{m}^{-1} = 1.097373157 \times 10^7\ \text{m}^{-1} = 4/K$ where the K is in Balmer's equation where $m = 2$, the $n = 3, 4, 5$ (the m and n have been interchanged from Balmer's formula.) Rydberg's formula was subsequently used to

predict spectral lines observed at the invisible infrared, ultraviolet and radio wavelengths.

The Danish physicist Niels Bohr (1885–1962) explained Balmer's equation by a model of the hydrogen atom, now known as the Bohr atom (Bohr 1913). In this model, a single electron in a hydrogen atom revolves about the nuclear proton in specific orbits with definite, quantized values of energy (Fig. 6.6). An electron only emits or absorbs radiation when jumping between those allowed orbits, each jump associated with a specific energy and a single wavelength. If an electron jumps from a low-energy to a high-energy orbit, it absorbs radiation at this wavelength; radiation is emitted at exactly the same wavelength when the electron jumps the opposite way. This unique wavelength is related to the difference between the two orbital energies. Bohr was awarded the 1912 Nobel Prize in Physics for these investigations of the structure of atoms and the radiation emanating from them.

Because only quantized orbits are allowed, spectral lines are produced only at specific wavelengths that characterize or identify an atom. An atom or molecule can absorb or emit a particular type of light only if it resonates to that light's energy. As it turns out, the resonating wavelengths or energies of each atom are unique.

A little more than a decade before Bohr introduced his model, Planck (1901) explained continuum thermal radiation, without lines, by supposing that the emission and absorption of light takes place only in the form of certain discrete portions, or quanta, of energy, now known as *photons* (See Sect. 2.4). He had quantized radiation, giving a photon an energy hv at frequency v. The constant h is the Planck constant, with the value $h = 6.626 \text{ x } 10^{-34}$ J s. The energy of each separate light quantum, or photon, is proportional to the frequency of light, or inversely proportional to its wavelength.

Bohr went one step farther and quantized the energy of motion of the electrons revolving in an atom. He proposed that the mechanical energy of any moving subatomic particle could take on only one of a certain set of discrete values, in an entirely new quantum mechanics. That is, he quantized the electron's angular momentum and energy, using the Planck constant, h. Because the quantum of an electron's orbital energy increases with the decreasing dimensions of the orbit, this suggested that quantum mechanics would become important only for very small, subatomic physical scales.

In the Bohr atom, the single electron of a hydrogen atom is said to orbit the atom's nuclear proton with an angular momentum, $m_e \text{V} r_n$, that is quantized, or (Bohr 1913):

$$m_e \text{V} r_n = hn/(2\pi), \tag{6.3}$$

where the mass of the electron $m_e = 9.109 \times 10^{-31}$ kg, the electron's orbital velocity is V, its orbital radius is r_n for the integer $n = 1, 2, 3 \ldots$, Planck's constant $h = 6.626 \times 10^{-34}$ J s, and $\pi = 3.14159$.

The radius, r_n, of the nth electron orbit is derived by equating the Coulomb force of attraction of the proton, of charge e, on the electron, also of charge e, to the centripetal force, or

$$\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r_n^2} = \frac{mV^2}{r_n} \tag{6.4}$$

to obtain

$$r_n = a_0 n^2 \tag{6.5}$$

where the radius, a_0, of the first Bohr orbit of hydrogen is:

$$a_0 = \frac{\varepsilon_0 h^2}{\pi m_e e^2} \approx 5.2918 \times 10^{-11}\ \text{m}, \tag{6.6}$$

with an electrical constant $\varepsilon_0 = 8.8542 \times 10^{-12}$ F m^{-1}, the Planck constant $h = 6.6261 \times 10^{-34}$ J s, the constant $\pi = 3.14159$, the electron mass $m_e = 9.1094 \times 10^{-31}$ kg, and the elementary charge $e = 1.6022 \times 10^{-19}$ C.

The electron orbital energy is also quantized, which can be seen by solving for its kinetic energy $m_e V^2/2$. That is, electrons can only occupy orbits with allowed orbital energy, E_n, given by:

$$E_n = \frac{1}{2}m_e V^2 = \frac{m_e e^4}{8\varepsilon_0^2 h^2 n^2} = \frac{hcR_\infty}{n^2} \tag{6.7}$$

where the Rydberg constant, R_∞, is:

$$R_\infty = \frac{m_e e^4}{8\varepsilon_0^2 h^3 c} \approx 10,973,731.568\ \text{m}^{-1} \approx 1.097 \times 10^7 \text{m}^{-1}, \tag{6.8}$$

and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

The constant $hcR_\infty = 13.\ 60569$ eV $= 1$ Ry is known as the Rydberg unit of energy, and is often used in atomic physics. The electron volt is equivalent to 1 eV $= 1.602176 \times 10^{-19}$ J.

The permitted orbital energy increases with decreasing integer, or quantum number, n, which corresponds to decreasing radius or size of the electron orbit. The closer an electron is to the nucleus, the greater is its allowed energy. Both the energy and radius of an electron's orbit vary as the inverse square of the quantum number.

The nth energy level of the hydrogen atom has energy $E_n = 13.6\ \text{eV}/n^2 = hcR_\infty/n^2$, where 13.6 eV is the ionization potential of hydrogen and 1 eV $= 1.602 \times 10^{-19}$ J. The temperature required to ionize hydrogen, removing its sole electron from its ground state is about 1.6×10^5 K, obtained from setting the temperature equal to 13.6 eV divided by the Boltzmann constant $k = 1.38 \times 10^{-23}$ J K^{-1}.

The energy difference between two energy levels specified by integers n and m is:

$$E_n - E_m = \left(\frac{1}{n^2} - \frac{1}{m^2}\right) \times 13.5609 \text{ eV}. \tag{6.9}$$

Radiation can be emitted or absorbed by a hydrogen atom when the electron changes between allowed orbits, with the frequency ν_{mn} of the $m - n$ transition given by

$$\nu_{mn} = \frac{1}{h}|E_m - E_n| = cR_\infty \left|\frac{1}{n^2} - \frac{1}{m^2}\right| \approx 2cR_\infty \frac{(m-n)}{n^3}, \tag{6.10}$$

where $||$ denotes the absolute value, and the wavelength λ_{mn} of the $m - n$ transition can be inferred from the frequency by $\lambda_{mn} = c/\nu_{mn}$ where $c = 2.9979 \times 10^8$ m s^{-1} is the speed of light.

The wavelength, λ_{mn}, of the radiation emitted or absorbed in a radiative transition between the two levels will be:

$$\lambda_{mn} = \frac{hc}{E_m - E_n} = \left(\frac{1}{n^2} - \frac{1}{m^2}\right)^{-1} \times 911.5 \text{ Å}, \tag{6.11}$$

where 1 Å $= 10^{-10}$ m $= 0.1$ nm is one Ångström.

The most intense spectral line for any given value of n is the transition with $m - n = 1$, known as the alpha, or α, transition, and the next most intense line is for the beta, or β, transition with $m - n = 2$, and so on for increasing $m - n$. The difference in orbital energies for the α transition is smaller than any other transition, and therefore the easiest to accomplish; more atoms will undergo this transition than other atoms, and the line is therefore the most intense. At higher $m - n$, the difference in orbital energies is greater and the transition is more energetic; but there are fewer atoms undergoing the transition and the intensity of the line is lower.

For any n, the frequency of the α transition is given by:

$$\nu_{mn} = \frac{2cR_\infty}{n^3}. \tag{6.12}$$

Example: Recombination lines at radio wavelengths

Electron transitions at high quantum numbers n have been detected from hydrogen atoms surrounding very hot stars. For the $m - n = 1$, or the α transition, radiation from the $n = 109$ transition has been observed. The frequency of this transition is $\nu_{mn} = 2cR_\infty/n^3 = 5.079 \times 10^9$ Hz $=$ 5,079 MHz, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1} and the Rydberg constant $R_\infty = 1.097 \times 10^7$ m^{-1}. The radius of the electron orbit is $r_n = a_0 n^2 \approx 6.29 \times 10^{-7}$ m, where the Bohr radius $a_0 = 5.2918 \times 10^{-11}$ m, so the orbit radius is much larger than the average radius of an atom, usually about 10^{-10} m.

The α transition detected at visible wavelengths is called the Balmer α transition, at a red wavelength of 6562.8 Å while the ultraviolet α transition is known as Lyman α at 1215.67 Å, where 1 Å $= 10^{-10}$ m.

The Lyman series, for $n = 1$, includes

Designation	m	λ_{mn}(Å)
$Ly\alpha$	2	1,216
$Ly\beta$	3	1,026
$Ly\gamma$	4	973
Lyman limit	∞	911.5

And the Balmer series, for $n = 2$, includes:

Designation	m	λ_{mn}(Å)
$H\alpha$	3	6,563
$H\beta$	4	4,861
$H\gamma$	5	4,341
Balmer limit	∞	3,646

The important hydrogen transitions between low n orbits have been given the last names of the persons who first observed them. They are called the Lyman transitions for $n = 1$ (Lyman 1906), the Balmer transitions for $n = 2$ (Balmer 1885), the Paschen transitions for $n = 3$ (Paschen 1908), the Brackett transitions for $n = 4$ (Brackett 1922), and the Pfund transitions for $n = 5$ (Pfund 1924).

The wavelengths of the relevant m to n transitions are given in Table 6.3 for m between 1 and 10, as well as for $m = \infty$, which is the limiting wavelength at which the lines merge together (Fig. 6.7). For wavelengths shorter than this limit,

Table 6.3 Wavelengths of the m to n transitions of hydrogen for $n = 1$ to $n = 5$ and $m = 2$ to $m = 10$[a]

Series (m)	Lyman ($n = 1$)	Balmer ($n = 2$)	Paschen ($n = 3$)	Brackett ($n = 4$)	Pfund ($n = 5$)
2	121.567				
3	102.572	656.280			
4	97.2537	486.132	1875.10		
5	94.9743	434.046	1281.81	4051.20	
6	93.7803	410.173	1093.81	2625.20 7	7457.8
7	93.0748	397.007	1004.94	2165.50	4652.5
8	92.6226	388.905	954.598	1944.56	3739.5
9	92.3150	383.538	922.902	1817.41	3296.1
10	92.0963	379.790	901.491	1736.21	3038.4
∞	91.15	364.6	820.36	1458.4	2278.8

[a] The wavelengths are given in nanometers where 1 nm $= 10^{-9}$ m

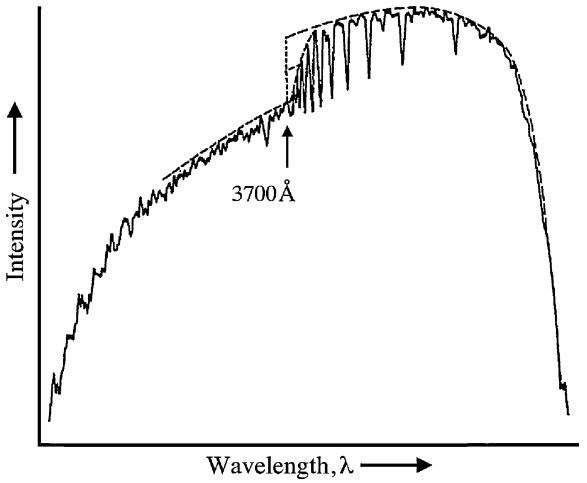


Fig. 6.7 Balmer limit As shown in this spectrum of a star, the lines of the Balmer series of hydrogen crowd together and merge at a limiting wavelength, the Balmer limit at 364.6 nm or approximately 3,700 Ångström. At wavelengths less than this limit, the radiation of hydrogen is emitted at a continuous range of wavelengths, known as the Balmer continuum, resulting from transitions between states with principal quantum number $n = 2$ and states in which the single electron is freed from the atom

the radiation is emitted at a continuous range of wavelengths, or as a continuum, resulting from transitions in which the electron leaves the atom.

For an atom of atomic number Z, which is the charge of the nucleus in units of the proton charge, the electron orbital energy, E_n, is given by

$$E_n = \frac{m_e e^4 Z^2}{8\varepsilon_0^2 h^2 n^2} = \frac{hcR_A}{n^2} \tag{6.13}$$

where n is an integer and the atomic Rydberg constant, R_A, is given by:

$$R_A = R_\infty \left(1 + \frac{m_e}{M_A}\right)^{-1} \approx R_\infty \left(1 - \frac{m_e}{M_A}\right) \tag{6.14}$$

where the electron mass $m_e = 5.4858 \times 10^{-4}$ u, the atomic mass M_A is often given in units of the atomic mass unit u $= 1.660539 \times 10^{-27}$ kg. The electron mass in atomic mass units is $m_e = 0.000548579867$ u. For hydrogen $Z = 1$ and for helium $Z = 2$. The frequency for the transition from an upper level m to a lower level n is given by

$$\nu_{mn} = cR_A Z^2 \left(\frac{1}{n^2} - \frac{1}{m^2}\right) \approx 2cR_A Z^2 \frac{(m-n)}{n^3} \tag{6.15}$$

Table 6.4 Atomic number Z, atomic mass M_A, and atomic Rydberg constant R_A for the most abundant atoms in the cosmos[a]

Atom	Atomic number, Z	Atomic mass, M_A (u)	Rydberg constant, R_A (10^7 m^{-1})
Hydrogen, H^1	1	1.007 825	1.096776
Helium, He^4	2	4.002 603	1.097223
Carbon, C^{12}	6	12.000 000	1.097323
Nitrogen, N^{14}	7	14.003 074	1.097330
Oxygen, O^{16}	8	15.994 914	1.097335
Neon, Ne^{20}	10	19.992 440	1.097343

[a] The atomic mass is given in atomic mass units u = 1.660539×10^{-27} kg

The atomic number and Rydberg constant for abundant atoms are given in Table 6.4.

6.3 Some Atoms are Excited Out of Their Lowest-Energy Ground State

Because atoms reside together in great numbers, we must use a statistical approach to determine their average properties. Their level of excitation depends on the temperature and the density, which influence how often the particles collide and become excited. The higher the temperature, the faster particles move, and the more frequent their collisions. When the particles are packed together in greater numbers per unit volume, with a greater density, collision frequency also increases.

The number of atoms in the lowest possible energy state, called the ground state, is greater than the number in any other state of energy, essentially because its easier to stay in the lowest energy state. Because it requires less energy, an atom acts as if it prefers the ground state. It is simply more difficult to enter or remain in an excited state of energy. Outside agencies are required to sustain an excited state of energy.

The allowed energy levels of the electron orbits within an atom can be compared to the rungs of a ladder. Electrons can climb the ladder's energy rungs when an atom either collides with other atoms or absorbs radiation. Once an electron is up on a rung, it can jump downward, releasing the energy it attained to get there. Since it is easier to stay on the ground and never climb the ladder, most atoms are usually in the ground state. Because each type of atom has a different type of ladder, with energy rungs located at different places, every element exhibits unique absorption or emission lines.

It takes more energy to excite the higher states, and the number of atoms that exist in a given energy level varies inversely and exponentially as a function of the energy. When collisions are the dominant process that influences the energy-level population, the ratio of the population of two energy levels of a given atom

depends on the temperature. At higher temperatures, more atoms are pumped up to the more energetic states.

Under conditions of local thermodynamic equilibrium, the ratio of the number of atoms at two different energies depends on their energy difference divided by the temperature. At a higher temperature, there are more atoms with higher energy; but there always are more atoms with the lowest possible energy, the ground state. The number ratio is known as the *Boltzmann distribution*, named after Ludwig Boltzmann (1894–1906) who derived it (Boltzmann 1872).

Individual atoms in a collection of atoms are always moving about, undergoing collisions, becoming excited and radiating that excitation away. The total energy emitted depends on the number of excited atoms, which varies with the temperature, the energy of the transition, and the transition probabilities, which are tabulated in spectroscopic data bases found on the Internet.

Under conditions of local thermodynamic equilibrium at a temperature T, the number densities, N_n, and N_m, of atoms in levels n and m with energies E_n and E_m are related by the Boltzmann distribution (Boltzmann 1872)

$$\frac{N_n}{N_m} = \frac{g_n}{g_m} \exp\left(-\frac{E_n - E_m}{kT}\right), \tag{6.16}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the g_n is the degeneracy of level n. The degeneracy of a level is the number of quantum states with the energy of that level. The quantity E_n is the energy of level n relative to a fundamental, unexcited ground state $n = 1$.

For the hydrogen atom, we have $g_n = 2n^2$ and

$$E_n = 13.6\left[1 - \frac{1}{n^2}\right] \text{eV}, \tag{6.17}$$

where n is the quantum number of the atomic energy level under consideration and 1 eV $= 1.602 \times 10^{-19}$ J.

Example: Exciting hydrogen atoms

Find the temperature at which the number density of hydrogen atoms in the fundamental ground state is equal to that of the second excited state for $n = 3$. From the Boltzmann equation, we set $N_1/N_3 = (g_1/g_3) \exp[-(E_1 - E_3)/(kT)] = 1$, and using $g_1 = 2$, $E_1 = 0$ for reference, $g_3 = 18$, $E_3 = 12.09$ eV $= 1.939 \times 10^{-18}$ J and $k = 1.38 \times 10^{-23}$ J K^{-1}, we obtain $(1.4036 \times 10^5)/T = \ln(9) = 2.197$, or $T \approx 63{,}900$ K. (The ln denotes the natural logarithm.)

If N_0 is the total number of atoms or ions per unit volume in the ground state with $n = 1$, the total number of excited atoms, N_E, with energy, E_n, above the ground state will be:

$$\frac{N_E}{N_0} = \frac{g_n}{U}\exp\left(-\frac{E_n}{kT}\right) \approx \exp\left(-\frac{E_n}{kT}\right), \tag{6.18}$$

where the partition function for the atom or ion under consideration is given by

$$U = \sum_1^\infty g_n \exp\left(-\frac{E_n}{kT}\right), \tag{6.19}$$

and g_n is the degeneracy of level n. The fraction of atoms or ions in a given energy level is equal to the portion of the partition function related to this level.

When the electron stays within the atom or ion, going from one bound state to another, then the transition is known as a bound–bound transition. If the electron breaks free of the atom or ion, then it is called a bound-free transition. The probability per unit time, P_{mn}, that an atom will undergo a bound–bound transition from a high state of energy, E_m, to a lower state of energy, E_n, is (Einstein 1917a):

$$P_{mn} = A_{mn} + B_{mn}U_\nu, \tag{6.20}$$

where A_{mn} is the Einstein coefficient for spontaneous transition between the two states (spontaneous emission), and B_{mn} is the Einstein stimulated coefficient for a transition induced by radiation of energy density U_ν in the frequency range ν to $\nu + d\nu$. The probability per unit time for a radiation-induced absorption is

$$P_{nm} = B_{nm}U_\nu \tag{6.21}$$

where B_{nm} is the Einstein coefficient for photo-absorption. The energy density of black body radiation at temperature, T, in the frequency range ν to $\nu + d\nu$ is (Planck 1901)

$$U_\nu = \frac{8\pi h\nu^3}{c^3}\left[\exp\left(\frac{h\nu}{kT}\right) - 1\right]^{-1} = \frac{4\pi}{c}B_\nu(T), \tag{6.22}$$

where $B_\nu(T)$ is the brightness of the radiator at frequency ν and the temperature is T. Because the number of downward transitions must equal the number of upward transitions, and because in thermodynamic equilibrium each state has a population determined by the Boltzmann distribution, it follows that

$$A_{mn} = \frac{8\pi h\nu^3}{c^3}B_{mn} \tag{6.23}$$

and

$$g_m B_{mn} = g_n B_{nm}, \tag{6.24}$$

where g_m is the statistical weight of the m level.

The A_{mn} is the probability per second that an atom with an electron in level m will spontaneously emit a photon of energy $h\nu_{mn} = E_m - E_n$ so the energy, E, emitted per unit volume by the spontaneous bound–bound transition is:

$$E = N_m h\nu_{mn} A_{mn}, \tag{6.25}$$

where N_m is the volume density of atoms in level m and $h\nu_{mn} = E_m - E_n$.

6.4 Ionization and Element Abundance in the Sun and Other Stars

Because a greater number of atoms will absorb more light, the relative darkness of the absorption lines in the Sun's spectrum should establish the relative abundance of the elements there. That is, darker, stronger absorption lines generally indicate high absorption and therefore larger amounts of the absorbing element. However, the strength of an element's absorption lines depends only to some extent on the element's abundance.

For example, atoms exist in altered physical states at the high temperatures that prevail within stars. This can result in a change in the wavelength and intensity of the spectral lines observed in stellar atmospheres. In the latter part of the 19th century, Joseph Norman Lockyer (1836–1920) had already shown experimentally that elements exhibit different spectra under varying conditions of temperature and pressure in the terrestrial laboratory, and that, in particular, the arc and higher temperature spark spectra of the same element differed (Lockyer 1887). The Sun is obviously hotter than typical laboratory temperatures, or it would not shine so brightly, and different temperatures and pressures prevail at various locations within the solar atmosphere. Astronomers eventually realized that an element displays different spectral lines depending on the physical conditions of the solar region in which it is located.

Moreover, some stars show conspicuous lines other than the dominant lines in the Sun's spectrum, suggesting that different stars have different compositions. Scientists began to think that this was instead related to the stellar temperature.

Example: Intensity of the red hydrogen alpha transition in the solar atmosphere

The intensity of a spectral line emitted by atoms or ions in an excited state s will depend on the number density, N_s, of atoms or ions occupying the excited state, the energy E_s of that state, and the temperature T. Under conditions of local thermodynamic equilibrium, the Boltzmann distribution indicates that (Boltzmann 1872):

$$\frac{N_s}{N_{tot}} = \frac{g_s}{U} \exp\left(\frac{-E_s}{kT}\right), \tag{6.26}$$

where g_s is the degeneracy, or the statistical weight, of level s, and the partition function for free particles of mass m is $U = (2\pi mkT)^{3/2} V/h^3$ where V is the

volume occupied by the gas, the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1} and the Planck constant $h = 6.626 \times 10^{-34}$ J s.

For the red hydrogen alpha transition, the wavelength $\lambda = 656.28$ nm $= 6.65628 \times 10^{-7}$ m, and the energy $E_s = hc/\lambda = 3.027 \times 10^{-19}$ J, where the speed of light $c = 2.9989 \times 10^8$ m s^{-1}. The number of atoms undergoing this transition in the solar chromosphere and the photosphere depends on the temperature, $T_P = 5{,}780$ K for the photosphere and $T_C = 2 \times 10^4$ K for the chromosphere. The relative number of atoms, N_{sC}/N_{sP}, in the two layers of the solar atmosphere is given by:

$$\frac{N_{sC}}{N_{sP}} = \left(\frac{T_P}{T_C}\right)^{3/2} \frac{\exp\left(\frac{-E_S}{kT_C}\right)}{\exp\left(\frac{-E_S}{kT_P}\right)} \approx 2.3. \tag{6.27}$$

So there are more atoms excited into this state in the chromosphere than the underlying photosphere, because at higher temperatures the exponential function is closer to one and at lower temperatures it is a smaller number.

When the theory of ionization in stellar atmospheres was developed, it became clear that the presence or absence of specific spectral lines did not necessarily indicate the chemical composition of a star's atmosphere. In 1920, Meghnad Saha (1893–1956), a young lecturer at the University of Calcutta, demonstrated that the spectral lines of different elements are excited under different conditions of temperature and pressure. This set the stage for showing that many stars have similar compositions (Saha 1920).

In his analysis, Saha demonstrated the analogy between the dissociation of molecules and the ionization of atoms. He replaced the mass of the atom with the mass of the electron in the expression for the degree of dissociation of a molecule, thereby obtaining his now-famous ionization equation. This formula, known as the Saha equation, relates the degree of ionization of an atom to temperature and pressure, and therefore indicates that the relative intensities of a star's different spectral lines are caused, in part, by differences in the pressure and temperature of the stellar atmosphere.

The fraction x of atoms that are ionized in a gas at a certain temperature T and pressure P is given by:

$$\frac{x^2}{1-x} = \frac{(2\pi m_e)^{3/2}}{h^3} \frac{(kT)^{5/2}}{P} \exp\left(-\frac{\chi}{kT}\right) \tag{6.28}$$

where χ is the ionization potential, or the amount of energy to be supplied to an atom to ionize it, h is the Planck constant and m_e is the mass of the electron. The ionization potentials of some cosmically abundant atoms are given in Table 6.5.

Table 6.5 Ionization potentials χ for different stages of ionization of the most abundant atoms in the cosmos

Z		Stage of Ionization							
		I	II	III	IV	V	VI	VII	VIII
1	H	13.598							
2	He	24.587	54.416						
6	C	11.260	24.383	47.887	64.492	392.077	489.981		
7	N	14.534	29.601	47.448	77.472	97.888	552.057	667.029	
8	O	13.618	35.116	54.934	77.412	113.896	138.116	739.315	871.4

Under conditions of local thermodynamic equilibrium, the number density, N_r, of atoms in the rth stage of ionization is related to that of the $(r + 1)$ state, N_{r+1}, by the Saha equation (Saha 1920, 1921):

$$\frac{N_{r+1}}{N_r} N_e = \frac{U_{r+1}}{U_r} \frac{2(2\pi m_e kT)^{\frac{3}{2}}}{h^3} \exp\left(\frac{-\chi_r}{kT}\right), \tag{6.29}$$

where N_e is the free electron density, U_r is the partition function of the rth stage, the χ_r is the energy required to remove an electron from the ground state of the r-times ionized atom, T is the temperature, m_e is the electron mass, and h and k are respectively the Planck and Boltzmann constants. The electron density N_e can be expressed in terms of the electron pressure $P_e = N_e kT$.

The population density, N_n, of the nth quantum level is given by:

$$N_n = N_e N_i \frac{h^3}{(2\pi m_e kT)^{\frac{3}{2}}} \frac{g_n}{2} \exp\left(\frac{\chi_r - \chi_n}{kT}\right), \tag{6.30}$$

where N_e and N_i are, respectively, the free electron and ion densities, g_n is the statistical weight of the nth level, and χ_n is the excitation energy of the nth level above ground level. For hydrogen like atoms,

$$\chi_n = I_H Z^2 \left(1 - \frac{1}{n^2}\right) \tag{6.31}$$

where $I_H = 13.6$ eV $= 2.179 \times 10^{-18}$ J is the ionization potential of the hydrogen atom, and

$$g_n = 2n^2. \tag{6.32}$$

Saha used his ionization equation in a physical theory for stellar spectra, specifying temperatures of stars of different spectral type. His result showed that differences in stellar spectra are caused by differences in excitation rather than in chemical composition. The English astrophysicists Ralph A. Fowler (1899–1944) and Edward Milne (1896–1950) then showed that the number of atoms or ions responsible for the production of a spectral line can be estimated from the line intensity once the

temperature and pressure of the stellar atmosphere are known (Fowler and Milne 1924). This paved the way for the work of the American astronomer Cecilia H. Payne (1900–1979), who showed that stars with different spectra have essentially the same composition, and it eventually led to the realization that the lightest element, hydrogen, is by far the most abundant element in most stars (Focus 6.1). This discovery also had a fundamental role in understanding how the Sun shines – by fusion reactions of the nucleus of the hydrogen atom, the proton.

Focus 6.1 Hydrogen, the most abundant element in the Sun and most stars

In a brilliant doctoral dissertation written in 1925, the American astronomer Cecilia H. Payne (1900–1979) showed that the atmospheres of virtually every luminous, middle-aged star have the same ingredients (Payne 1925). Her calculations also indicated that hydrogen is by far the most abundant element in the Sun and most other stars. However, she could not believe that the composition of stars differed so enormously from that of the Earth, where gaseous hydrogen is rarely found, so she mistrusted her understanding of the hydrogen atom. Prominent astronomers of the time also did not think that hydrogen was the main ingredient of the Sun and other stars.

Subsequent detailed investigations of the Sun's absorption-line intensities, by the German astronomer Albrecht Unsöld (1905–1995), the British astronomer William H. McCrea (1904–1999) and the American astronomer Henry Norris Russell (1877–1957) showed that the Sun is composed mainly of the lightest element, hydrogen, accounting for 92.1 % of the number of atoms in the Sun (Unsöld 1928; McCrea 1929; Russell 1929). Hydrogen is a million times more abundant than any other element in solar atmosphere, and the number density of hydrogen atoms at the base of the chromosphere is about 10^{18} atoms m^{-3}.

The Danish astronomer Bengt Strömgren (1908–1987) next calculated the hydrogen content in the interior of stars, assuming that they are chemically homogeneous, and showed that their observed luminosities require that the entire star, not just its atmosphere, be composed predominantly of hydrogen (Strömgren 1932).

We now know that very old stars have very few elements other than hydrogen and helium; these stars probably have existed since our Galaxy formed. Middle-aged stars, like the Sun, contain noticeable, but still small, amounts of heavier elements.

Hydrogen is the most abundant element in the stellar universe, so there was nothing wrong with Miss Payne's calculations. The Earth just does not have sufficient gravity to retain hydrogen in its atmosphere for any length of time. Any hydrogen gas that our young planet might have once had must have evaporated away while the Earth was forming and has long since escaped, or become locked into water or surface rocks.

In addition to specifying the compositions of stars, detailed observations of absorption or emission lines yield information about the temperature, density, motion, and magnetism of the Sun and other stars, as well as rarefied nebulae, interstellar matter, and a various cosmic objects.

6.5 Wavelengths and Shapes of Spectral Lines

6.5.1 Radial Motion Produces a Wavelength Shift

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 (See Sect. 4.2). Such a wavelength change is named the *Doppler effect* after the Austrian scientist, mathematician and schoolteacher, Christiaan Doppler (1803–1853) who first explained how it works (Doppler 1842).

The Doppler effect shows that the wavelength of a spectral line that an astronomer observes can differ from the emitted line wavelength. The size of the wavelength change depends on the relative speed of the radiating source along the observer's line of sight, known as the radial velocity. The greater the radial velocity, the larger the Doppler shift. In 1868, the English astronomer William Huggins (1824–1910) was the first to use this method to determine the velocity of a star moving away from the Earth (Huggins 1868).

For a source radial velocity directed away from the observer along the line of sight, the shift is toward longer, redder wavelengths in the visible part of the electromagnetic spectrum, and therefore also is known as a *redshift*. When the motion is toward the observer, there is a *blueshift* to shorter, bluer wavelengths. For a source radial velocity, V_r, away from the observer along the line of sight, the Doppler redshift formula is applicable:

$$z = \frac{\Delta\lambda}{\lambda_L} = \frac{\lambda_{observed} - \lambda_L}{\lambda_L} = \frac{v_L - v_{observed}}{v_{observed}} \approx \frac{V_r}{c} \text{ for } V_r \ll c, \tag{6.33}$$

or

$$\frac{\lambda_{obs}}{\lambda_L} = 1 + z \tag{6.34}$$

and

$$\lambda_{obs} \approx \lambda_L \left(1 + \frac{V_r}{c}\right) \text{ for } V_r << c, \tag{6.35}$$

where the observed line wavelength is, $\lambda_{observed}$, the emitted line wavelength is λ_L, the z is known as the redshift since visible spectral lines are Doppler shifted to longer, redder wavelengths, v denotes frequency and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

For large radial velocities approaching c, the relativistic Doppler effect yields:

$$1 + z = \left[\frac{c + V_r}{c - V_r}\right]^{\frac{1}{2}} \text{ for } V_r \approx c \tag{6.36}$$

and

$$\frac{V_r}{c} = \frac{(z+1)^2 - 1}{(z+1)^2 + 1} \text{ for } V_r \approx c. \tag{6.37}$$

Example: Redshifting Lyman α into the visible

The Lyman α transition from hydrogen atoms in a star that is moving at relatively low speeds with respect to the Earth will be emitted at the rest wavelength of $\lambda_L = 121.567$ nm, in the ultraviolet part of the electromagnetic spectrum. If a galaxy is moving at a high enough speed, it might redshift the emission into a red wavelength of $\lambda_{obs} = 600$ nm, with a redshift $z = (\lambda_{obs} - \lambda_L)/\lambda_L = 3.936$. For redshifts greater than one, the velocities approach that of light, and the radial velocity $V_r = 2.76 \times 10^8$ m s^{-1} using the high-speed equation $\frac{V_r}{c} = \frac{(z+1)^2-1}{(z+1)^2+1}$ for the radial velocity V_r with the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

6.5.2 Gravitational Redshift

For massive, collapsed stars, there also is a detectable gravitational redshift caused by the loss of radiation photon energy in overcoming the immense gravitational pull of a star. This is a small effect for stars like the Sun, whose gravitational redshift is about 2×10^{-6}, but it increases for collapsed stars.

When a photon of energy $hv = mc^2$ leaves the surface of a massive body of mass M and radius R, it loses an energy ΔE given by

$$\Delta E = h\Delta v = \frac{GMm}{R} = \frac{GMhv}{Rc^2} \tag{6.38}$$

where the radiation moves at the speed of light c and is imagined to have an effective mass m. The gravitational redshift is given by

$$z_g = \frac{\Delta v}{v} = \frac{\Delta \lambda}{\lambda} = \frac{GM}{Rc^2}. \tag{6.39}$$

The gravitational redshift for the Sun is $z_g \approx 2 \times 10^{-6}$, which is a value that would be produced by the Doppler effect at a radial velocity of only 0.64 km s^{-1}, and so is very difficult to measure. Nevertheless, such a measurement has been

reported (Blamont and Roddier 1961; Forbes 1970; Snider 1972). The effect increases for collapsed stars, and it has been measured for white dwarf stars with a mass about equal to that of the Sun and a size about equal to that of the Earth (Adams 1925; Popper 1954; Greenstein et al. 1971; Greenstein and Trimble 1972; Shipman 1972, Sect. 13.3)

The most accurate tests of the gravitational redshift are made in the Earth's gravity, which is much weaker than that of the Sun or white dwarf stars. The greater terrestrial accuracy in the measurement of this effect is possible because the wavelengths of nuclear spectral lines are more precisely known than those of the spectral lines of any celestial object. The discovery of narrow gamma ray spectral lines of nuclei (Mössbauer 1958) led to the first detection of the terrestrial gravitational redshift. Using the Mössbauer effect, Pound and Rebka (1959, 1960) measured a value of $z_g = (2.57 \pm 0.26) \times 10^{-15}$ as opposed to the predicted value of $z_g = 2.46 \times 10^{-15}$. The agreement between theory and experiment was subsequently improved to about one percent (Pound and Snider 1964, 1965)

Example: Measuring the gravitational redshift on the Earth and stars

Due to the Earth's relatively low gravity, the gravitational redshift measurement requires radiation at a very precise gamma-ray wavelength, generated by a nuclear resonance Mössbauer effect. In going from the ground to a height H, the wavelength change due to the gravitational redshift will be $\Delta\lambda/\lambda = GM_E H/(R_E^2 c^2) \approx 10^{-16} H$, where $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}, the mass of the Earth $M_E = 5.974 \times 10^{24}$ kg, the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, and the radius of the Earth $R_E = 6.371 \times 10^6$ m. Pound and Rebka (1959, 1960) measured the change over a distance of 22.5 m, requiring a measurement precision of 2×10^{-15}.

For the Sun, the gravitational redshift is $\Delta\lambda/\lambda = GM_\odot/(R_\odot c^2) \approx 2 \times 10^{-6}$, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg and the Sun's radius $R_\odot = 6.955 \times 10^8$ m. By way of comparison, the thermal broadening $\Delta\lambda/\lambda_L = V_{thermal}/c = (3kT/m)^{1/2}/c = 4 \times 10^{-5}$ for a spectral line of hydrogen, of mass $m = 1.67 \times 10^{-27}$ kg, emitted from the visible disk of the Sun at temperature $T = 5780$ K, where the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1} (also see next example). So the thermal broadening effect is about 16 times greater than the gravitational redshift effect, making the latter very difficult to detect.

A white dwarf star has about the same mass as the Sun, but a radius that is comparable to that of the Earth and 100 times smaller than the Sun, at about 6.37×10^6 m, so the gravitational redshift is about 100 times larger and easier to detect (Sect. 13.3).

A more exact, relativistic expression for the gravitational redshift, which is applicable to a very massive, collapsed object is (Einstein 1911, 1916):

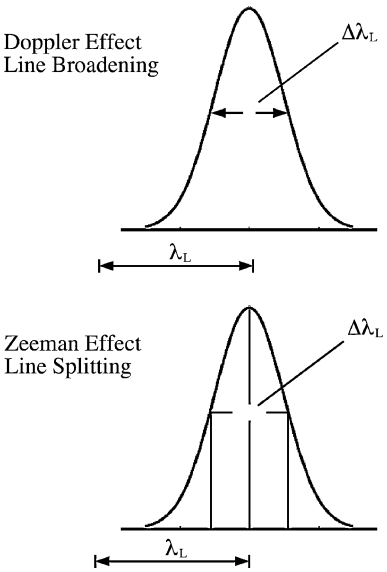


Fig. 6.8 Effects that broaden a spectral line The motion of absorbing or emitting atoms can broaden a line to the short-wavelength and long-wavelength side of the resting, or nonmoving, wavelength, here denoted by λ_L (*top*). The Doppler effect describes the broadening (see Fig. 4.2). When the motion is due to the heat or temperature of the radiating atoms, the effect produces thermal broadening; when the average temperature of a collection of atoms increases, the thermal broadening becomes wider. The Doppler effect broadening can also be caused by the rotation or expansion of the source. An intense magnetic field can split a single line at wavelength λ_L into two components by the Zeeman effect (*bottom*). The wavelength difference $\Delta\lambda_L$ between the split lines is proportional to the strength of the magnetic field. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

$$z_g = \left(1 - \frac{2GM}{Rc^2}\right)^{-1/2} - 1. \tag{6.40}$$

Vessot et al. (1980) have compared the time of a hydrogen maser clock in a rocket with a similar clock on the ground, confirming the relativistic gravitational redshift within the uncertainties of the measurement.

6.5.3 Thermal Motion Broadens Spectral Lines

Any observed spectral line is the superposition of the lines emitted by many individual atoms in different physical conditions. Rather than appearing at a single wavelength, the observed line therefore is broadened over a range of wavelengths (Fig. 6.8). Van Vleck and Haber (1977) have reviewed absorption and emission lines and line breadths.

The amount of wavelength broadening increases with the temperature of the source or, to be exact, it varies as the thermal velocity of the moving atoms and the square root of the temperature.

For atoms in thermodynamic equilibrium at temperature T, a line emitted at wavelength λ_L will be Doppler broadened by an amount $\Delta\lambda_D$ given by:

$$\Delta\lambda_D = \lambda_L\left(\frac{V_{thermal}}{c}\right), \tag{6.41}$$

where $c = 2.9979 \times 10^8$ m s^{-1}, and the thermal velocity $V_{thermal}$ of an atom of mass m at temperature T is given by:

$$V_{thermal} = \sqrt{\frac{3kT}{m}} = \left[\frac{3kT}{m}\right]^{1/2}, \tag{6.42}$$

where the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1}.

Example: Thermal broadening of atomic hydrogen lines in the photosphere

The temperature of the visible solar disk, the photosphere, is $T = 5{,}780$ K, so the thermal velocity, $V_{thermal}$ of a hydrogen atom of mass $m_H = 1.007825$ u $= 1.6735 \times 10^{-27}$ kg, will be $V_{thermal} = (3kT/m_H)^{1/2} = 1.20 \times 10^4$ m s^{-1}, where the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1}. The wavelength broadening $\Delta\lambda$ in wavelength λ is given by $\Delta\lambda/\lambda = V_{thermal}/c = 4.0 \times 10^{-5}$, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

6.5.4 Rotation or Expansion of the Radiating Source can Broaden Spectral Lines

If a source is rotating, the Doppler effect of the object's side rotating toward the observer produces a blueshift to shorter wavelengths; the other side, which is rotating away, shifts a line to longer wavelengths. The combined effect produces a line broadening that increases with the rotation velocity and that depends on the projected linear equatorial velocity or the observed rotational velocity of the line.

If a source is rotating with an equatorial rotation velocity V_{rot} it will produce a line broadening by an amount $\Delta\lambda_{rot}$ given by:

$$\Delta\lambda_{rot} = \lambda_L\left(\frac{V_{rot}\sin i}{c}\right), \tag{6.43}$$

where i is the inclination of the source's equator to the celestial equator. The quantity of V_{rot} sin i is the projected linear equatorial velocity or the observed rotational velocity of the line.

A similar broadening applies to an expanding source, which exhibits a line broadening that increases with the expansion velocity, $V_{\exp}$, producing

$$\Delta\lambda_{\exp} = \lambda_L \left(\frac{V_{\exp}}{c} \right). \tag{6.44}$$

6.5.5 Curve of Growth

The Belgian astronomer Marcel Minnaert (1893–1970) and his Dutch colleague Gerard F.W. Mulders (1908–1993) introduced the equivalent width, W_λ, or the area of an absorption line profile, as a measure of line intensity (Minnaert and Mulders 1930). By plotting the equivalent widths against the number of absorbers (in a logarithmic form), they showed for the first time the empirical curve of growth, which permits the comparisons of line intensities and widths with theoretical expectations. It is mainly of concern for advanced studies of stellar atmospheres.

The two classical model atmospheres are known as the Schuster-Schwarzschild atmosphere (Schuster 1905; Schwarzschild 1906) and the Milne-Eddington atmosphere (Milne 1921, 1930; Eddington 1917, 1926, b). In local thermodynamic equilibrium they both lead to the approximate relation for the equivalent width, W_λ, of a line of intensity I_λ at wavelength, λ. If we define

$$W_\lambda = \int \frac{I_C - I_\lambda}{I_C} d\lambda, \tag{6.45}$$

where the subscripts C and λ denote the continuum adjacent to the line and the line wavelength, we have:

$$\frac{W_\lambda}{\lambda} = \frac{e^2}{4\varepsilon_0 m_e c^2} N f \lambda, \tag{6.46}$$

where the electron charge is e, the electric constant is ε_0, the electron mass is m_e, the speed of light is c, the N is the column density of the atoms producing the spectral line and the oscillator strength f is related to the absorption cross section σ by

$$\sigma = \frac{e^2}{4\varepsilon_0 m_e c} f. \tag{6.47}$$

Each spectral line is characterized by the oscillator strength *f*, and the larger the *f* the stronger the spectral line that is seen. We can determine *N* once the oscillator strength is known and the equivalent width measured.

Observations of the intensities and widths of spectral lines can be compared with theoretical expectations in order to determine the excitation temperature, the turbulent velocity, the electron and gas pressures, the surface gravity, and the abundance of the elements in a stellar atmosphere. Nowadays there are complex theoretical models for the atmospheres, and relevant data is on the Internet.

6.5.6 Magnetic Fields Split Spectral Lines

When an atom is placed in a magnetic field, it acts like a tiny compass, and it adjusts the energy levels of its electrons. If the atomic compass is aligned in the direction of the magnetic field, the electron's energy increases. If it is aligned in the opposite direction, the energy decreases. Because each energy change coincides with an alteration in the wavelength or frequency of the radiation emitted by that electron, a spectral line emitted at a single wavelength by a randomly oriented collection of atoms becomes a group of three lines of slightly different wavelengths in the presence of a magnetic field (Fig. 6.9). The size of an atom's internal

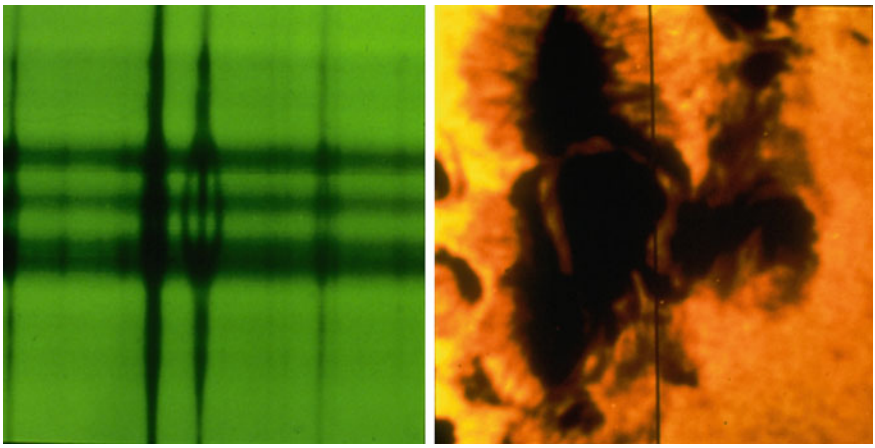


Fig. 6.9 The Zeeman effect The magnetic field in a sunspot can be measured using the Zeeman effect. In a sunspot (*right*) the spectral lines that are normally at a single wavelength become split into two or three components (*left*), depending on the orientation of the field with respect to the line of sight. The vertical line crossing the sunspot denotes the alignment of the observing instrument slit. The separation of the outermost components is proportional to the strength of the magnetic field, in this sunspot about 0.4 tesla, or 4,000 gauss. The components also have a circular polarization, which indicates the direction of the longitudinal magnetic field. (Courtesy of NOAO.)

adjustments, and the extent of its spectral division, increases with the strength of the magnetic field.

This magnetic transformation was named the *Zeeman effect*, after Pieter Zeeman (1865–1943), who first noticed it in a laboratory (Zeeman 1896, 1897a, b). His Dutch colleague Hendrik A. Lorentz (1853–1928) predicted the effect (Lorentz 1898). The pair received the 1902 Nobel Prize in Physics for their work.

We can gain some insight to the Zeeman effect by considering the motion of a free electron in the presence of a magnetic field. A charged particle cannot move straight across a magnetic field, but instead gyrates around it. If the particle approaches the magnetic field in a perpendicular direction, a magnetic force pulls it into a circular motion about the magnetic field line. Because the particle can move freely in the direction of the magnetic field, it spirals around it with a helical trajectory.

The size of the circular motion, called the radius of gyration and designated R_g, depends on the velocity, $V_\perp$, of the particle in the perpendicular direction, the magnetic field strength, B, and the mass, m, and charge, Ze, of the particle. That gyration radius is described by the equation (Larmor 1897):

$$R_g = \left(\frac{mV_\perp}{ZeB}\right), \tag{6.48}$$

provided that the velocity is not close to the speed of light, c. An electron will circle about the magnetic field with a radius, R_g, and with a period, P, given by:

$$P = \frac{2\pi R_g}{V_\perp} = \frac{2\pi m_e}{eB}. \tag{6.49}$$

At the velocity $V_\perp$, the electron goes once around the circumference $2\pi R$ in the period P. The rotational frequency, ν_g, of this circular motion, in a plane perpendicular to the magnetic field, is

$$\nu_g = \frac{1}{P} = \frac{eB}{2\pi m_e} \approx 2.80 \times 10^{10} B \,\text{Hz}, \tag{6.50}$$

where the elementary charge of the electron is $e = 1.6022 \times 10^{-19}$ C, the mass of the electron is $m_e = 9.1094 \times 10^{-31}$ kg, and B is the magnetic field strength in tesla.

When an atom is placed in a magnetic field, a very similar thing happens to its electrons and the spectral lines they emit. A line that radiates at a wavelength λ_L without a magnetic field becomes split into two or three components depending on the orientation of the magnetic field. This has been detected in sunspots (Fig. 6.8).

We can represent the orbital motion of the electron using the Bohr model of the atom for the electron in the ground state orbit with quantum number $n = 1$. The orbital angular momentum in a circular orbit will be $m_e VR = nh/(2\pi) = h/(2\pi)$ for orbital velocity V, orbital radius R, and h denotes the Planck constant with $h = 6.6261 \times 10^{-34}$ J s. When the atom is placed in a magnetic field of strength B, the magnetic moment, designated μ, is the Bohr magneton, denoted by μ_B, given by:

$$\mu_B = \frac{e}{2m_e}(mVR) = \frac{eh}{2\pi m_e} = 9.274 \times 10^{-24}\ \text{J T}^{-1}, \tag{6.51}$$

where J denotes the joule units of energy and T designates the tesla unit of magnetic field strength.

When the atom is placed in the magnetic field, a single spectral line is split into three closely spaced spectral lines. The adjacent energy levels differ in energy, ΔE_B, corresponding to a photon energy difference $h(\nu_1 - \nu_2)$ related by:

$$\Delta E_B = \mu_B B = \frac{eh}{4\pi m_e} B = h(\nu_1 - \nu_2). \tag{6.52}$$

The two outer components of the three Zeeman-split spectral lines will be separated by an energy difference $2\Delta E_B$, and a total frequency separation $\Delta\nu$ given by:

$$\Delta\nu = \frac{2\mu_B}{h} B \approx 2.80 \times 10^{10} B\ \text{Hz}. \tag{6.53}$$

For the three component split, the shift, $\Delta\lambda$, in wavelength of the two outer components is given by:

$$\Delta\lambda = \frac{\lambda^2}{c}\Delta\nu \approx 93.3\lambda^2 B\ \text{m}, \tag{6.54}$$

where the shift in frequency is $\Delta\nu$, and the speed of light is $c = 2.9979 \times 10^8\ \text{m s}^{-1}$, both $\Delta\lambda$ and λ are in meters, and B is in tesla. The separation is thus proportional to the magnetic field strength B.

Example: Zeeman effect in sunspots and interstellar space

The magnetic field strength in a sunspot measured by Hale (1908a, b) was $B = 0.3$ tesla. For the red Hα transition at a wavelength of $\lambda = 656.3\ \text{nm} = 6.56 \times 10^{-7}$ m, the Zeeman effect will give a line splitting of $\Delta\lambda = 93.3\ \lambda^2\ B \approx 1.20 \times 10^{-11}$ m or $\Delta\lambda/\lambda \approx 1.80 \times 10^{-5}$. The Doppler broadening of this line due to the thermal motion of the hydrogen atoms at a temperature of 5,780 K is $\Delta\lambda/\lambda = V_{thermal}/c = 4.0 \times 10^{-5}$ (see previous example), only twice the expected.

The interstellar spaces are very cold, about 100 K or less, with less thermal broadening, but an accuracy of $\Delta\lambda/\lambda = 7 \times 10^{-10}$ is required to make the measurement at a wavelength of the $\lambda = 21$ cm transition of interstellar hydrogen.

The American solar astronomer George Ellery Hale (1868–1938) first made measurements of this Zeeman splitting in sunspots, showing that they have magnetic field strengths of about 0.3 tesla (Hale 1908a, b). This is about 10,000 times the strength of the terrestrial magnetic field that orients our compasses. The split lines

Table 6.6 Cosmic magnetic fields

Object	Magnetic field strength, B (tesla)[a]
Earth (equator to pole)	3×10^{-5}–6×10^{-5}
Solar wind (at Earth orbit)	2.5×10^{-9}
Sunspot	0.3
Sun (global)	3×10^{-4}
Interstellar space	10^{-10}
White dwarf star	10^{2}
Pulsar	10^{8}

[a] 1 tesla = 1 T = 10^4 gauss = 10^4 G

are circularly polarized, and the direction of polarization indicates the direction of the magnetic field, pointing into or out of the Sun (Hale et al. 1919).

As suggested by Bolton and Wild (1957), the magnetic field strength in interstellar space, with about $B = 10^{-10}$ tesla, can be observed at radio wavelengths (Verschuur 1968, 1971; Troland and Heiles 1977).

Ledoux and Rensen (1966) have provided an early summary of magnetic stars; Donati and Landstreet (2009) reviewed magnetic fields in nondegenerate stars. Angel (1978) provided a review of magnetic white dwarf stars. Crutcher (2012) discusses magnetic fields in molecular clouds; Beck et al. (1996) review galactic magnetism; Sofue et al. (1986) reviewed the global structure of magnetic fields in galaxies, and Carilli and Taylor (2002) reviewed observations of magnetic fields in clusters of galaxies.

The magnetic field strengths, B, of various cosmic objects are listed in Table 6.6.

All of this refers only to the outer atmosphere of the Sun and other stars. As it turns out, atoms do not exist in most of the Sun, except in the cool visible layer in which dark absorption lines are formed. It is too hot everywhere else for whole atoms to survive. Innumerable collisions fragment the abundant hydrogen atoms into their constituent pieces. Their protons and electrons have been set free from their atomic bonds, wandering throughout the solar material unattached to each other. Some of these protons merge together near the Sun's center, but an understanding of these nuclear fusion reactions had to await investigations of how nuclear particles occasionally escape the tight confines of radioactive elements.

Chapter 7
Transmutation of the Elements

7.1 The Electron, X-rays and Radium

When voltage is applied to the ends of a glass tube filled with rarefied gas, an arc of light passes from the cathode – negative end – of the gas tube to the other end – the anode, which is positive. If vacuum pumps are used to reduce the gas pressure in the glass tubes, they cease to glow inside, but the glass shines where the invisible cathode rays apparently are striking it.

The French scientist Jean Perrin (1870–1942) used a magnet to deflect the location of the glowing spot on the glass, suggesting that the unseen streams of "rays" were beams of negatively charged particles (Perrin 1895). A magnetic field deflects a negative charge in one direction and a positive charge in the other direction, and the size of the deflection depends on the mass of the charged particle, its velocity, and the strength of the magnetic field. The British physicist Joseph John Thomson (1856–1940) subsequently used both electric and magnetic fields to measure the deflection of the beams of charged particles, now called *electrons*.

The deflection experiments worked this way: an electrical field produces a force that bends a beam of electrons in the direction of the field, toward positive and away from negative electrodes. By measuring the deviation from the original direction of motion, the charge to mass ratio e/m can be determined. Because the deflection also depends on the velocity of the electrons, another measurement was needed, which was accomplished by placing a magnet near the electron beam. Combining both the electrical and magnetic experiments determines the charge to mass ratio of the electron. So, it was from this ratio and the fundamental unit of charge that the mass of the electron was determined. Thomson concluded that the electron is much less massive than any atom. In fact, it is roughly 1,000 times less massive than the lightest atom, hydrogen (Thomson 1897a, b, 1903).

Meanwhile, the German physics professor Wilhelm Röntgen (1845–1923) inadvertently left some wrapped, unexposed photographic plates near a glowing gas tube that he was studying. Later, he found that the plates were fogged and that

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_7, © Springer-Verlag Berlin Heidelberg 2013

this always happened when other new plates were left near the apparatus. Röntgen concluded that invisible rays were passing out of the tube and fogging the plates.

In those days, photographs were taken with glass plates covered with a light-sensitive emulsion of silver salts. This preceded the use of photographic film, which has been replaced by the charge-coupled devices used in digital cameras.

To remove any light, Röntgen enclosed the electrical discharge tube in black cardboard, and noticed a glow coming from a nearby sheet of paper coated with a substance (barium platinocyanide) known to shine in strong light – but there was no light to make it glow. The cause seemed to be the same invisible rays that fogged the photographic plates. Moreover, when Röntgen's wife placed her hand between the electrified gas tube and a photographic plate, the developed photograph showed the bones of her hand and the ring she was wearing (Röntgen 1896).

The report of these previously unknown and penetrating rays caused a public sensation, for they could see inside humans and reveal the invisible. The rays were able to penetrate skin and muscle, detecting human bones, which revolutionized medicine.

No one knew what these mysterious, penetrating emissions were, so Röntgen called them x-rays, using the mathematical designation x for something unknown. Subsequent investigations eventually showed that the x-rays are electromagnetic radiation of very short wavelength and high photon energy.

In 1896, at the Sorbonne in Paris, Henri Becquerel (1852–1908) was investigating fluorescent substances that could collect the energy of light and remain luminous after the light source was removed. Uranium salts, for example, glowed in the dark, and Becquerel thought that they also might emit x-rays after being stimulated by sunlight. However, clouds shut out the Sun, so he tossed the packet of uranium salts into a drawer in his worktable.

A box of unexposed photographic plates had been left in the drawer, wrapped in thick black paper and never opened; this led to another accidental discovery. When Becquerel took out the plates a few days later and developed photographs taken with them, he found that they were ruined, as if they had previously been exposed to light. The uranium salts were emitting unseen rays that could affect the plates, even in pitch darkness. Becquerel called the invisible, highly penetrating phenomenon uranic rays (Becquerel 1896). For a time they were called Becquerel rays, and eventually they became known as radioactive rays, but for decades, no one knew exactly what they were.

The new type of rays passed without difficulty through a covered box of photographic plates, but they might not penetrate metal. Becquerel repeated the experiment with an iron key placed between new photographic plates and the uranium. When the plates were developed, they showed the silhouette of the key. This indicted that the uranium was emitting rays that were unable to pass through the iron, even though they could penetrate dark paper that blocks ordinary light. In this respect, the uranium rays resembled x-rays; but unlike x-rays, the uranium was emitting rays spontaneously without any previous excitation by sunlight or electricity.

Hearing of Becquerel's discovery, Pierre Curie (1859–1906), also a professor of physics at the Sorbonne, and the young graduate student he recently had married, Manya (Marie) Curie (1867–1934), began to investigate the new type of rays. Madame Curie wanted to know if uranium was the only element that emitted the mysterious rays, and developed methods to measure the amounts being released. To her surprise, she found that impure uranium ores emitted more rays than could be explained in terms of the uranium they contained. The couple began a laborious two-year search for the unknown emitters; from one ton of the uranium ore known as pitchblende they extracted just a few grams of powerful new elements that had not been known previously (Curie 1898; Curie and Curie 1898). One was called radium and the other polonium, after Marie's native Poland.

Radium is one million times more radioactive than uranium, which – in the terminology of the time – meant that radium is emitting the penetrating rays more intensely than uranium, not that either substance emits radio waves. Crystals containing radium can light up an otherwise dark room, and also burn the skin, as Curie discovered to his dismay.

7.2 Radioactivity

At the time of their discovery, no one knew exactly what radioactive rays were, where their energy came from, or why the radioactive materials kept pouring out energy. Moreover, the amount of energy being released by radium was difficult to explain, for it far surpassed anything that had been achieved by chemical reactions. And, it is a natural process that happens all the time in the ground on which we stand.

It was the English physicist Ernest Rutherford (1871–1937) who found that the source of radioactive energy must come from the interior of the radioactive atoms. These very heavy atoms were unstable, disintegrating all by themselves and slowly leaking energy from their interiors in spontaneous transmutation (Rutherford 1904).

Rutherford previously found that the radioactive rays emitted by uranium included at least two distinct types, termed *alpha rays* and *beta rays* (Rutherford 1899). These rays are not waves of radiation; they instead are beams of energetic, fast-moving particles. By using electrical and magnetic fields, the two types of particles could be separated and their physical properties examined. The directions in which the beams were deflected indicated the sign of their electrical charge, and the magnitude of the deflection provided a measure of both the charge and the mass.

The most energetic particles emitted by radioactive substances are alpha particles (Ramsay and Soddy 1903). Alpha particles carry a double dose of positively charged protons, and they move at astonishingly high speeds – about 1.5×10^7 m s^{-1}, or one-twentieth the speed of light (Rutherford 1911). This corresponds to a typical kinetic energy of about 5 MeV, or five mega-electron-volts and 8.01×10^{-13} J. Rutherford and his colleagues eventually showed that an alpha particle is nothing

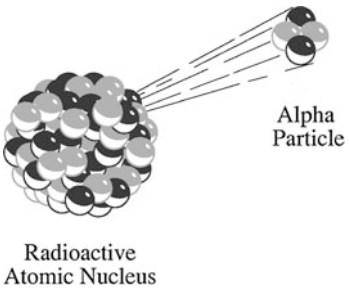


Fig. 7.1 Radioactive alpha decay An unstable, heavy nucleus of a radioactive element can disintegrate or decay into a stable, lighter nucleus, with the emission of an alpha particle that carries mass away from the heavy nucleus during its decay. The subatomic alpha particle consists of two protons and two neutrons. The nucleus of a helium atom is an alpha particle. Radioactive alpha decay of an individual heavy element such as uranium does not occur very often, on average. We would have to wait 4.47 billion years for half of a rock of uranium to change into lead by emitting alpha particles. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2012. Reprinted with permission.)

more than the nucleus of the helium atom, containing two protons and two neutrons. The alpha particles were being emitted by a heavy, unstable atom that was turning spontaneously into a slightly lighter atom, without any interactions with another particle or radiation from outside the atom (Fig. 7.1).

Radioactive decay is occasionally accompanied by the emission of charged beta rays, which make an electronic adjustment to an atom's nucleus without significantly changing its mass. Beta rays consist of negatively charged particles of low mass, which eventually were identified as high-speed electrons.

The ejection of alpha and beta particles often is accompanied by the emission of powerful electromagnetic radiation, akin to x-rays but with even shorter wavelength and greater energy (Villard 1900; Rutherford and Andrade 1914). Because the energetic radiation, called gamma rays and designated γ, is not charged, electrical or magnetic fields do not deflect it.

Working with the young chemist Frederick Soddy (1877–1956), Rutherford found that radioactive atoms continued to disintegrate into other pieces after the emission of an alpha particle (Rutherford and Soddy 1902, 1903). Uranium, for example, initially turned into thorium, which also was radioactive and released other substances, including gaseous radon.

However, the progressive disintegration of heavy, unstable elements does not continue forever. As radioactive decay progresses, the inner parts of an atom rearrange into greater stability, eventually reaching an equilibrium that does not decay. For uranium, this stable endpoint is lead.

A simplified notation, described in Focus 7.1, clarifies how these complex chains of radioactive decay work.

Focus 7.1 Nuclear nomenclatures

The number of protons in a nucleus is denoted by the atomic number Z. They account for the charge of the nucleus but not for all of its mass. A nucleus of any element except hydrogen has about twice the mass of the sum of its protons. The extra mass is due to neutral, or uncharged, particles called neutrons, each with about the same mass as a proton. The mass of the nucleus is specified by the mass number $A = N + Z$, where N is the number of neutrons; the mass number A also gives the number of nucleons in the nucleus, which is the sum of the number of protons and the number of neutrons.

Letters denote the nuclei and other subatomic particles. Both a letter and a superscript, the mass number A, designate a nucleus. An arrow $\rightarrow$ specifies the reaction. Nuclei on the left side of the arrow react to form products given on the right side of the arrow. The alpha decay of uranium, for example, is given by

$$^{238}\mathrm{U} \rightarrow {}^{234}\mathrm{Th} + \alpha \tag{7.1}$$

or

$$^{238}\mathrm{U} \rightarrow {}^{234}\mathrm{Th} \ + {}^{4}\mathrm{He}, \tag{7.2}$$

where U denotes a nucleus of uranium, Th indicates a nucleus of thorium, and the alpha particle α is the nucleus of the helium atom, also designated ^{4}He.

Another way of describing the alpha decay of a nucleus of mass number A and atomic number Z, the number of protons, is:

$$(A, Z) \rightarrow (A - 4,\, Z - 2) + \alpha, \tag{7.3}$$

where the helium nucleus α, with $A = 4$ and $Z = 2$, balances out both the mass and charge on each side of the reaction. The total sum of neutrons and protons on the left side of the arrow is equal to the total sum on the right side; it's just that they have been redistributed.

The spontaneous decay of the uranium, U, radium, Ra, and polonium, Po, nuclei can be represented by:

$$\begin{aligned} ^{238}\mathrm{U} &\rightarrow {}^{234}\mathrm{Th} + {}^{4}\mathrm{He} + \mathrm{Energy}, \\ ^{226}\mathrm{Ra} &\rightarrow {}^{222}\mathrm{Rn} + {}^{4}\mathrm{He} + \mathrm{Energy}, \end{aligned} \tag{7.4}$$

and

$$^{210}\mathrm{Po} \rightarrow {}^{206}\mathrm{Pb} + {}^{4}\mathrm{He} + \mathrm{Energy}, \tag{7.5}$$

where the arrow denotes the decay of the nucleus on the left side to the nuclei on the right side, the superscripts denote the number of protons and

neutrons in the nucleus, *Th* denotes the thorium nucleus, *Rn* denotes the nucleus of radon, *Pb* denotes the lead nucleus, and *He* is the helium nucleus, the alpha particle. Both radium and polonium are also natural products in the decay chain of radioactive uranium.

7.3 Tunneling Out of the Atomic Nucleus

Why don't the nuclei of the radioactive atoms decay completely all at once? Or to ask a related question, how have the nuclei of so many uranium atoms managed to retain their alpha particles? After all, there is still plenty of uranium around billions of years after the Earth formed, continuing to make the rocks and soil around us radioactive. The reason is that it takes significant energy to break free of the strong forces that bind the protons and neutrons together in the nucleus of an atom. They are locked so firmly within the nucleus that exceptional force must be applied to dislodge them.

In fact, the escape of an alpha particle from an atomic nucleus seemed impossible from the viewpoint of classical physics. The forces holding the particle inside the nucleus are so strong that the energy required to overcome them is enormous. It is as if the atomic nucleus is surrounded by tall walls with an energy much higher than that of the alpha particles.

The Russian physicist George Gamow (1904–1968) resolved this paradox in 1928, by using the uncertain, probabilistic nature of quantum theory to explain the mechanism of radioactive decay. Quantum theory indicated that the location of a tiny subatomic particle is not defined precisely. It might be anywhere, although with decreasing probability at regions far from the most likely location.

This explains the escape of alpha particles from the nuclei of radioactive atoms such as uranium. These particles usually lack the energy to overcome the nuclear barrier, but some have a small probability of escaping to the outside world. And because there is some probability that they will "leak through" the nuclear walls, eventually some do. The chance of escaping the atomic nucleus increases with the kinetic energy of the subatomic particle and decreases with its electrical charge.

Not all nuclei are radioactive. For a nucleus to be unstable there has to be available a final state in which the sum of the masses of the decay products is less than the mass of the initial nucleus.

So, how fast is the rate of decay? Gamow's (1928) quantum–mechanical formula for the "transparency" of the nuclear walls was in good agreement with Rutherford's suggestion that the number of radioactive atoms, N, in a rock changes with time, t, since its solidification with a constant rate of decay according to the equation (Rutherford 1900):

$$\frac{dN}{dt} = -\lambda N \tag{7.6}$$

where the radioactive decay constant $\lambda = 0.693/\tau_{1/2} = \ln 2/\tau_{1/2}$ and $\tau_{1/2}$ is the half-life of the radioactive species.

The decay rates of various radioactive substances differ. The nuclei of uranium can retain their alpha particles for billions of years, whereas other radioactive nuclei eject them in a matter of seconds. These rates are quantified in terms of the nuclear half-life, which is the time needed for a given amount of a radioactive substance to decay to half of its initial value.

The decay is probabilistic, governed by the rules of quantum mechanics that describe a random, slow, and statistical decay. On the level of a single atom, it is impossible to predict when a given atom will decay, and the probability that a given unstable atom decays is the same for all atoms of that type, independent of age. For numerous identical atoms, the decay rate is predictable using quantum theory, and that rate depends on the radioactive element under consideration.

At almost the same time as Gamow's discovery, the British physicist Ronald W. Gurney (1898–1953) and the American physicist Edward U. Condon (1902–1974) developed a similar explanation for spontaneous radioactive decay (Gurney and Condon 1928).

The decay equation integrates to give the number of radioactive atoms, $N(t)$ at time t:

$$N(t) = N_0 \exp(-\lambda t) = N_0 \exp\left(\frac{-0.693\, t}{\tau_{1/2}}\right), \tag{7.7}$$

where N_0 is the number of atoms at time $t = 0$, the time of solidification, λ is the radioactive decay constant, $\tau_{1/2}$ is the half-life of the radioactive atom, and 0.693 is the natural logarithm of 2 denoted ln2.

The mean lifetime, τ, of a radioactive atom is the inverse of the decay constant, or

$$\tau = \frac{1}{\lambda} = \frac{\tau_{1/2}}{\ln 2} = \frac{\tau_{1/2}}{0.693}, \tag{7.8}$$

where the radioactive decay constant $\lambda = 0.693/\tau_{1/2} = \ln(2)/\tau_{1/2}$, and $\tau_{1/2}$ is the half-life of the atom.

The half-life is the time required for a quantity of radioactive atoms to fall to half its value as measured at the beginning of the time period, following the exponential decay equation. For example, uranium 238 decays into thorium 234 with the emission of an alpha particle and a half-life of 4.46×10^9 years, or 4.46 billion years. The thorium eventually decays into radium 226, which itself decays into radon gas with the emission of another alpha particle and a half-life of 1,601 years. The radon eventually decays into polonium 210, which decays into stable lead 206 with the emission of another alpha particle and a half-life of just 138.4 days.

Table 7.1 Long-lived radioactive isotopes used for dating

Radioactive parent	Stable daughter	Half-life (years)
Rubidium (Rb) 187	Strontium (Sr) 87	48.8 billion
Rhenium (Re) 187	Osmium (Os) 187	44 billion
Lutetium (Lu) 176	Halfnium (Hf) 176	35.7 billion
Thorium (Th) 232	Lead (Pb) 208	14.05 billion
Uranium (U) 238	Lead (Pb) 206	4.47 billion
Potassium (K) 40	Argon (Ar) 40	1.27 billion
Samarium (Sm) 146	Neodymium (Nd) 142	0.10 billion
Uranium (U) 235	Lead (Pb) 207	704 million
Plutonium (Pu) 244	Thorium (Th) 232	83 million
Iodine (I) 129	Xenon (Xe) 129	16 million
Palladium (Pd) 107	Silver (Ag) 107	6.5 million
Manganese (Mn) 53	Chromium (Cr) 53	3.7 million
Aluminum (Al) 26	Magnesium (Mg) 26	0.72 million

The parent radioactive atoms eventually decay into stable daughters, and the amount of the daughter steadily increases at the expense of the parent.

Half-lives for the decay of long-lived radioactive parent isotopes are listed in Table 7.1 together with the stable daughter isotopes. As an example, it takes 4.47 billion years for half the atoms in a lump of uranium to change into lead.

This provides us with a way of measuring the age of the solar system. You just measure the relative amounts of radioactive parents and non-radioactive daughters. 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 daughters can escape easily when the rock is molten; only when it cools and solidifies do they start to accumulate. For this reason, the age determined for a rock is actually the time since the rock became solid. If the rock is remelted, for example by the impact of a meteorite, its radioactive clock is reset and the age will measure the time since the last solidification.

Radioactive dating of primitive meteorites, ancient rocks returned from the Moon, and deep ocean sediments indicates an age of about 4.6 billion years. These relics have remained unaffected by the geological erosion processes that removed the primordial record from most terrestrial rocks. If the solar system originated as one entity, then this also should be the approximate age of the Sun and the rest of the solar system (Focus 7.2).

Focus 7.2 The age of the solar system

Radioactive elements can be used to clock the age of rocks on the Earth's surface, meteorites, and lunar rock samples. The number of radioactive atoms in the rock will be halved in a time equal to the half-life. Radioactive uranium, U^{238}, decays, for example, into lead, Pb^{206} (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 U^{238},

and express the abundance in terms of another kind of lead, Pb^{204}, which 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 abundances of lead and uranium are related by the equation:

$$\left(\frac{Pb^{206}}{Pb^{204}}\right)_t = \left(\frac{U^{238}}{Pb^{204}}\right)_t [\exp(\lambda_{238}\, t) - 1] + \left(\frac{Pb^{206}}{Pb^{204}}\right)_0, \tag{7.9}$$

where the subscripts t and 0 denote the present and initial abundance, respectively.

If all of the rock samples have the same initial Pb^{206}/Pb^{204} abundance, and if all of them have the same age, t, then a plot of $(Pb^{206}/Pb^{204})_t$ against $(U^{238}/Pb^{204})_t$ 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.

Radioactive dating has been used to determine an age of carbonaceous chondrite meteorites of $t = 4.566 \pm 0.002$ billion years, where one billion years $= 10^9$ years = one Gigayear, abbreviated Gyr (Patterson 1956, Birck 1990). Radioactive dating of the oldest rocks returned from the Moon indicate an age of about 4.5 billion years (Wasserburg et al. 1977), and deep sediments in the Earth's oceans are dated at 4.55 billion years. Rounding off the numbers and allowing for possible systematic errors, we can say that the Earth, Moon and meteorites solidified at 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.

Despite Gamow's tunneling discovery, our understanding of radioactivity was still incomplete; there was something wrong with the way the beta rays were behaving. This resulted in the discovery of an entirely new particle – the electron neutrino.

7.4 The Electron and the Neutrino

When first discovered, the electrons emitted from radioactive elements were called beta rays, to distinguish them from alpha rays (helium nuclei) and gamma rays (high-energy radiation) that also are emitted during radioactive decay processes. From their measured charge and mass, it was discovered that the betas are not rays at all but instead ordinary electrons moving at nearly the speed of light. The emission of high-speed electrons by a radioactive element is known as beta decay.

Detailed measurements of the high-speed electrons, given off during radioactive decay, seemed to violate a fundamental principle of physics known as the conservation of energy. According to this rule, the total energy of a system must remain unchanged unless acted on by an outside force. We know of no process that disobeys this principle.

Nevertheless, the sum of the energy of the beta decay nucleus and the energy of the emitted electrons sometimes turned out to be less energy than the amount of energy lost by the initial nucleus. Careful measurements failed to turn up the missing energy, which seemed to have vanished into thin air, suggesting that energy might not be conserved during beta decay. The eminent Danish physicist Niels Bohr (1885–1962) even proposed that the conservation of energy law was being violated on the atomic scale, suggesting that the beta decay observations may force us to renounce the very idea of energy balance (Bohr 1930).

However, it turned out that a mysterious, invisible particle was spiriting away the missing energy. It was the elusive neutrino, whose existence was postulated by Wolfgang Pauli (1900–1958), a brilliant Austrian physicist. Pauli proposed a "desperate way out" of the energy crisis, speculating that an electrically neutral particle, produced at the same time as the electron, carried off the remaining energy (Pauli 1930, 1933). The sum of the energies remains constant during the beta decay, so the energy is balanced and the principle of conservation of energy is saved.

Pauli thought he had done "a terrible thing", for his desperate remedy postulated an invisible particle that could not be detected. Dubbed the *neutrino*, or "little neutral one" by the Italian physicist Enrico Fermi (1901–1954), the new particle could not be observed with the technology of the day, since the neutrino is electrically neutral, has almost no mass, and moves at nearly the speed of light. Therefore the hypothetical neutrinos were removing energy that would never be seen again. (Even in Pauli and Fermi's time, the observed high-energy shape of the emitted electron's energy spectrum indicated that the mass of the neutrino is either zero or very small relative to the mass of the electron).

Unlike light or any other form of radiation, neutrinos can move nearly unimpeded through any amount of material, even the entire universe. In the parlance of modern physics, neutrinos are characterized by a weak interaction with anything in the material world.

As beautifully described by Fermi, the decay process occurs when the neutron in a radioactive nucleus transforms into a proton with the simultaneous emission of an energetic electron and a high-speed neutrino (Fermi 1934). When left alone outside a nucleus, a neutron will, in fact, self-destruct in about 10 min into a proton, plus an electron to balance the charge, and a neutrino to help remove the energy.

We can write the radioactive beta decay reaction as

$$n \rightarrow p + e^{-} + \bar{\nu}_e, \tag{7.10}$$

where n denotes a neutron, p is a proton, e^- is an electron, and $\bar{\nu}_e$ denotes the anti-electron neutrino, or the anti-particle of the electron neutrino. The proton and electron have equal charges of opposite sign, so their charges cancel to match the uncharged neutron – provided that the neutrino has no charge. The electron e^- is sometimes denoted beta minus (β^-) to denote the emission of the beta particle, the electron. The reaction was also known as negative beta decay, with the negative standing for the negative charge of the negative beta particle, the electron.

The nuclear reaction for negative beta decay can be written:

$$(Z-1,\ A) \rightarrow (Z,\ A) + e^- + \bar{\nu}_e, \tag{7.11}$$

where (Z, A) denotes a nucleus of atomic number or charge Z and mass number A.

If a positron, denoted e^+, is emitted, the reaction is written:

$$\text{Energy} + p \rightarrow n + e^+ + \nu_e, \tag{7.12}$$

where energy has been supplied to fuel the reaction and ν_e denotes the electron neutrino. Sometimes the positron in this reaction is denoted beta plus (β^+). Unlike β^- decay, the positive beta decay is not a spontaneous reaction that occurs in isolation because it requires energy, the mass of the neutron being slightly greater than the mass of the proton. The nuclear reaction for positive beta decay is

$$\text{Energy} + (Z+1,\ A) \rightarrow (Z,\ A) + e^+ + \nu_e. \tag{7.13}$$

As far as anyone could tell, an atomic nucleus consists only of neutrons and protons, so the electron and neutrino seemed to come out of nowhere. They do not reside within the nucleus and are created at the time of nuclear transformation. No one knew exactly how the neutrinos formed.

How do you observe something that spontaneously appears out of nowhere and is so close to being nothing at all? Calculations suggested that the probability of a neutrino interacting with matter, so that the effect might be seen, is so incredibly small that no one could ever detect it. To see one neutrino, we would have to produce enormous numbers at about the same time, and build a massive detector to increase the chances of detecting it. Although almost all of the neutrinos still would pass through any amount of matter unhindered and undetected, a rare collision with other subatomic particles might leave a trace.

Nuclear reactors produce large numbers of neutrinos, and if a massive detector is placed near a large nuclear reactor, with appropriate shielding from extraneous signals, the telltale sign of the hypothetical neutrino may be barely observed.

The existence of the neutrino was finally proven by Project Poltergeist, an experiment designed by Clyde L. Cowan (1919–1974) and Frederick Reines (1918–1998) of the Los Alamos National Laboratory in New Mexico. They placed a 10-ton (10,000 L) tank of water next to a powerful nuclear reactor engaged in making plutonium for use in nuclear weapons. After shielding the neutrino trap underground and running it for about 100 days, Reines and Cowan detected a few synchronized flashes of gamma radiation that signaled the interaction of a few

neutrinos with the nuclear protons in water (Reines and Cowan 1953, Cowan et al. 1956).

The neutrinos themselves were not observed, and they never have been. Their presence was inferred by an exceedingly rare interaction. One of every billion billion, or 10^{18}, neutrinos that passed through the water tank hit a proton, producing the telltale burst of radiation. Nearly four decades later, Reines received the 1995 Nobel Prize in Physics for the detection of the neutrino; but by that time however, Cowan had died and therefore could not share in the award.

As discussed in Chap. 8, the Sun emits copious amounts of neutrinos. Every second, trillions upon trillions of neutrinos that were produced inside the Sun pass right through the Earth without even noticing that it is there. The indestructible neutrinos interact so rarely with the material world that almost nothing ever happens to them. Billions of ghostly neutrinos from the Sun are passing right through us every second, even in our bedrooms at night, and they did not come through the door. The solar neutrinos travel right through the Earth, a building, and us, without our body noticing them or them noticing our body.

Moreover, when a minute number of the Sun's neutrinos were snared in massive underground detectors, fewer than expected were observed; this eventually led to a new understanding of the neutrino. For the time being, however, let us move on to energetic cosmic rays that are always entering the atmosphere from outer space.

7.5 Cosmic Rays

Subatomic particles are entering our atmosphere from all directions in interstellar space and moving at nearly the speed of light. The perpetual high-energy rain was discovered about a century ago, when the Austrian physicist Victor Franz Hess (1883–1964), an ardent amateur balloonist, measured the amount of ionization at different heights within our atmosphere (Hess 1912).

It was already known that radioactive rocks at the Earth's surface were emitting energetic "rays" – the alpha and beta particles – that ionize molecules in the atmosphere near the ground. It was expected that the ionizing rays would be absorbed completely after passing through sufficient quantities of the atmosphere. The measured ionization at first decreased with altitude, as expected from atmospheric absorption of energetic particles emitted by radioactive rocks. However, the ionization rate measured by Hess increased at even higher altitudes to levels exceeding that at the ground when the balloons rose to above 1 km in altitude (Hess 1912). This meant that some penetrating source of ionization came from beyond the Earth. By flying his balloons at night and during a solar eclipse, when the high-altitude signals persisted, Hess showed that they could not come from the Sun but rather from some other source.

The American physicist Robert A. Millikan (1868–1953) subsequently used high-altitude balloon measurements to confirm that the "radiation" comes from

beyond the terrestrial atmosphere, and he gave it the present name of *cosmic rays* (Millikan 1926; Millikan and Cameron 1926). Millikan believed the cosmic rays were gamma rays associated with the synthesis of heavy elements deep in space, the "birth cries" of new matter. We now know that cosmic rays are energetic charged particles, not radiation, and more likely the "death cries" of massive exploding stars.

Global measurements showed that cosmic rays are electrically charged. During an ocean voyage in 1927 and 1928, the Dutch physicist Jacob Clay (1882–1955), for example, found lower cosmic-ray intensity near the Earth's Equator than at higher terrestrial latitudes; his results were confirmed and extended between 1930 and 1933 by Arthur H. Compton (1892–1962) of the University of Chicago (Clay 1932; Compton 1932). Compton conclusively demonstrated an increase in cosmic-ray intensity with terrestrial latitude, and also made measurements at mountain altitudes, where the increase with latitude was even stronger. His results indicated that cosmic rays must be electrically charged particles deflected by the Earth's magnetic field toward its magnetic poles, which are close to the geographic ones.

The amount of cosmic rays entering the Earth's atmosphere varies with the amount of solar activity over an 11 year cycle; the amount of cosmic rays increase when the solar activity decreases and vice versa. This inverse correlation has been named the Forbush effect, after Scott Forbush (1904–1984) who discovered it (Forbush 1950). It has been attributed to an enhancement in the interplanetary magnetic field originating from the Sun at times of increased solar activity (Davis 1955; Meyer et al. 1956); these magnetic fields divert the cosmic rays and keep them from encountering the Earth.

The charged cosmic-ray particles also are deflected and change direction during encounters with the interstellar magnetic field that winds its way among the stars. Therefore, we cannot look back along their incoming path and tell where cosmic rays originate; the direction of arrival shows only where they last changed course. The favored hypothesis, proposed by Walter Baade (1893–1960) and Fritz Zwicky (1898–1974), is that cosmic rays are accelerated to their tremendous energy during the supernova explosion of massive stars that have run out of thermonuclear fuel (Baade and Zwicky 1934a, b, c, d). Kulsrud et al. (1972) and Blandford and Ostriker (1980) have discussed the acceleration of cosmic rays in supernovae.

By the late 1940s, instruments carried by high-altitude balloons established that the most abundant cosmic ray particles arriving in the Earth's upper atmosphere are protons – the nuclei of former hydrogen atoms – and the second most abundant particles are helium nuclei – the alpha particles. Cosmic-ray electrons arriving near the top of the atmosphere took longer to discover (Earl 1961), mainly because they are far less abundant than the cosmic-ray protons at a given energy.

About 79 % of the cosmic rays arriving at the top of the Earth's atmosphere are protons, about 14.7 % are nuclei of helium, and roughly 1 % is carbon and oxygen nuclei (Table 7.2).

Although they are relatively few in number, cosmic rays contain phenomenal amounts of energy. That energy usually is measured in units of electron volts, abbreviated eV – for conversion use 1 eV $= 1.602 \times 10^{-19}$ J. The greatest flux of

Table 7.2 Average fluxes of primary cosmic rays at the top of the atmosphere[a]

Type of nucleus	Flux (particles m^{-2} s^{-1})
Hydrogen (protons)	640
Helium (alpha particles)	94
Carbon, oxygen	6

[a] The flux is in units of nuclei per square meter per second for particles with energies greater than 1.5 billion (1.5×10^9) electron volts per nucleon, denoted 1.5 GeV per nucleon, arriving at the top of the atmosphere from directions within 30° of the vertical

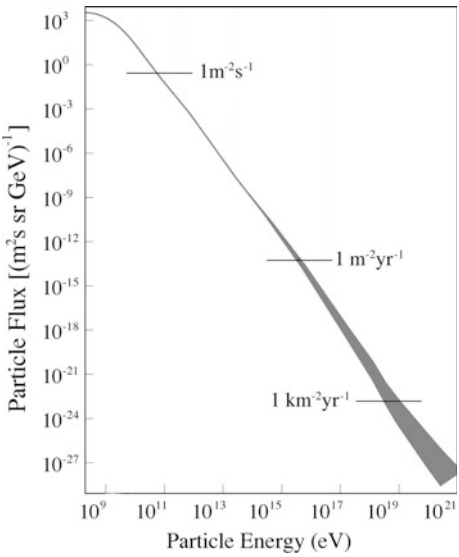


Fig. 7.2 Flux of cosmic rays The energy spectrum of cosmic-ray particles striking the outer atmosphere of the Earth. The particle flux is plotted as a function of the particle energy in units of electron volts, abbreviated eV, where $1\ \text{eV} = 1.602 \times 10^{-19}$ J and $1\ \text{GeV} = 10^9$ eV, or 1 billion eV. The most abundant cosmic-ray particles are protons with energies of about 1.5×10^9 eV. Every second about 640 enter every square meter of the Earth's outer atmosphere. They probably are accelerated to high energy during the supernova explosions of massive stars. One cosmic-ray proton of 10^{10} eV in energy enters each square meter of the Earth's outer atmosphere every second. The more energetic cosmic ray particles of 10^{14} eV are less abundant, with one per square meter every year. Solar flares can emit protons with energies of 10^{10} eV or less, and these solar energetic particles can strike the Earth when the solar active region is on the near side of the Sun. Cosmic rays with low flux and very high energy, greater than 1 million billion eV, or 10^{15} eV, may be of extragalactic origin. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2012. Reprinted with permission.)

cosmic-ray protons arriving at the Earth have energies of 1–10 GeV, or at 1 billion (10^9) to 10 billion (10^{10}) electron volts of energy. By way of comparison, a helium nucleus, or alpha particle, emitted during radioactive decay reaches no more than a million eV in energy, or a thousand times less than that of a cosmic ray proton.

Table 7.3 Particle speeds at different particle energies, expressed as fractions of the speed of light, c[a]

Particle kinetic energy (keV)	Electron speed (times c)	Proton speed (times c)
1 keV	0.063	0.0015
1,000 keV = 1 MeV	0.94	0.046
100,000 keV = 100 MeV	0.999987	0.43
1,000,000,000 keV = 1 GeV	0.99999987	0.88

[a] An energy of one kilo-electron volt is 1 keV = 1.6022×10^{-16} J, and the speed of light $c = 2.99792458 \times 10^{8}$ m s^{-1}

Cosmic rays do not all have the same energy, and some reach an energy of 10^{20} eV, more than 10 billion (10^{10}) times the abundant ones; Kotera and Olinto (2011) have reviewed the astrophysics of ultra high-energy cosmic rays. The flux of the most common, lower-energy cosmic rays at 10^{9}–10^{10} eV is greatest (Fig. 7.2).

Even at an energy of 10^{9} eV, a cosmic-ray proton must be traveling at 88 % of the speed of light (Table 7.3).

Cosmic rays enter the atmosphere with such great energies that they act like colossal atom destroyers, hitting molecules and their component atoms in the upper atmosphere and producing showers of subatomic debris, known as secondary cosmic-ray particles. This eventually led to the first observation of an energetic particle that does not belong to the atom. But first a method needed to be developed to detect the then-unknown particle.

Subatomic particles coming down through the atmosphere are detected near or at the ground by tracks in a cloud chamber, which creates a cloudy mist that precipitates as long thin bands of fog, along the trajectory of the particles. This is somewhat similar to the white vapor trails of jet aircraft, which record an airplane's movement in the sky. Fine water droplets condense from the jet exhaust fumes and create the elongated clouds.

The first cloud chamber, invented by Charles Thomas Rees Wilson (1869–1859), was very simple, consisting of a metallic cylinder with a glass cover and a piston that could be moved up and down from below, permitting air filled with water vapor to enter the space above it (Wilson 1911). When the piston was lowered quickly, the sudden expansion cooled the gas so that a mist formed in the chamber, like the foggy mist found high in the mountains. The water vapor in the chamber condensed or precipitated out on any ions present, making the ionized tracks of cosmic rays visible and showing where they had moved.

When the cloud chamber is placed between the poles of a strong magnet, the magnetic field exerts a force on any charged particle entering the chamber, which produces a curved particle track. When Carl Anderson (1945–2004), who was Millikan's student at the California Institute of Technology, built such a device, using a powerful electromagnet, he found in 1932 that a few of the cosmic ray showers produce two similar curved trajectories in opposite directions (Fig. 7.3). Further experiments revealed that an electron, which has a negative charge, was producing one of the curved tracks, whereas a particle with about the same mass as

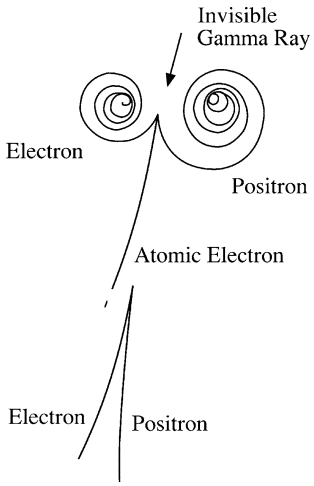


Fig. 7.3 The electron and the positron An invisible gamma-ray photon (*top*) produces an *electron* and a *positron* (short for positive electron), seen by *curved tracks* in a bubble chamber. Both the electron and the positron are bent into circular tracks by the instrument's magnetic field, moving in opposite directions because of their opposite electrical charge and spiraling into a smaller circular motion as they lose energy. In this *upper pair*, some of the photon's energy is taken up in displacing an atomic electron, which shoots off toward the *bottom left*. In the lower example, all of a gamma ray's energy goes into the production of the electron−positron pair. As a result, these particles are more energetic than the *upper pair*, and their tracks do not curve so tightly in the chamber's magnetic field. (Schematic of a Lawrence Berkeley Laboratory bubble-chamber image, reproduced by Frank Close, Michael Marten, and Christine Sutton in *The Particle Explosion*, New York: Oxford University Press 1987. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

the electron and a positive charge of the same amount but opposite sign as the electron was producing the track curved in the opposite direction. Anderson had discovered the positron, short for "positive electron" and the anti-matter counterpart of the electron (Anderson 1932a, 1933). The energetic cosmic rays had produced a new type of particle that had never been seen before!

Anderson received the Nobel Prize in Physics in 1936 for his discovery of the positron, sharing the award with Hess for his discovery of cosmic rays. Wilson had already received the recognition 9 years earlier, for his method of making the paths of electrically charged particles, the cosmic rays, visible by condensation of vapor.

As it turned out, Paul Adrien Maurice Dirac (1902–1984), then at Cambridge University, had predicted the existence of anti-matter (Dirac 1928). For Dirac, mathematical beauty was the most important aspect of any physical law describing nature. He noticed that equations that describe the electron have two solutions. Only one was needed to characterize the electron; the other solution specified a sort of mirror image of the electron – that is, an anti-particle, now called the

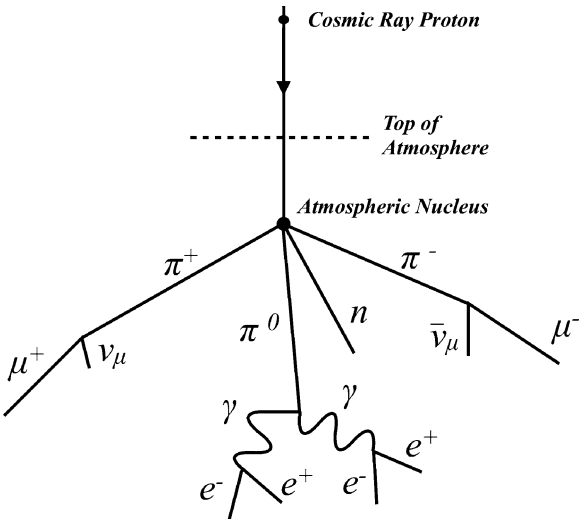


Fig. 7.4 Cosmic ray shower When a primary cosmic ray enters the Earth's atmosphere and collides with the nucleus of an atom in an atmospheric molecule, it can produce a shower of secondary subatomic particles. Here we show the most abundant cosmic ray particle, the proton, producing a neutron, designated n; pions denoted by π, muons denoted by μ, with the anti-muon-neutrino, $\bar{\nu}_\mu$, gamma rays, electrons e^- and positrons e^+

positron. At the time of his discovery, Anderson nevertheless was unaware of Dirac's theoretical prediction of the positron.

The nuclear reaction that describes the creation of a positron is known as the inverse beta decay or positive beta decay, and it occurs when the energy of a colliding cosmic-ray particle is used to convert a proton, p, into a neutron, n. The reaction, written

$$\text{Energy} + p \rightarrow n + e^+ + \nu_e, \tag{7.14}$$

includes the production of the positron, denoted e^+ or sometimes β^+, which carries away the charge of the proton, leaving a neutron with a slight increase in mass, and the electron neutrino, designated ν_e, to balance the energy books on the two sides of the reaction.

Once created, anti-matter does not stay around for very long, for any anti-matter will promptly self-destruct when it encounters ordinary matter. When an electron and positron meet, they annihilate one another and disappear in a puff of energetic radiation. The electron–positron annihilation reaction is:

$$e^- + e^+ \rightarrow \gamma + \gamma, \tag{7.15}$$

where e^- is the electron, and γ denotes a gamma-ray photon, each with energy of 0.511 MeV equal to $m_e c^2$, which corresponds to the complete destruction of an electron of rest mass m_e where c is the speed of light. As discussed subsequently, this reaction helps to produce radiation in the core of the Sun, and also is observed during explosive flares on the visible solar disk.

When examining the cloud-chamber tracks of secondary particles produced by cosmic rays entering our atmosphere, scientists discovered other previously unknown subatomic particles (Fig. 7.4). There is the muon, denoted μ (Anderson

Table 7.4 Nobel Prizes related to experimental investigations of subatomic matter

Years	Person	Accomplishment
1927	C. T. R. Wilson (Scottish physicist)	For his method of making the paths of electrically charged particles visible by condensation of vapor
1936	Victor Franz Hess (Austrian physicist)	For his discovery of cosmic radiation [now called cosmic rays]
	Carl Anderson (American physicist)	For his discovery of the positron
1938	Enrico Fermi (Italian physicist)	For his demonstrations of the existence of new radioactive elements produced by neutron irradiation, and for his related discovery of nuclear reactions brought about by slow neutrons
1939	Ernest Lawrence (American physicist)	For the invention and development of the cyclotron and for results obtained with it, especially with regard to artificial radioactive elements
1948	Patrick M. S. Blackett (English physicist)	For his development of the Wilson cloud chamber method, and his discoveries therewith in the fields of nuclear physics and cosmic radiation
1950	Cecil Powell (English physicist)	For his development of the photographic method of studying nuclear processes and his discoveries regarding mesons made with this method
1951	John Cockcroft (English physicist)	For their pioneering transmutation of atomic nuclei by artificially accelerated atomic particles
	Ernest T. S. Walton (English physicist)	
1959	Emilio Segrè (Italian-born American physicist)	For their discovery of the antiproton
	Owen Chamberlain (American physicist)	
1995	Frederick Reines (American physicist)	For the detection of the neutrino
2002	Raymond Davis Jr. (American astrophysicist)	For their pioneering contributions to astrophysics, in particular for the detection of cosmic neutrinos
	Masatoshi Koshiba (Japanese astrophysicist)	

and Neddermeyer 1937; Neddermeyer and Anderson 1937) and the pion, designated π, discovered using photographic emulsions to detect secondary cosmic ray particles at high altitudes, including the Pic Du Midi astronomical observatory (Lattes et al. 1947). The muon has a mass between that of the electron and the positron. The pion also has an intermediate mass. Hideki Yukawa (1907–1981) had predicted the existence of the pion more than a decade before its discovery, in his theory for the nuclear force (Yukawa 1935, 1937). Altogether, the investigations of x-rays, radioactivity, and cosmic rays resulted in quite a lot of recognition by the Nobel Prize in Physics (Table 7.4).

Atomic nuclei from outer space hit the upper atmosphere and produce a debris of pions. These soon decay into muons, which are always raining down to the

Earth. Many of the muons decay into electrons, positrons, and neutrinos during their flight through the atmosphere, and the very energetic muon neutrinos penetrate deep underground.

In 1998 a Japanese group reported that observations of muon neutrinos, generated by cosmic rays interacting with the atmosphere, indicate that neutrinos change type or flavor, oscillating between types as they travel through matter (Fukada et al. 1998a). There were roughly twice as many muon neutrinos coming from the atmosphere directly over their detector than those coming from the other side of the Earth. The muon neutrinos are produced in the atmosphere above every place on our planet, but some of them apparently disappeared while traveling through the Earth to arrive at the detector from below (Sect. 8.4).

In 2011, the *Alpha Magnetic Spectrometer* was carried by *Space Shuttle* to the *International Space Center*, where it will search for evidence of dark matter and anti-matter by measuring cosmic rays above the Earth's atmosphere.

During the first half of the 20th century, investigations of cosmic rays revealed many unexpected aspects of subatomic particles, and this has inspired the construction of particle accelerators, which have been used to discover many other subatomic particles.

7.6 Nuclear Transformation by Bombardment

What happens if we turn radioactivity around and instead of watching the nucleus of an unstable heavy atom decay, we bombard a perfectly normal, lighter nucleus with very energetic particles? Perhaps this normally stable element could be transformed artificially on the Earth through such a nuclear bombardment. After all, that is what the cosmic rays were doing in the atmosphere, resulting in all kinds of surprises, turning some atoms into previously unknown, fundamental particles.

Medieval alchemists had been trying to change one element into another, like lead into gold, for centuries, but they always failed in their attempts because the chemical and thermal reactions they employed were nowhere near energetic enough to crack open the nucleus of an atom.

The first successful attempts to transform elements in the terrestrial laboratory occurred when Patrick M. S. Blackett (1897–1974), a recent graduate of Cambridge University, directed a beam of fast alpha particles, ejected by radioactive decay, into Wilson's cloud chamber. The chamber was filled with normal atmospheric air, which is composed mainly of nitrogen molecules. Blackett improved the cloud chamber so that the air expanded and cooled automatically, and he took automatic photographs of the alpha-particle tracks. Most of them passed straight through the chamber. However, after more than 23,000 photographs of alpha particles bombarding nitrogen in the cloud chamber, during a three-year period from 1921 to 1924, Blackett finally succeeded in recording just eight head-on collisions of alpha particles with the nuclei of nitrogen atoms (Blackett 1925).

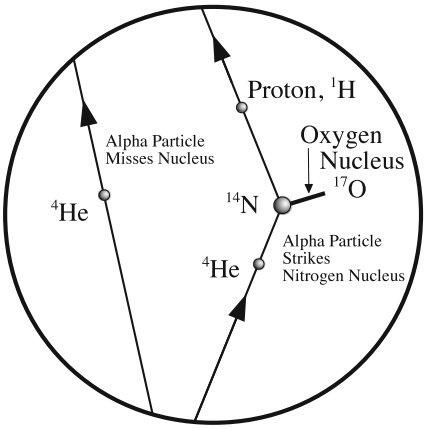


Fig. 7.5 Nuclear transformation When an alpha particle, or helium nucleus denoted ^{4}He, is sent through a cloud chamber, it usually passes right through it, with a trajectory that marks out a straight line. Occasionally, the alpha particle will strike the nucleus, ^{14}N, of a nitrogen atom in the air within the chamber, transforming it into the nucleus, ^{17}O, of an oxygen atom with the emission of a proton, the nucleus of a hydrogen atom and denoted ^{1}H. Such a nuclear transformation was first observed in cloud chamber photographs taken in the early 1920s by Blackett (1897–1974). (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

On each photograph, the track of an alpha particle suddenly stopped, being replaced with the fine, straight track of an ejected proton and the short, stubby recoil track of the struck nucleus (Fig. 7.5). However, there was no sign of the recoiling alpha particle. The collision had brought the alpha particle into the nitrogen nucleus, forming a nucleus of a form of oxygen. The reaction can be written as follows:

$$^{4}\mathrm{He} + {}^{14}\mathrm{N} \rightarrow {}^{17}\mathrm{O} + {}^{1}\mathrm{H}, \tag{7.16}$$

where the collision of an alpha particle, or helium nucleus ^{4}He, with a nitrogen nucleus, ^{14}N, gave rise to the nucleus of oxygen, ^{17}O, and a proton, the nucleus of hydrogen, ^{1}H. The old alchemist's dream finally had been realized in a laboratory on the Earth, in which nuclear transformation had been induced and recorded.

Enthusiastic scientists directed beams of alpha particles into many other elements, creating nuclear transformations similar to the one observed for nitrogen. When it came to heavier elements, with greater nuclear charge, however, a nuclear transformation could not be produced. The greater electrical repulsion of nuclei with atomic number Z greater than 18 always withstood the bombardment by alpha particles.

That is when Gamow's paper on the decay of heavy radioactive nuclei had a decisive role (Gamow 1928). His calculations indicated that on rare occasions alpha particles could tunnel through the positively charged wall of a nucleus, but

that fast protons more easily overcome the barrier than slow, heavier nuclei. Because it has a smaller electrical charge, a proton suffers less nuclear repulsion when approaching a given nucleus, and therefore has a greater probability of penetrating it. Moreover, because the proton is four times less massive than an alpha particle, it might be easier to accelerate it to high speed.

Rutherford's student John Cockcroft (1897–1967) had studied electrical engineering, and used his background to build a machine that would accelerate the hydrogen nuclei, the protons, in very intense electrical fields. When applying 500,000 V, Cockcroft was able to produce a parallel beam of protons traveling at the speed of 10^7 m s^{-1} or 1/30th the speed of light.

When bombarding lithium with high-energy protons, Cockcroft and his colleague Ernest T. S. Walton (1903–1995) turned the lithium nucleus into two alpha particles by the following nuclear reaction (Cockcroft and Walton 1932a):

$$^{1}\mathrm{H} + {}^{7}\mathrm{Li} \rightarrow {}^{4}\mathrm{He} + {}^{4}\mathrm{He}, \tag{7.17}$$

where ^{1}H denotes the proton, the nucleus of a hydrogen atom, ^{7}Li designates the lithium nucleus, and ^{4}He denotes an alpha particle, the nucleus of the helium atom.

At about the same time, Ernest Lawrence (1901–1958), located at Berkeley University, invented the cyclotron, which used magnets to bend the path of a charged particle into a circular orbit that passed across an alternating and accelerating voltage. As the radius of the spiraling orbit increased, so did the particle's speed; therefore, the time to complete each orbit remained constant, and the particles were repeatedly accelerated in equal time intervals before directing them into a collision with something else. This is similar to pushing a child in a swing at the same part of its swinging motion, pumping it up to greater and greater speed.

Lawrence built increasingly larger cyclotrons at Berkeley's Radiation Laboratory, whirling the protons and other particles to faster and faster speeds. A cyclotron of just 0.69 m in diameter was able to accelerate protons, denoted ^{1}H to 5 MeV in energy (Lawrence and Livingston 1932, 1934; Lawrence et al. 1932), and a 1.5 m cyclotron was used to accelerate deuterons, the nuclei of heavy hydrogen designated ^{2}H, to an energy of 16 MeV (Lawrence and Cooksey 1936). His final cyclotron, with a diameter of nearly 5 m, could accelerate the deuterium nuclei of heavy hydrogen to energies of 195 MeV.

Deuteron bombardment of various elements, including beryllium, resulted in the creation of numerous neutrons and artificial radioactive isotopes. In 1939, the Nobel Prize in Physics was awarded to Lawrence for the invention and development of the cyclotron and for results obtained with it, especially with regard to artificial radioactive elements. It wasn't until 1951 that John Cockcroft and Ernest Walton received the prize for their pioneer work on the transmutation of atomic nuclei by artificially accelerated atomic particles.

These early accomplishments stimulated the construction of increasingly powerful particle accelerators that accelerated particles to more and more energy. Eventually, accelerators were built that reached cosmic-ray energies, and this resulted in the discovery of new, previously unknown particles, such as the tau

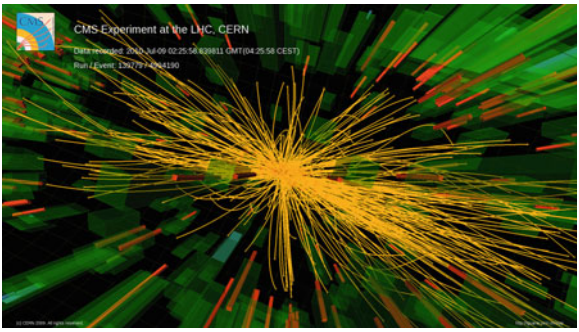


Fig. 7.6 Proton-proton collision Two beams of protons have been whirled in opposite directions to nearly the speed of light, each with an energy of 7 trillion electron volts, or 7 TeV = 7×10^{12} eV = 1.12×10^{-6} J, and directed into collision with one another at CERN's Large Hadron Collider (LHC). This image displays the tracks of more than 100 charged particles as they fly away from the point of proton collision. Experiments with this instrument have provided evidence that is consistent with a new, previously unseen particle, named the *Higgs boson*, in the mass-energy range of 10^{11} eV. By studying the collision particle debris, including correlations among them, scientists hope to gain an improved knowledge of how subatomic particles interact at extremely high energies, including the hot, dense conditions only a small fraction of a second after the "big bang". At the point of proton impact, temperatures of more than 1 million, or 10^{12}, K are generated, exceeding 100,000 times the temperature at the center of the Sun. In particle physics, a hadron is a composite particle made of quarks held together by the strong force; the best-known hadrons are the protons and neutrons, which are components of atomic nuclei. CERN is a French acronym for the Conseil Européen pour la Recherche Nucléaire (the European Organization for Nuclear Research). The Compact Muon Solenoid (CMS) particle detector created this image. (Courtesy of CERN.)

lepton found as the result of particle collisions using the Stanford Linear Particle Accelerator (Perl et al. 1975, 1976).

Eventually, the 2 km-diameter Tevatron, an accelerator at the Fermi National Accelerator Laboratory, abbreviated Fermi lab, used thousands of electromagnets to whirl protons up to 1,000 GeV, or 1 TeV, of energy. By 2009 the Large Hadron Collider at the European Organization for Nuclear Research (known as CERN, an acronym for Conseil Européan pour la Recherche Nucléaire) was using superconducting magnets in a circular tunnel 27 km in circumference to produce two beams of protons moving in opposite directions, and eventually directed into collision with one another (Fig. 7.6). It has provided evidence consistent with another previously unknown particle, the Higgs boson.

In the meantime, on the eve of World War II (1939–1945), the German radiochemist Otto Hahn (1879–1968), who had been working with Fritz Strassmann (1902–1980) and Lise Meitner (1878–1968), showed that when uranium is bombarded with neutrons it could be split into two nearly equal fragments (Hahn and Strassmann 1939a, b). The process is known as nuclear fission, analogous to binary fission in the biological sciences. This was altogether different from the proton bombardment of much lighter nuclei, for the heavy uranium was

broken in two and released enormous amounts of energy in the process. Moreover, it also freed additional neutrons, which in turn can trigger the fission of neighboring nuclei, resulting in a runaway chain reaction if not properly controlled. When controlled, a nuclear chain reaction can be used to generate electricity in nuclear reactors; when uncontrolled, it has applications in nuclear weapons. Hahn received the 1944 Nobel Prize in Chemistry for the discovery of nuclear fission, but some historians believe that Meitner should have shared the award.

Meitner, who was in exile in Copenhagen, confirmed the fission of uranium by neutron bombardment (Meitner and Frish 1939), and the Danish physicist Niels Bohr described it to Albert Einstein (1875–1955), who had immigrated to the United States. By the time Einstein heard about uranium fission, World War II (1939–1945) had begun, and scientists feared that Nazi Germany would use the discovery to build an atomic bomb to conquer the world. In 1939 and 1940 Einstein wrote to the President of the United States, Franklin Roosevelt (1882–1945), encouraging a program that would achieve a nuclear chain reaction and to consider the development of "extremely powerful bombs" that the Germans also might be constructing. The concern was real, for it was later discovered that prominent German physicists, including Werner Heisenberg (1901–1976) and Carl von Weizsäcker (1912–2007), helped the Germans investigate the feasibility of constructing nuclear weapons during World War II.

In 1942, the famed American physicist J. Robert Oppenheimer (1904–1967) invited a small group of theoretical physicists to the University of California at Berkeley to discuss how an atomic bomb might be assembled. Within a year, they had all moved to Los Alamos, the secret laboratory in New Mexico where thousands of scientists, technicians, and military personnel worked under Oppenheimer's enthusiastic direction to create the first atomic bomb.

Many of the best scientific minds in the country were involved, including Hans Bethe (1906–2005), head of the Los Alamos theoretical division; Richard Feynman (1918–1988), who worked on numerical calculations of bomb performance; and the Italian immigrant Enrico Fermi (1901–1954), who helped produce the first self-sustaining nuclear chain reaction. Another famous physicist, Philip Morrison (1915–2005), accompanied the first bombs all the way to their final flights, caring for them before they were dropped on Japan in 1945.

Some of the same scientists who developed the atomic bomb also showed how nuclear reactions deep inside the Sun and other stars makes them shine, while also producing most of the elements heavier than helium that are now found in the universe.

Chapter 8
What Makes the Sun Shine?

8.1 Can Gravitational Contraction Supply the Sun's Luminosity?

When we measure the total amount of sunlight that illuminates and warms our globe, and extrapolate back to the Sun, we find that it is emitting an enormous power of 385.4 million, million, million, million, or 3.828×10^{26}, watts, where one watt $= 1\ \mathrm{J\ s^{-1}}$. This brilliance is far too great to be perpetually sustained, and we therefore wonder what heats the Sun, and how long that heat will last.

In the mid-nineteenth century, the German physicist Hermann von Helmholtz (1821–1894) proposed that the Sun's luminous energy is due to its gravitational contraction (Helmholtz 1856, 1908). If the Sun were gradually shrinking, the compressed matter would become hotter and the solar gases would be heated to incandescence; in more scientific terms, the Sun's gravitational energy would be converted slowly into the kinetic energy of motion and heat up the Sun, so that it would continue to radiate. This follows from the principle of conservation of energy, which Helmholtz was one of the first to propose (Helmholtz 1847). It states that energy can be neither created nor destroyed; it can only change form.

The Irish physicist William Thomson (1824–1907), later Lord Kelvin, subsequently showed that the Sun could have illuminated the Earth at its present rate for about 100 million years by slowly contracting (Kelvin 1862, 1899; Burchfield 1975). We can follow his reasoning by noting that the gravitational potential energy, Ω, of a star of mass M_S and radius R_S is given by (Sect. 3.2):

$$\Omega = -\frac{3GM_S^2}{5R_S}, \tag{8.1}$$

where the gravitational constant $G = 6.674 \times 10^{-11}\ \mathrm{N\ m^2\ kg^{-2}}$. The change, $\Delta\Omega$, in gravitational potential energy created by a decrease in radius, ΔR_S, is:

$$\Delta\Omega = \frac{3GM_S^2}{5R_s^2}\Delta R_S \approx 3.27 \times 10^{32}\left(\frac{M_S}{M_\odot}\right)^2\left(\frac{R_\odot}{R_s}\right)^2 \Delta R_s\ \mathrm{J}, \tag{8.2}$$

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_8, © Springer-Verlag Berlin Heidelberg 2013

where the radius decrease is in meters, and we have normalized the mass and radius in terms of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg and the Sun's radius $R_\odot = 6.955 \times 10^8$ m.

If the energy change provides an absolute luminosity, L_S, in a time interval, Δt, then $L_S = \Delta\Omega/\Delta t$, and the rate of change in radius is:

$$\frac{\Delta R_S}{\Delta t} = \frac{5 L_S R_S^2}{3 G M_S^2} \approx 1.17 \times 10^{-6} \left(\frac{L_S}{L_\odot}\right) \left(\frac{M_\odot}{M_S}\right)^2 \left(\frac{R_S}{R_\odot}\right)^2 \text{m s}^{-1}, \tag{8.3}$$

where the Sun's absolute luminosity is $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. Since one year is 3.156×10^7 s, this shows that a contraction of only 36.9 m per year will power the Sun at the present rate. That is a very small change considering the much larger radius of the Sun.

The problem with this mechanism is the long duration of the Sun and other stars. If the source of a star's present luminosity were gravitational potential energy, then the current radius would shrink to zero in the Kelvin–Helmholtz time, denoted by the symbol τ_{K-H}, given by Kelvin (1863):

$$\tau_{K-H} = \frac{R_S}{(\Delta R_S/\Delta t)} = \frac{\Omega}{L_S} = \frac{3 G M_S^2}{5 R_S L_S} = 5.95 \times 10^{14} \left(\frac{M_S}{M_\odot}\right)^2 \left(\frac{R_\odot}{R_S}\right) \left(\frac{L_\odot}{L{-}_S}\right) \text{s}. \tag{8.4}$$

Using 1 year $= 3.156 \times 10^7$ s, the Kelvin–Helmholtz time for the Sun is about 1.89×10^7 years. A contraction of 36.9 m per year will power the Sun at its present rate by converting gravitational potential energy into heat. However, if the Sun continues to shine this way, it will shrink down to practically nothing and vanish from sight in 18.9 million years. The astonishing thing, which was not realized at the time Lord Kelvin wrote his articles, was the Sun's durability. It has lasted much longer than he envisioned.

The problem is much worse for a giant star that has both a larger radius and a greater luminosity. As shown by the British astronomer Arthur Stanley Eddington (1882–1944), gravitational contraction can only keep a giant star shining for no more than 100,000 years, and he therefore proposed that energy that is locked up inside the atom was a likely alternative candidate for making the stars shine (Eddington 1920).

Kelvin had assumed the Sun would be about as old as the Earth, and calculated the age of the Earth under the assumption that it began in an initially molten state and cooled from the outside in. Using the equation of heat conduction with the known conductivity of rock, he calculated that it would take about 100 million years to reach the then observed temperature gradient between the hot lower levels of mines and the cold surface of the Earth (Kelvin 1863, 1899).

The discovery of radioactivity provided an entirely new perspective on our planet's internal heat and age. Radioactive elements could heat the planet from inside, emitting energetic particles that produced a rise in internal temperature (Rutherford 1905), so the Earth's hot interior is not a result of cooling from an earlier, hotter state. Radioactivity also clocked the Earth's age, by establishing the

relative amounts of radioactive parent elements and stable, non-radioactive daughters. When this ratio is combined with the known rates of radioactive decay, they indicated that the Earth is at least 2–3 billion years old (Boltwood 1907; Rutherford 1929). Modern measurements using this method establish an age of about 4.6 billion years for the Earth, and presumably for the Sun.

Moreover, in looking back at the Earth's history, we find that the Sun has been shining steadily and relentlessly for eons, with a brilliance that could not be substantially less than it is now. The radioactive clocks in rock fossils indicate, for example, that the Sun was hot enough to sustain primitive creatures on the Earth 3.4 billion years ago (Tice and Lowe 2004).

Even in the early 20th century, no one had any clue as to why the Sun, or any other star, could shine so brightly for billions of years. That understanding had to await the discovery of subatomic particles, and the realization that the Sun is composed mainly of hydrogen. Of equal importance was the fact that the center of the Sun is much hotter than an ordinary fire, enabling it to consume atomic nuclei.

8.2 How Hot is the Center of the Sun?

The most abundant atom in the Sun is hydrogen, with a single proton at its nuclear center and one electron outside of the nucleus. It is so hot within most of the Sun, except its cool outer atmosphere, that all of the protons and electrons have been liberated from their atomic bonds and move about unattached to one another.

Protons are 1,836 times more massive than electrons; therefore they dominate the gravitational effects inside the star. The temperature, $T_{C\odot}$, at the center of the Sun can be estimated by assuming that each proton down there is hot enough and moving fast enough to counteract the gravitational compression it experiences from the rest of the star. That is we can equate the thermal energy of the proton to the gravitational energy, expressed by the relation:

$$\frac{3}{2}kT_{C\odot} = \frac{Gm_P M_\odot}{R_\odot}, \tag{8.5}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}, the mass of the proton is $m_P = 1.6726 \times 10^{-27}$ kg, the mass of the Sun $M_\odot = 1.989 \times 10^{30}$ kg and the radius of the Sun $R_\odot = 6.955 \times 10^8$ m.

Solving for the central temperature of the Sun we obtain:

$$T_{C\odot} = \frac{2Gm_P M_\odot}{3kR_\odot} \approx 1.5 \times 10^7 \text{ K}. \tag{8.6}$$

So deep down inside the Sun, within its dense central core, the protons have a temperature of 15 million K. This and other physical parameters of the Sun are given in Table 8.1.

Example: Gas pressure and mass density at the center of the Sun
The gas pressure $P_{C\odot}$ at the center of the Sun, will be the force per unit area due to the gravitational force per unit area of the material above it. To a rough approximation:

$$P_{C\odot} \approx \frac{M_\odot}{R_\odot^2}\frac{GM_\odot}{R_\odot^2} \approx \frac{GM_\odot^2}{R_\odot^4} \approx 10^{15}\,\text{pascal}, \tag{8.7}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}, the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, and the Sun's radius $R_\odot = 6.955 \times 10^8$ m. A more exact calculation gives a value about ten times as large or $P_{C\odot} \approx 10^{16}$ pascal.

From the ideal gas law (Sect. 5.4):

$$P_{C\odot} = NkT_{C\odot} = \frac{\rho_{C\odot} kT_{C\odot}}{m_P}, \tag{8.8}$$

where N *i*s the number density of protons, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the $\rho_{C\odot}$ is the central mass density of the Sun, and the mass of the proton $m_P = 1.6726 \times 10^{-27}$ kg. Solving for the central mass density we obtain:

$$\rho_{C\odot} = \frac{P_{C\odot} m_P}{kT_{C\odot}} \approx 10^5\,\text{kg m}^{-3}, \tag{8.9}$$

using a central temperature of $T_{C\odot} \approx 1.5 \times 10^7$ K.

Such extreme central conditions were recognized more than a century ago, when Jonathan Homer Lane (1819–1880), an American astronomer and inventor working at the U.S. Patent Office, assumed that gas pressure supports the weight of the Sun (Lane 1870; Ritter 1898). Although no one knew about nuclear protons at the time, Lane's basic reasoning still applies. The hot protons move about with high speeds, frequently colliding with one another and creating the gas pressure that holds up the Sun. For the Sun, a central temperature of 15 million degrees K establishes equilibrium between the outward pressure of the moving protons and the inward gravitational pull at the Sun's center.

In any layer within the Sun, the weight of the overlying gas must be equal to the outward-pushing pressure; otherwise the Sun would expand or contract, which is not observed. At greater distances from the center, there is less overlying material to support and the compression, pressure, and temperatures are less, so the solar material becomes progressively thinner and cooler.

As one might suspect, a more massive star produces greater compression at its center, and a higher central temperature is required to hold it up. The central temperature, T_{CS}, of a star of mass, M_S. and radius, R_S, is given by:

Table 8.1 Physical properties of the Sun

$M_\odot$ = mass of Sun = 1.989×10^{30} kg
$R_\odot$ = radius of Sun = 6.955×10^8 m
$\rho_\odot$ = mean mass density of Sun = $3M_\odot/(4\pi R_\odot^3) \approx 1{,}409$ kg m^{-3}
$\rho_{c\odot}$ = central mass density of Sun = 151,300 kg m^{-3}
$T_{C\odot}$ = central temperature of Sun = $2Gm_pM_\odot/(3kR_\odot) \approx 1.5 \times 10^7$ K
$V_{esc\odot}$ = escape velocity from photosphere of Sun = $(2GM_\odot/R_\odot)^{1/2} \approx 6.177 \times 10^5$ m s^{-1}
$D_\odot$ = 1 AU = mean Earth–Sun distance = 1.4959787×10^{11} m $\approx 1.496 \times 10^{11}$ m
$\theta_\odot = R_\odot/D_\odot$ = angular radius of Sun = 959.63″ where 1″ = 1 s of arc
(At the Sun 1 s of arc = 1″ = 7.253×10^5 m)
$P_{r\odot}$ = sidereal rotation period of visible solar disk at the equator = 25.67 days
$V_\odot$ = rotation velocity of visible solar disk at the equator = 1,971 m s^{-1}
$f_\odot$ = solar constant = 1,361 J s^{-1}m^{-2}
$L_\odot$ = absolute luminosity of Sun = $4\pi f_\odot D_\odot^2 \approx 3.828 \times 10^{26}$ J s^{-1}
$T_\odot$ = effective temperature of visible solar disk = $[L_\odot/(4\pi\sigma R_\odot^2)]^{1/4} \approx 5{,}780$ K
$m_{v\odot}$ = apparent visual magnitude of the Sun = −26.74 mag
$m_{bol,\odot}$ = apparent bolometric magnitude of the Sun = −26.83 mag
$M_{v\odot}$ = absolute visual magnitude of the Sun = +4.83 mag
$M_{bol\odot}$ = absolute bolometric magnitude of Sun = +4.74 mag
$B_\odot$ = magnetic field strength at visible solar disk = 100–1,000 G = 0.01–0.1 T
X = mass fraction of hydrogen = 0.7154
Y = mass fraction of helium = 0.2703
Z = mass fraction of all other atoms = 0.0142
Age = 4.6×10^9 year

$$T_{CS} = T_{C\odot}\left(\frac{M_S R_\odot}{M_\odot R_S}\right). \tag{8.10}$$

It is the moving particles inside any star, including the Sun, which holds up the star. This motion, pushing, and pressure of the particles prevent a star from collapsing under its enormous weight. What keeps the particles down there hot, to sustain their rapid motion? It is nuclear-fusion reactions in the compact, dense core of a star that energizes the particles there, sustaining their heat and making them move rapidly. Once the nuclear reactions begin, the subatomic energy that is liberated keeps the nuclei sufficiently hot to ensure continuation of the reactions.

8.3 Nuclear Fusion Reactions in the Sun's Core

8.3.1 Mass Lost is Energy Gained

The only known method for keeping the Sun shining with its present luminosity for billions of years involves nuclear fusion reactions under the intense pressures and exceptionally high temperatures at great depths within the Sun. They are termed "nuclear" because it is the interaction of atomic nuclei that powers the

Sun. In nuclear-fusion reactions, two or more atomic nuclei fuse together to produce a heavier nucleus, releasing energy, subatomic particles, and radiation. For the Sun, it is protons, the nuclei of abundant hydrogen atoms that fuse together to make the nuclei of helium atoms, the next most abundant element in the Sun.

Energy can be derived only from energy, and the source of energy in nuclear fusion is mass loss. The basic idea was provided by Albert Einstein (1879–1955) in his *Special Theory of Relativity*, which included the famous formula $E = mc^2$ for the equivalence of mass, m, and energy, E (Einstein 1905a, b, 1906, 1907). Because the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, or about 300 million meters per second, is a very large number, only a small amount of mass is needed to produce a huge amount of energy.

Important evidence for the then unknown source of the Sun's energy was being obtained at the Cavendish Laboratory at Cambridge University in England about a decade after Einstein's seminal work. Here Ernest Rutherford (1871–1937) showed that the massive nuclei of all atoms are composed of hydrogen nuclei, which he named protons. At about the same time, the chemist Francis Aston (1877–1945), also working at the Cavendish, invented the mass spectrograph and used it to show that the mass of the helium nucleus is slightly less massive, by a mere 0.7 %, than the sum of the masses of the four hydrogen nuclei, or protons, that enter into it (Aston 1919, 1920).

While Rutherford and Aston were discovering the inner secrets of the atoms, the astronomer Arthur Stanley Eddington (1882–1944), director of the nearby Cambridge Observatory, was examining the internal workings of the Sun and other stars, and reasoned that the stars are the crucibles in which the heavier elements are made from lighter ones.

Realizing that such stellar alchemy would release energy, Eddington (1920) proposed that hydrogen is transformed into helium inside stars, with the resultant mass difference released as energy to power the Sun. The mass that is lost goes into energizing the Sun and other stars. The mass difference, Δm, is converted into energy, ΔE, to power the Sun, all in accordance with Einstein's equation $\Delta E = \Delta mc^2$. Eddington rightly concluded that this could supply the Sun's current luminous output for an estimated 15 billion years.

During the ensuing decade it was realized that the lightest known element, hydrogen, is the most abundant element in the Sun, so hydrogen nuclei, or protons, must play the dominant role in nuclear reactions within our daytime star. Physicists were nevertheless convinced that protons could not react with each other inside the Sun. Even in a high-speed, head-on collision at the enormous central temperature of the Sun, two protons did not have sufficient energy to overcome their mutual electrical repulsion and merge together. The thermal velocity of the average proton at a temperature of 15 million K was far too slow, and the central temperature had to be at least a thousand times hotter to raise the average kinetic energy of the protons above the electrical barrier (Focus 8.1).

Focus 8.1 The temperatures necessary for thermonuclear reactions
Protons are positively charged, and like charges repel one another with an electrostatic force given by Coulomb's law (Coulomb 1785), discovered by the French physicist Charles Augustin de Coulomb (1736–1806). This means that there is an electrified barrier that prevents protons from becoming too close to each other.

The electrical force, F, on a charge q_1 due to the presence of another charge q_2 separated from it by a distance, D, is given by Coulomb's law:

$$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{D^2} \approx 8.9875 \times 10^9 \frac{q_1 q_2}{D^2}, \tag{8.11}$$

where $\pi \approx 3.14159$ and the electric constant $\varepsilon_0 = 8.8542 \times 10^{-12}$ F m^{-1}. A positive force implies an electrical repulsion and a negative force implies an electrical attraction.

The repelling force is proportional to the square of the electrical charge of the protons, and it is inversely proportional to their separation. Therefore, the force of repulsion between two protons becomes increasingly larger as they are brought closer together. Stated another way, the protons do not move fast enough, with enough kinetic energy, to overcome the electrical barrier and merge together.

The potential energy U_{21} at charge 1 by charge 2 is given by:

$$U_{21} = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{D}, \tag{8.12}$$

where for two protons $q_1 = q_2 = e$, the elementary charge. To determine the velocity that is just fast enough for one proton to move into another, we equate the kinetic energy of motion to the electrical potential energy, or:

$$\frac{1}{2} m_P V^2 = \frac{e^2}{4\pi\varepsilon_0 D}, \tag{8.13}$$

and solve for the velocity, V,

$$V = \left[\frac{2e^2}{4\pi\varepsilon_0 m_P D}\right]^{1/2} = 1.66 \times 10^7 \text{ m s}^{-1}, \tag{8.14}$$

where the charge of a proton is $e = 1.6022 \times 10^{-19}$ C, the mass of the proton is $m_P = 1.6726 \times 10^{-27}$ kg, and we have assumed that the separation of two protons is comparable to the size of an atomic nucleus, or $D = 10^{-15}$ m, when they touch and merge together.

The mean speed of a proton inside the Sun is determined by equating the kinetic energy of the moving proton to the thermal energy that keeps it hot (Sect. 5.2), or

$$\frac{1}{2}\, m_p V_{th}^2 = \frac{3}{2}\, kT\,, \tag{8.15}$$

where T is the temperature and the Boltzmann constant $k = 1.38066 \times 10^{-23}$ J K^{-1}. Solving for the thermal velocity, V_{th}:

$$V_{th} = \left[\frac{3kT}{m_P}\right]^{1/2} \approx 157\, T^{1/2} \text{ m s}^{-1}. \tag{8.16}$$

For the center of the Sun, where the temperature $T = 15$ million K, the speed is only $V_{th} \approx 6.1 \times 10^5$ m s^{-1}, more than twenty times lower than that required to overcome the electrical repulsion between two protons. A proton would have to be at a temperature of $T \approx 10^{10}$, or 10 billion, K for the nuclear fusion to occur at the speed of 1.6×10^7 m s^{-1}.

Even at the enormous central temperature of the Sun, two protons do not seem to have enough energy to overcome their electrical repulsion and move into each other. As subsequently described in the text, when the quantum–mechanical penetration probability of a nucleus is combined with the Maxwellian distribution of particle speeds, a few protons at the higher speeds can fuse together in the center of the Sun.

As it turned out Eddington was correct, for the Russian physicist George Gamow (1904–1968) already had provided an explanation for this paradox. While at the University of Göttingen, in what is now Western Germany, Gamow showed how a subatomic particle could escape from the nucleus of a radioactive atom. He used the quantum theory of the very small, in which a subatomic particle can act like a spread-out wave with no precisely defined position, to determine the penetrability of the barrier surrounding a nucleus during radioactive decay (Gamow 1928; Gurney and Condon 1928). This tunnel effect works in the opposite way when nuclear particles merge rather than separate, and it explains why nuclei can fuse together inside a star.

With Gamow's encouragement, two young students, the English astronomer Robert d'Escourt Atkinson (1898–1982) and the Austrian physicist Fritz Houtermans (1903–1966), applied and extended his quantum-tunneling theory to the process of nuclear fusion in stars (Atkinson and Houtermans 1929). Using Gamow's penetration probability with the Maxwellian distribution of particle speeds, they showed that the fusion of light nuclei could create stellar energy in accordance with Einstein's formula connecting mass loss to energy gained.

It was immediately clear that the most effective nuclear interactions were those involving light nuclei with low charge and that only a few particles in the high-speed tail of the Maxwellian speed distribution would be able to penetrate nuclei. For this reason, nuclear reactions proceed slowly in the Sun and other stars. Atkinson and Houtermans also demonstrated that the rate of nuclear reactions substantially increases with the increasing central temperature of stars.

Atkinson then addressed this problem in far greater detail, arguing that the observed relative abundances of the elements might be explained by the synthesis of heavy nuclei from lighter ones within stars (Atkinson 1931). By this time the great stellar abundance of hydrogen had been established, and Atkinson subsequently demonstrated that the most likely nuclear reaction within stars is the collision of two protons to form a deuteron and a positron (Atkinson 1936).

The probability of penetration depends on the kinetic energy, or speed, and the electrical charges of the colliding particles. They have a greater impact when moving at faster speeds, but the electrical repulsion increases with the charge. For this reason, the lightest nuclei, the protons, are more likely to fuse together than the heavier ones, which have greater nuclear charge and mass. That is, the lightest elements carry the smallest charge, with less electrical repulsion between them, and they also move faster than heavier nuclei at a given temperature.

Even with this enhanced penetration probability, the average kinetic energy of two colliding protons is not enough for fusion to occur at the center of the Sun. However, the particles in a hot gas do not all move at the same average velocity. There is a relatively small number, in the high-speed tail of the Maxwellian speed distribution, which moves at much faster speeds, permitting fusion once the penetration probability also is considered.

The number of high-speed protons decreases exponentially with increasing speed and kinetic energy, whereas the tunneling probability increases exponentially with the energy (Fig. 8.1). In the overlap region, where the exponential decline meets the exponential rise, there are protons that can participate in the nuclear fusion reactions that make the Sun shine. Thus, protons do sometimes get close enough to move into each other and fuse together, even though their average energy is well below that required to overcome their electrical repulsion.

If two nuclei, designated by 1 and 2, undergo fusion, the nuclear reaction is written

$$1 + 2 \rightarrow 3 + 4 + Q, \tag{8.17}$$

where Q is the amount of energy released during this reaction, often given in units of MeV with 1 MeV $= 1.602 \times 10^{-13}$ J. The reaction rate, r_{12}, is given by

$$r_{12} = N_1 N_2 < \sigma V > \ \mathrm{m^3 s^{-1}}, \tag{8.18}$$

where N_1 and N_2 are the number densities of the two nuclei and $<\sigma V>$ denotes an averaged product of the interaction cross-section σ and the relative velocity V of the two nuclei. It takes into account both the tunneling cross section and the Maxwellian speed distribution, which depends on the temperature.

As an example, the fusion of two protons at the center of the Sun, where the temperature is 15.6 million K, has $<\sigma V> = 1.19 \times 10^{-49}$ m^3 s^{-1}, which is determined from both the theory of thermonuclear reactions and laboratory measurements of the proton–proton reaction.

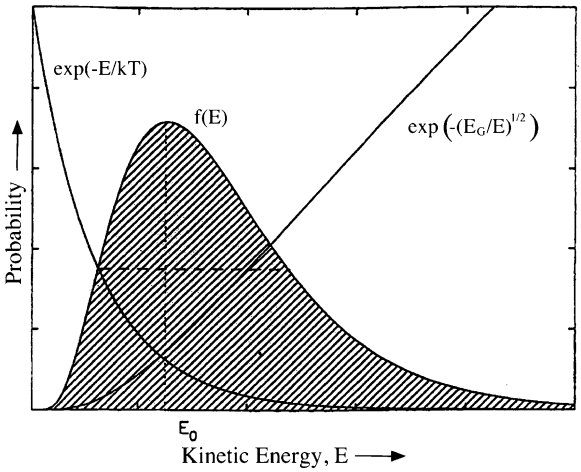


Fig. 8.1 Nuclear tunneling in the Sun's core The high-speed tail (*left*) of the Maxwellian distribution of nuclear particle speeds plotted as a function of kinetic energy, E, for protons near the center of the Sun. The P(E) function (*right*) describes the quantum mechanical probability of two protons overcoming the electrical repulsion between them; it depends on the Gamow energy E_G and the relative energy E of the two colliding protons. The shaded area (*center*) illustrates the function f(E), which is the product of the speed and penetration function; it determines the nuclear reaction rate in the core of the Sun

In stellar model calculations one uses the mass density, ρ, and the mass fraction X_i of the nuclei i, which are related to the number density N_i by:

$$N_i = \rho N_A \frac{X_i}{A_i} \tag{8.19}$$

where the Avogadro constant $N_A = 6.022141 \times 10^{23}$ mol^{-1}, and A_i is the atomic mass of nuclear species i in atomic mass units.

The mean lifetime τ_1 of nucleus 1 until destruction by fusion with nucleus 2 is:

$$\tau_1 = \frac{1}{N_2 < \sigma V >}. \tag{8.20}$$

where N_2 is the number density of nucleus 2 and $<\sigma V>$ is the average product of the interaction cross section σ and the relative velocity V of the two nuclei.

At the center of the Sun the mass density is $\rho = 1.513 \times 10^5$ kg m^{-3}, and since the mass is established by the protons with a mass $m_P = 1.6726 \times 10^{-27}$ kg, we have $N_1 = N_2 = \rho / m_P \approx 10^{32}$ m^{-3}. The mean lifetime for destruction of a proton by fusion with another proton at the center of the Sun is therefore $\tau \approx 10^{17}$ s $\approx$ 3 billion years, and since the density decreases with distance from the very center of the Sun there is plenty of time to keep the Sun shining by this reaction.

The power generated per unit mass, ε_{12}, by this reaction is given by:

$$\varepsilon_{12} = \frac{r_{12} Q}{\rho} \text{ J s}^{-1} \text{ kg}^{-1}. \tag{8.21}$$

The total energy released by the proton–proton chain of reactions that make the Sun shine is $Q \approx 26$ MeV $\approx 4 \times 10^{-12}$ J, and $\varepsilon_{12} \approx 10^{-3}$ J s^{-1} kg^{-1}. If the core of the Sun has a mass of $0.2\, M_\odot \approx 0.4 \times 10^{30}$ kg, the product of $\varepsilon_{12} \times 0.2\, M_\odot \approx 4 \times 10^{26}$ J s^{-1}, which is the present luminosity of the Sun, $L_\odot$.

Example: How frequent are proton collisions at the center of the Sun? The number density, N_P, of protons at the center of the Sun is $N_P = \rho_{C\odot}/m_P \approx 10^{32}$ m^{-3}, where the central density $\rho_{C\odot} = 1.5 \times 10^5$ kg m^{-3} and the proton mass $m_P = 1.6726 \times 10^{-27}$ kg. That is 100 million trillion trillion protons per cubic meter at the center of the Sun. The mean free path l between collisions is $l = 1/(N_P \pi R^2)$ (Sect. 5.2), the reciprocal of the product of the proton's area, πR^2 and the number density of protons. With a proton number density of 10^{32} m^{-3} a proton with a radius of 10^{-15} m will move about 0.003 m before striking another proton. The thermal velocity, $V_{th} = (3kT/m_P)^{1/2} \approx 6.1 \times 10^5$ m s^{-1} for a proton at a temperature of $T_{C\odot} = 15$ million degrees, where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, so the time between collisions is $l/V_{th} \approx 5 \times 10^{-9}$ s and there are about 200 million collisions occur every second.

Despite the exceptionally large number of collisions, a fusion reaction between two colliding protons does not happen very often. The protons nearly always bounce off one another without triggering a nuclear reaction during a collision. Even with the help of tunneling, the average proton must make about 10 trillion trillion, or 10^{25}, collisions before nuclear fusion can happen. It only occurs when the collision is almost exactly head on, and between exceptionally fast protons. This explains why the Sun does not expend all of its nuclear energy at once, like an immense hydrogen bomb.

8.3.2 Understanding Thermonuclear Reactions

What is of primary interest in fueling stars is the rate at which the nuclear reactions occur and the power they generate. But before considering these details, it is useful to know the units that nuclear astrophysicists commonly use when considering thermonuclear reactions. The include:

$$\begin{aligned}
\text{Size} &= 1\,\text{Fermi} = \ 1\ \text{fm} = 10^{-15}\ \text{m} \\
\text{Cross section} &= 1\,\text{barn} = 10^{-28}\ \text{m}^2 = \ 100\ \text{fm}^2 \\
\text{Energy} &= 1\,\text{MeV} = 1.60217646\text{x}\ 10^{-13}\ \text{J} = 1000\ \text{keV} \\
\text{Mass} &= 1\,\text{atomic mass unit} = \text{u} = 1.66053886 \times 10^{-27}\ \text{kg} \\
\text{Rest mass energy} &= \ \text{Mass} \times c^2
\end{aligned} \tag{8.22}$$

$$\begin{aligned}
\text{Proton} &= m_P = 938.2720\,\text{MeV} \\
\text{Neutron} &= m_n = 939.5654\,\text{MeV} \\
\text{Helium nucleus} &= 3727.379\,\text{MeV} = \text{Alpha particle} \\
\text{Electron} &= m_e = 0.511\,\text{MeV} \\
\text{Atomic mass unit} &= \text{u} = 931.494\,\text{MeV} \\
\text{Boltzmann constant} &= k = 1.38065 \times 10^{-23}\ \text{J K}^{-1} \\
&= 8.61733 \times 10^{-11}\,\text{MeV K}^{-1}
\end{aligned} \tag{8.23}$$

The mass of an atom's nucleus is, for example, always less than the sum of the individual masses of its constituent protons and neutron, or nucleons. Energy is removed to bind the subatomic nucleons together and form a nucleus, and this energy has mass. This mass is removed from the total mass of the original particles, and it is missing in the resulting nucleus. The missing mass is known as the nuclear *mass defect*, and represents the energy released when the nucleus formed. The mass defect, ΔM, for a nucleus containing A nucleons, Z protons, and A–Z neutrons is:

$$\Delta M = [Zm_p + \ (A - Z)\, m_n - m_{nuc}], \tag{8.24}$$

where A is the mass number of the nucleus, Z is the atomic number, m_P is the mass of the proton, m_n is the mass of the neutron, and m_{nuc} is the mass of the nucleus.

The *binding energy*, E_B, used to assemble the nucleus from its constituent nucleons is

$$E_B = \Delta M\, c^2. \tag{8.25}$$

The binding energy measures how tightly bound a nucleus is. The binding energy is the energy required to separate the nucleus into its constituent nucleons, and it also is a measure of the energy released during nuclear fusion of light nuclei into heavier ones. The binding energy released during nuclear fusion is responsible for energy production in the interior of stars.

The *binding energy per nucleon*, f, is given by $f = E_B/A$ and illustrated in Fig. 8.2. It shows that nuclei near iron have the largest nuclear binding, while lighter and heavier nuclei are less tightly bound. The graph indicates that energy can be released by combining lighter nuclei into heavier ones, known as *nuclear fusion*, provided they are less massive than iron. For example, it shows that about 7 MeV × 4 = 28 MeV in energy would be released during the fusion of four

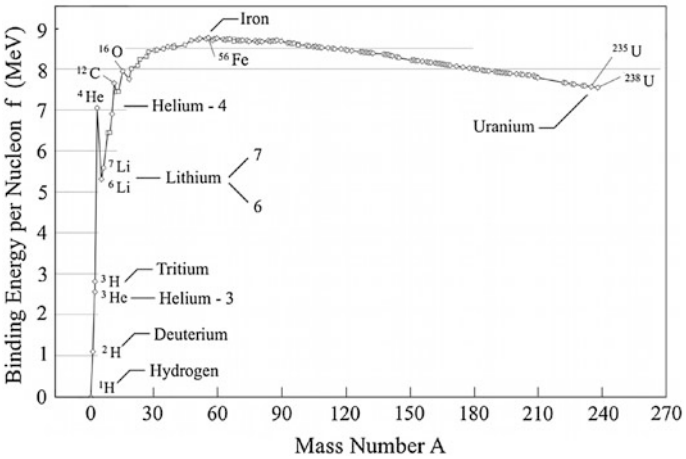


Fig. 8.2 Binding energy The binding energy per nucleon, denoted *f*, is shown as a function of atomic mass number, *A*. The highest nuclear binding energy is found near iron, denoted ^{56}Fe. Energy can be released by the nuclear fusion of lighter nuclei into heavier ones, provided they are less massive than iron. For example, the Sun shines by fusing four protons, denoted ^{1}H, into one helium nucleus, designated ^{4}He. About 7 MeV is released for each nucleon, and since four nucleons are involved the total energy release in the synthesis of one helium nucleus is about 28 MeV. Further energy is released in giant stars by the nuclear fusion of helium into carbon, denoted ^{12}C, but note that energy would not be released by fusing helium into lithium, denoted ^{6}Li or ^{7}Li, due to its lower binding energy per nucleon

protons, denoted ^{1}H, into one helium nucleus, designated ^{4}He; the exact value of the energy release is Q = 26.73 MeV.

The amount of energy released during nuclear reactions, or the Q value, results from the mass difference between the initial and final nuclei. The assumption that nuclear mass values are equal to the measured mass values usually makes little difference, for the binding energy and mass defect contribution of electrons are negligible compared to those of the nucleons. (As discussed subsequently, there is a small difference owing to the production of positrons.)

The helium is synthesized by fusing four protons together, so the fraction of mass converted during each one of these transmutations is $Q/(4m_Pc^2) = 0.007$, since Q = 26.73 MeV and the energy equivalent of the mass of the proton is $m_Pc^2 = 938$ MeV. Under the assumption that the Sun continues to emit its current luminosity $L_\odot$ by this process, it could last a time of $0.007\,M_\odot c^2/L_\odot = 3.2 \times 10^{18}$ s $\approx 10^{11}$ years, but since the nuclear reactions are limited to the core of the Sun, the nuclear lifetime is closer to 10^{10} years.

Example: Binding energy of a deuteron
A deuteron, denoted ^2H or ^2D, is the nucleus of the deuterium atom, and consists of one proton and one neutron. The mass of the deuterium atom, which is the mass of the deuteron for the accuracy needed, is $m_D = 2.013553$ u, and the mass of the proton and neutron are, respectively, $m_P = 1.007276$ u and $m_n = 1.008665$ u, where the atomic mass unit u $= 1.66053886 \times 10^{-27}$ kg. The mass defect $\Delta M = m_P + m_n - m_D = 0.002388$ u $= 3.9654 \times 10^{-30}$ kg, and the binding energy $E_B = \Delta M c^2 = 3.56 \times 10^{-13}$ J $= 2.224$ MeV, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1} and 1 J $= 6.24150974 \times 10^{12}$ MeV.

The binding energy and binding energy per nucleon for several nuclei are given in Table 8.2.

As suggested by Table 8.2 and Fig. 8.2, the binding energy per nucleon $f = E_B/A$ for cosmically abundant elements exhibits a nearly smooth increase with increasing A up until iron, with a steady decrease beyond that. This means that binding energy is released in the fusion of very light nuclei into somewhat heavier nuclei, and that iron is the most tightly bound, abundant nucleus. The reason the trend reverses after iron is the growing positive charge of the nuclei. It means that you cannot gain energy by synthesizing elements heavier than iron inside stars. Energy can only be released from these very heavy nuclei by nuclear fission into intermediate-mass nuclei.

The calculation of thermonuclear reaction rates is enormously complex, for it depends upon the specific reaction, whether or not there is a resonance in the reaction cross section, and on accelerator measurements of the reaction, carried out for decades by William A. "Willy" Fowler (1911–1995) and his colleagues; Fowler was awarded the 1983 Nobel Prize in Physics for his theoretical and experimental studies of nuclear reactions of importance in the formation of the

Table 8.2 Binding energy, E_B, and binding energy per nucleon, $f = E_B/A$, for some nuclei of atomic mass number A[a]

Nucleus	Symbol	A	E_B (MeV)	$f = E_B/A$ (MeV)
Proton	^{1}H	1	0.0000136	0.0000136
Deuteron	^{2}H or ^{2}D	2	2.22452	1.11226
Tralphium	^{3}He	3	7.7181	2.5727
Helium	^{4}He	4	28.3007	7.0752
Carbon	^{12}C	12	92.1617	7.6801
Oxygen	^{16}O	16	127.619	7.9762
Iron	^{56}Fe	56	492.258	8.7903
Uranium	^{238}U	238	1801.7688	7.5704

[a] The number preceding the letter symbol of the nucleus is the atomic mass number A, the number of nucleons

chemical elements in the universe. The fascinating details of the calculations can be found in Clayton (1968, 1984) and Rolfs and Rodney (1988), with summaries of the various reactions in Lang (1999). The details for non-resonant reactions are discussed in Focus 8.2.

Focus 8.2 Non-resonant thermonuclear reaction rates

All the nuclei in a star are positively charged, and they must tunnel through each other's Coulomb barrier of electrical repulsion for fusion to occur, even at the high temperatures within stars. Quantum mechanical considerations show that in the absence of resonances the cross section for the tunneling, $\sigma(E)$, can be written (Gamow 1928; Gurney and Condon 1928):

$$\sigma(E) = \frac{S(E)}{E} \exp\left[-\left(\frac{E_G}{E}\right)^{1/2}\right], \tag{8.26}$$

where the kinetic energy, E, of two nuclei of masses M_1 and M_2 in their center-of-mass system is:

$$E = \frac{\mu V^2}{2} = \frac{M_1 M_2}{M_1 + M_2}\left(\frac{V^2}{2}\right), \tag{8.27}$$

for a reduced mass $\mu = M_1 M_2/(M_1 + M_2)$ and relative velocity V of the two nuclei. The Gamow energy E_G is given by:

$$E_G = (\pi\alpha Z_1 Z_2)^2 2\mu c^2 = \left[0.98948 Z_1 Z_2 A^{1/2}\right]^2 \text{MeV}. \tag{8.28}$$

The fine structure constant $\alpha = e^2/(2\varepsilon_0 hc) = 7.29735 \times 10^{-3} = [137.0356]^{-1}$, for elementary charge e, electric constant ε_0, Planck constant h and speed of light c, the charges of the two nuclei, in units of the proton charge, are Z_1 and Z_2, and the reduced nuclear mass number $A = A_1 A_2/(A_1 + A_2)$, where the mass number A_1 denotes the number of nucleons, or the number of protons plus the number of neutrons, in nucleus 1. For the fusion reaction of two protons, $Z_1 = Z_2 = 1$, the $\mu = m_P/2$ for proton mass m_P, the $A = 0.5$, the proton rest energy $m_P c^2 = 938.2723$ MeV, and the Gamow energy is $E_G = (\pi/137)^2 \times 938.2723 = 0.494$ MeV ≈ 0.5 MeV $=$ 500 keV.

For most nuclear reactions in stars, the strength factor $S(E)$ is between 10 MeV-barns and 1 keV-barns, but it usually cannot be calculated from theoretical considerations. Far from a nuclear resonance, the factor $S(E)$ is a slowly varying function of E and may conveniently be expressed as a power series expansion:

$$S(E) = S_0\left[1 + \frac{S'(0)}{S_0}E + \frac{S''(0)}{2S_0}E^2\right]. \tag{8.29}$$

References to the experimental measurements of the constants in this expression are provided in Lang (1999).

For a number density N_1 and N_2 of the reacting nuclei, the reaction rate r_{12} is given by:

$$r_{12} = N_1 N_2 < \sigma V > \quad \mathrm{m}^{-3}\mathrm{s}^{-1}, \tag{8.30}$$

where the Gamow tunneling cross section is combined with the Maxwellian speed distribution at temperature, T, to obtain:

$$\langle \sigma V \rangle = \left(\frac{8}{\pi\mu}\right)^{1/2} \frac{S_0}{(kT)^{3/2}} \int_0^\infty \exp\left(\frac{-E}{kT}\right) \exp\left[-\left(\frac{E_G}{E}\right)^{1/2}\right]. \tag{8.31}$$

The integration is over the function:

$$f(E) = \exp\left(\frac{-E}{kT}\right) \exp\left[-\left(\frac{E_G}{E}\right)^{1/2}\right], \tag{8.32}$$

where the first term is due to the fall in the number of particles with increasing energy, found in the Maxwellian distribution, and the second term expresses the exponential rise in tunneling probability with increasing energy. The overlap of these two terms, previously shown in Fig. 8.1, defines the region of energy in which the nuclear reactions occur.

Most of the reactions take place at the effective thermal energy, E_0, or Gamow peak, given by Fowler and Hoyle (1964):

$$E_0 = \left(\frac{kT}{2}\right)^{2/3} E_G^{1/3} = 0.1220\left(Z_1^2 Z_2^2 A\right)^{1/3} T_9^{2/3}\,\mathrm{MeV} \tag{8.33}$$

where T_9 is the temperature in billions of K, or $T_9 = T/10^9$. For light nuclei and temperatures of some tens of millions of degrees, the most effective energy E_0 is usually 10 to 30 keV. This energy is greater than $kT = 86\ T_9$ keV, reflecting the fact that the particles in the high-speed tail of the Maxwellian distribution contribute to the reactions. For the proton–proton reaction at the center of the Sun, $Z_1 = Z_2 = 1$, the $A = 0.5$, and $T_9 = 0.0156$ for a central temperature of 15.6 million K, with $kT = 1.3$ keV, so the $E_0 \approx 0.006$ MeV $= 6$ keV for this solar reaction.

Nuclei with energies close to E_0 and spread over the energy range ΔE_0 contribute mainly to the total rate of the thermonuclear reactions, where ΔE_0 is given by full width to half maximum of the function $f(E)$, or

$$\Delta E_0 = 4\left(\frac{E_0 kT}{3}\right)^{1/2} \approx 0.65 E_G^{1/6} (kT)^{5/6} \approx 0.237\left(Z_1^2 Z_2^2 A\right)^{1/6} T_9^{5/6}\,\mathrm{MeV}. \tag{8.34}$$

The spread ΔE_0 is usually between 4 and 10 keV, just a bit smaller than E_0. For the proton–proton fusion reaction, the two are about equal.

The reaction rate is therefore approximated by (Burbidge et al. 1957; Fowler et al. 1975; Clayton 1968, 1984; Rolfs and Rodney 1988)

$$r_{12} = N_1 N_2 \left(\frac{2}{\mu}\right)^{1/2} \frac{S_0 \Delta E_0}{(kT)^{3/2}} \exp\left[-\frac{3E_0}{kT}\right], \tag{8.35}$$

or equivalently

$$r_{12} \approx 0.65 N_1 N_2 \left(\frac{2}{\mu}\right)^{1/2} \frac{S_0 E_G^{1/6}}{(kT)^{2/3}} \exp\left[-3\left(\frac{E_G}{4kT}\right)^{1/3}\right]. \tag{8.36}$$

For the fusion of two protons, $S_0 = 3.89 \times 10^{-25}$ MeV barns $= 3.89 \times 10^{-53}$ MeV m^2 (Kamionkowski and Bahcall 1994) and for the Sun center where $T = 15.6$ million K, this reaction has a rate of $r_{12} \approx 10^{-49} N_1 N_2$ m^3 s^{-1} $\approx 10^{15}$ m^3 s^{-1}, where $N_1 = N_2 \approx 10^{32}$ m^{-3}.

The energy generation, ε_{12}, or the power generated per unit mass, for a reaction that generates an energy Q is

$$\varepsilon_{12} = \frac{r_{12} Q}{\rho} \quad \text{J s}^{-1}\,\text{kg}^{-1}, \tag{8.37}$$

where ρ is the mass density at the reaction site. In the core of the Sun, the complete chain of thermonuclear reactions releases a $Q \approx 26$ MeV $\approx 4.2 \times 10^{-12}$ J and $\varepsilon_{12} \approx 10^{-3}$ J s^{-1} kg^{-1}.

8.3.3 Hydrogen Burning

The detailed nuclear reactions inside stars could not be understood until the 1930s when several subatomic particles were known, including the neutron, discovered in 1932, the positron, detected in cosmic ray showers in 1932, and the neutrino, hypothesized in 1933.

The German physicist Carl Friedrich von Weizsäcker (1912–2007) proposed that the solution to the solar-energy problem lay in the fusion of protons, which Atkinson previously suggested (Weizsäcker 1937; Atkinson 1936). Then Gamow, who had immigrated to the United States, suggested to one of his graduate students, Charles Critchfield (1910–1994), that he calculate the details of the reaction. The results were sent to the German-born American physicist Hans A. Bethe (1906–2005) at Cornell University, who found them to be correct, and the two

published a joint paper titled "The Formation of Deuterons by Proton Combination" (Bethe and Critchfield 1938).

In April 1939, Gamow, who was teaching physics at George Washington University in Washington, DC, organized a conference to bring astronomers and physicists together to discuss the problem of stellar energy generation, under the sponsorship of the Department of Terrestrial Magnetism of the Carnegie Institution. At this conference, the astronomers told the physicists what they knew about the internal constitution of stars, which was quite a bit.

By then, it was known that the lightest element, hydrogen, is by far the most abundant element in the outer atmosphere of the Sun (Unsold 1928; McCrea 1929; Russell 1929), as well as its interior (Strömgren 1931, 1932). At Gamow's conference, the Danish astronomer Bengt Strömgren (1908–1987) additionally reported that because the Sun was predominantly hydrogen, it would have a central temperature of about 15 million K rather than 40 million K, as estimated by Eddington under the assumption that the Sun had approximately the same chemical composition as the Earth, with a preponderance of heavy elements rather than hydrogen. The lower temperature meant that the calculations of Bethe and Critchfield correctly predicted the Sun's luminosity. Bethe, who attended the conference, was so stimulated by the meeting that within six months he had published a paper titled "Energy Production in Stars," which explains how the Sun fuses hydrogen into helium, releasing energy to heat the Sun's core and generate the radiation that makes it shine (Bethe 1939, 1967).

At about the same time, Weizsäcker showed how other nuclear reactions could fuel stars using carbon as a catalyst in the synthesis of helium from hydrogen, but he did not investigate the rate of energy production or its temperature dependence, which Bethe subsequently did (Bethe 1939; Weizsäcker 1938). He eventually received the Nobel Prize in Physics, in 1967, for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars.

The sequence of nuclear reactions that make the Sun shine is called the proton–proton chain because it begins with the fusion of two protons. The complete chain of nuclear reactions also is known as the hydrogen-burning reaction – for it is hydrogen nuclei, protons, which are being consumed to make helium. However, it is not combustion in the ordinary chemical sense; in the proton–proton chain (Fig. 8.3), four protons are fused together to form a helium nucleus that contains two protons and two neutrons.

Nuclear reactions often are written in shorthand notation using letters to denote the nuclei and other subatomic particles. An arrow → specifies the reaction. Nuclei on the left side of the arrow react to form the products given on the right side. The amount of energy released during the reaction can also be given on the far right side of the arrow, and is often specified in units of MeV where $1\ \text{MeV} = 1.692 \times 10^{-13}$ J. A letter and a preceding superscript designate a nucleus. For historical reasons, the nuclei of the hydrogen isotopes ^{1}H, ^{2}H, and ^{3}H, also are named protons, deuterons and tritons, and the nucleus of ^{4}He is called an alpha particle. A Greek letter γ denotes energetic gamma-ray radiation. A positron

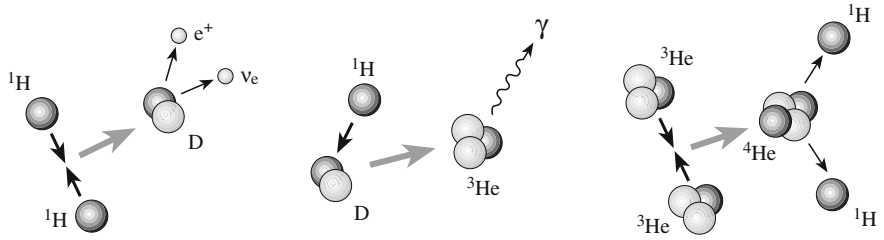


Fig. 8.3 Proton-proton chain Hydrogen nuclei, or protons, are fused together to form helium nuclei within the solar core, providing the Sun's energy. In 1939, the German-born American physicist Hans Bethe (1906–2005) described the detailed sequence of nuclear-fusion reactions, called the proton–proton chain. It begins when two protons, here designated by the letter ^{1}H, combine to form the nucleus of a deuterium atom, the deuteron that is denoted by D, together with the emission of a positron, denoted by e^+, and an electron neutrino, designated by ν_e. Another proton collides with the deuteron to make a nuclear isotope of helium, denoted by ^{3}He, and then a nucleus of helium, designated by ^{4}He, is formed by the fusion of two ^{3}He nuclei, returning two protons to the gas. Overall, this chain fuses four protons together to make one helium nucleus. Even in the hot, dense core of the Sun, only rare, fast-moving particles can take advantage of the tunnel effect and fuse in this way

is denoted by e^+ and an electron is designated by e^-. The symbol v_e denotes an electron neutrino.

The superscript that precedes a nucleus letter is the mass number, A, which is the sum of the neutrons and the protons and the total number of nucleons in the nucleus. Different isotopes of an element have the same number of protons and the same letter symbol but a different number of neutrons and a different superscript. For instance, a rare isotope of helium, designated ^3He, has two protons and one neutron in its nucleus, whereas the nucleus of the common form of helium, ^4He, has two protons and two neutrons.

In the first step of the proton–proton chain, two protons, each designated by either ^1H or p, meet head on and merge into each other, tunneling through the electrical barrier separating them. The two protons combine to make a deuteron, ^2D, the nucleus of a heavy form of hydrogen known as deuterium.

Because a deuteron consists of one proton and one neutron, one of the protons must be neutralized. It is turned into a neutron, n, with the ejection of a positron, e^+, to carry away the proton's charge, and an electron neutrino, v_e, to balance the energy in the reaction. This is the positive beta decay reaction, denoted by:

$$p \rightarrow n + e^+ + v_e, \tag{8.38}$$

which applies to only one of the two protons making the deuteron. The initiating proton–proton reaction that involves both protons therefore is written:

$$p + p \rightarrow {}^2D + e^+ + v_e, \tag{8.39}$$

which releases 0.425 MeV in energy.

Each proton inside the Sun is involved in a collision with other protons millions of times every second, but only exceptionally hot ones can tunnel through their electrical repulsion and fuse together. Only one collision in every ten trillion trillion initiates the proton–proton chain.

The electron neutrinos produced in the first step of the proton–proton chain escape from the Sun without reacting with matter, carrying energy away. However, the positron or positive electron – the anti-matter particle of the electron – is consumed immediately. Anti-matter and matter cannot coexist. As soon as any anti-matter is produced, it is immediately wiped out of existence by colliding with an electron, and the reason why we live in a material world is simply because there is more matter than anti-matter.

The positron, e^+, created in the first step of the proton–proton chain almost instantly encounters a free electron, e^-, and both become pure energy. The two subatomic particles collide and annihilate one another in a flash of radiation at gamma ray wavelengths, denoted γ. This energy-producing pair-annihilation reaction can be written symbolically as:

$$e^- + e^+ \rightarrow 2\gamma, \tag{8.40}$$

where each gamma-ray photon has an energy of 0.511 MeV, corresponding to the rest mass energy of an electron, for a total reaction energy release of 1.022 MeV.

The next step follows with little delay. In less than 1 s, the deuteron collides with another proton to form a nucleus of light helium, ^3He, and releases yet another gamma-ray photon, with about 5.49 MeV in energy. In symbolic terms, the second step of the proton–proton chain is written:

$$^2D + p \rightarrow {}^3He + \gamma. \tag{8.41}$$

This reaction occurs so easily that deuterium cannot be synthesized inside stars; it is consumed quickly to make heavier elements.

In the final part of the proton–proton chain, two such light helium nuclei meet and fuse together to form a nucleus of normal heavy helium, ^4He, and return two protons to the solar gas. This step takes about 1 million years on average and is written:

$$^3He + {}^3He \rightarrow {}^4He + 2p, \tag{8.42}$$

generating another 12.86 MeV of energy. This normal helium nucleus contains two protons and two neutrons. So, two of the protons that contributed to the formation of helium were converted into neutrons by the positive beta-decay reaction. The total energy released in the proton–proton chain is the sum of that released by all of the contributing reactions,

$$Q = 2 \times 1.022 + 2 \times 0.425 + 2 \times 5.49 + 12.86 = 26.73 \text{ MeV}. \tag{8.43}$$

This last reaction in the proton–proton chain happens 86 % of the time. Less frequent terminations involve the interaction of light helium with heavy helium to form beryllium (Focus 8.3).

Focus 8.3 Secondary nuclear fusion reactions in the Sun

The proton–proton chain will end by making beryllium about 14 % of the time. A nucleus of light helium, ^3He, will fuse with a nucleus of heavier helium, ^4He, to form a nucleus of light beryllium, ^7Be, according to the nuclear fusion reaction:

$$^3He + {}^4He \rightarrow {}^7Be + \gamma, \tag{8.44}$$

where γ denotes energetic gamma-ray radiation.

Most of the time, the light beryllium will combine with an electron, e^-, to make a nucleus of lithium, ^7Li, which then joins a proton, ^1H, to make two nuclei of heavy helium, ^4He. The reactions are (Parker et al. 1964):

$$^7Be + e^- \rightarrow {}^7Li + \nu_e, \tag{8.45}$$

and

$$^7Li + {}^1H \rightarrow {}^4He + {}^4He, \tag{8.46}$$

where ν_e denotes an electron neutrino.

About 0.02 % of the time, the light beryllium combines with a proton to make boron, ^8B. The boron is a radioactive nucleus that decays in just one second into beryllium, ^8Be, together with the emission of a positron, e^+, and an electron neutrino. The heavy beryllium then decays to make two nuclei of heavy helium, completing the conversion of protons into helium. These secondary nuclear fusion reactions are:

$$^7Be + {}^1H \rightarrow {}^8B + \gamma \tag{8.47}$$

and

$$^8B \rightarrow {}^8Be + e^+ + \nu_e \tag{8.48}$$

where the positron annihilates with an electron to create a gamma ray γ, and

$$^8B \rightarrow {}^4He + {}^4He \tag{8.49}$$

A total of six protons are required to produce the two ^3He nuclei that go into this last reaction, but two protons are returned to the solar interior to be reused later. Because two protons and a helium nucleus are produced, the net result of the proton–proton chain is:

$$4p \rightarrow {}^4He + \text{gamma ray radiation} + \text{neutrinos}, \tag{8.50}$$

which releases a net energy of 26.73 MeV or 4.28×10^{-12} J each time it occurs, using 1 MeV $= 1.602 \times 10^{-13}$ J. That corresponds to the energy $\Delta E = \Delta mc^2$, where Δm is the mass loss in converting four protons into one helium nucleus and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The energy released in each reaction is very small, but there are a lot of them, about 10^{38} every second.

Deducting the 2×0.511 MeV $= 1.022$ MeV from the annihilation of the two pre-existing electrons by interaction with protons, we have an energy of 25.71 MeV, which corresponds to the rest mass difference between four protons and a ^{4}He nucleus. That is:

$$25.71 \text{ Mev } = (4m_P - m_{He})c^2 = 0.007\big(4m_Pc^2\big), \tag{8.51}$$

where m_P denotes the proton mass, m_{He} designates the mass of a helium nucleus, the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, the $m_Pc^2 = 938.2720$ MeV and $m_{He}c^2 = 3727.379$ MeV. Thus, the rest-mass-to-energy conversion of the proton–proton chain is 0.007 or 0.7 %.

This energy leaves the Sun as radiation, and the part of this radiation that constitutes visible light is what makes the Sun shine. The subatomic energy that is liberated also keeps the core of the Sun hot, assuring continuation of the nuclear reactions. The time, τ, to radiate away just 10 % of the energy available from this source is:

$$\tau = \frac{0.1(0.007)M_\odot c^2}{L_\odot} \approx 3.26 \times 10^{17}\text{s} \approx 10^{10} \text{ years}, \tag{8.52}$$

where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and 1 year $= 3.1557 \times 10^7$ s.

Example: The Sun is losing mass

The Sun shines by the energy released, ΔE, from the mass loss, Δm, every time four protons are converted into a helium nucleus. That is, the helium nucleus is slightly less massive, by a mere 0.007, or 0.7 %, than the four protons that combine to make it, so there is an energy, ΔE, released given by:

$$\Delta E = \Delta m\, c^2 = (4m_p - m_{He})c^2 = 0.007\big(4m_p\big)c^2 \approx 4.2 \times 10^{-12} \text{ J} \tag{8.53}$$

where the mass of the proton is $m_p = 1.6726 \times 10^{-27}$ kg, the mass of the helium nucleus is $m_{He} = 6.644 \times 10^{-27}$ kg, and $c = 2.9979 \times 10^8$ m s^{-1} is the speed of light.

The number of reactions, N, that must occur every second to make the Sun shine with its present luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, is:

$$N = \frac{L_\odot}{\Delta E} \approx 10^{38} \text{ reactions per second}, \tag{8.54}$$

and the mass loss, ΔM, from the Sun in just one second is

$$\Delta M = \frac{\Delta m L_\odot}{\Delta E} = \frac{L_\odot}{c^2} \approx 4.26 \times 10^9 \text{ kg s}^{-1}. \quad (8.55)$$

Since 1 ton is equal to 1,000 kg, about 4 million tons of matter disappears from the Sun every second, vanishing to provide the Sun's energy. Provided the Sun has been shining at the same rate ever since it formed 4.6 billion years ago, it has lost about 6.2×10^{26} kg, using 1 year $= 3.1557 \times 10^7$ s. However, this mass loss is trivial compared to the Sun's total mass of $M_\odot = 1.989 \times 10^{30}$ kg. The Sun has lost only 0.0003 of its original mass in all that time.

The time, t, required to consume the entire solar mass this way would be:

$$t = \frac{M_\odot}{\Delta M} = 0.47 \times 10^{21} \text{ s} \approx 10^{13} \text{ years}, \quad (8.56)$$

but since the nuclear reactions are confined to the hot, central core, the hydrogen is depleted in a lifetime of about one-tenth this value.

8.3.4 Why Doesn't the Sun Blow Up?

Unlike a nuclear bomb on the Earth, the temperature-sensitive reactions inside the Sun act like a thermostat, releasing energy in a steady, controlled manner at exactly the rate needed to keep the Sun in equilibrium between the inward pull of gravity and the outward pressure of the moving subatomic particles. If the rate of the thermonuclear reactions in the central regions of the Sun rises as the result of a temperature increase, nuclei move faster and create more pressure. The entire body of the Sun then would expand and thereby bring down its central temperature. If the rate of core nuclear reactions were to drop, gravity would pull the Sun inwards, and the resulting increase in central temperature would bring the rate of energy production back into equilibrium. So the pendulum continues to swing between gravity and fusion, with no winner. That is how the Sun harnesses its nuclear energy, which it has been doing for 4.6 billion years.

8.4 The Mystery of Solar Neutrinos

8.4.1 The Elusive Neutrino

Neutrinos, or "little neutral ones," are tiny, invisible packets of energy with no electrical charge and almost no mass, traveling at nearly the speed of light (see Sect. 7.4). They are produced in great profusion by thermonuclear reactions in the

Sun's core, removing substantial amounts of energy that is never seen again. Every second, trillions upon trillions of the solar neutrinos pass right through the Earth without even noticing that it is there (Focus 8.4). At night, the solar neutrinos travel through the Earth before passing through the walls of our houses and even through our bodies, without us ever noticing them.

Focus 8.4 Trillions upon trillions of neutrinos

The aggregate number of solar neutrinos is staggering. Every time the proton–proton chain creates one helium nucleus, it releases an energy, $\Delta E = \Delta m c^2 = 0.007\ (4m_p)\ c^2 = 4.2 \times 10^{-12}$ J, where Δm is the mass difference between the helium nucleus and the four protons that went into making it. The mass of the proton is $m_p = 1.6726 \times 10^{-27}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The energy released by the Sun's ongoing nuclear fusion reactions works its way out of the Sun to provide its present luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. The total number of helium-producing proton–proton chains required to fuel the Sun's energy every second is $L_\odot/\Delta E \approx 10^{38}$, and since two neutrinos are emitted every time one helium nucleus is made, we conclude that the Sun emits 2×10^{38} neutrinos every second. If the Sun has been shining at the same rate for the past 4.6 billion years, it has emitted an astonishing 3×10^{55} neutrinos.

The number of neutrinos passing through the Earth each second is less than those emitted by the Sun, diminished by the ratio of the Earth's cross sectional area to the area of a sphere with a radius equal to the mean distance from the Sun to Earth. Thus, the number of neutrinos passing through Earth per second is:

$$\left(\frac{2L_\odot}{\Delta E}\right)\left(\frac{\pi R_E^2}{4\pi (D_\odot)^2}\right) \approx 10^{29}, \tag{8.57}$$

where the radius of the Earth is $R_E = 6.378 \times 10^6$ m and the mean distance between the Earth and the Sun is $D_\odot = 1$ AU $= 1.496 \times 10^{11}$ m.

So, there are 0.1 million trillion trillion neutrinos passing through the Earth every second. To obtain the number of neutrinos passing through every square meter of the side of the Earth facing the Sun, just divide by the Earth's area πR_E^2 to get about 7×10^{14}, or 700 thousand billion, neutrinos per square meter.

The flux of solar neutrinos expected at the Earth is calculated using supercomputers that culminate in the Standard Solar Model that describes the Sun's luminous output, size, and mass at its present age. Such calculations have been developed and refined by John N. Bahcall (1934–2005) of the Institute for Advanced Study at Princeton (Bahcall 1964, 1978; Bahcall and

Pinsonneault 2004), and other astrophysicists throughout the world, such as Sylvaine Turck-Chièze (1951–) at Saclay, France (Turck-Chièze et al. 1988).

The computer models always include three basic assumptions:

(1) Energy is generated by hydrogen-burning reactions in the central core of the Sun, and there is no mixing of material between the core and overlying regions. The nuclear reaction rates depend on the density, temperature, and composition of the core, as well as coefficients extrapolated from laboratory experiments.
(2) The outward thermal pressure, due to the energy-producing reactions, just balances the inward pressure due to gravity, thereby keeping the Sun from either collapsing or blowing up.
(3) Energy is transported from the deep interior to the visible solar disk via radiation and convection (see the next Sect. 8.5). The great bulk of energy is carried by radiative transport with an opacity determined from atomic physics calculations.

One begins with a newly formed Sun having a uniform composition, and an element abundance that is observed in the visible solar disk today. The model then imitates the evolution of the Sun to its present age of 4.6 billion years by slowly converting hydrogen into helium within the model core. The central nuclear reactions supply both the radiated luminosity and the local heat or pressure, while also creating neutrinos and producing composition changes in the core. The Sun's current luminosity, size, and neutrino flux are obtained after 4.6 billion years when about 37 % of the hydrogen in the core has now been transformed into helium. Once the Standard Solar Model has specified the neutrino flux the predictions are extended to specific experiments that detect neutrinos of different energies.

8.4.2 Solar Neutrino Detectors Buried Deep Underground

Unlike a conventional optical telescope, which is placed as high as possible to minimize distortion by the Earth's obscuring atmosphere, a solar neutrino detector is buried beneath a mountain or deep within the Earth's rocks inside mines. This shields the instrument from deceptive signals caused by cosmic rays. There, beneath tons of rock that only a neutrino can penetrate, detectors unambiguously measure neutrinos from the Sun. If neutrino detectors were placed on the Earth's surface, they would detect high-energy particles and radiation produced by cosmic rays interacting with the Earth's atmosphere.

Thus, solar-neutrino astronomy involves massive subterranean detectors that look right through the Earth and observe the Sun at night or day. The first such neutrino telescope, constructed in 1967 by Raymond Davis, Jr. (1914–2006), was a 615-ton tank located 1.5 km underground in the Homestake Gold Mine near Lead, South Dakota. The huge cylindrical tank was filled with 378,000 liters of cleaning fluid, technically called perchloroethylene or "perc" in the dry-cleaning trade;

each molecule of the stain remover consists of two carbon atoms and four chlorine atoms (Davis 1964; Davis et al. 1968).

Most solar neutrinos passed through the tank unimpeded. Occasionally, however, a neutrino scored a direct hit with the nucleus of a chlorine atom, turning one of its neutrons into a proton, emitting an electron to conserve charge, and transforming the chlorine atom into an atom of radioactive argon. The new argon atom rebounded from the encounter with sufficient energy to break free of the perc molecule and enter the surrounding liquid. Because argon is chemically inert, it can be culled from the liquid by bubbling helium gas through the tank. The number of argon atoms recovered in this way measured the incident flux of solar neutrinos.

Every few months, Davis and his colleagues flushed the tank with helium, extracting about 15 argon atoms from a tank the size of an Olympic swimming pool. That was a remarkable achievement considering that the tank contained more than 1 million trillion trillion, or 10^{30}, chlorine atoms. The scientists persisted for nearly thirty years, capturing signs of only 2,000 neutrinos in all that time. However, the consequences were enormous. The measurements implied not only that nuclear-fusion reactions indeed were providing the Sun's energy, making it shine, but there also was a small unexpected problem with the result that led to a new understanding of the physics of neutrinos.

The neutrino reaction rate with atoms in neutrino detectors is so slow that a special unit was invented to specify the experiment-specific flux. The Solar Neutrino Unit, abbreviated SNU and pronounced "snew," is equal to one neutrino interaction per second for every trillion trillion trillion, or 10^{36}, atoms. Even then, the predictions were only a few SNU per month for even the largest, most-massive, detectors first constructed.

The Homestake detector always yielded results in conflict with the most accurate theoretical calculations. The final experiment value was 2.55 ± 0.25 SNU, where the $\pm$ value denotes an uncertainty of one standard deviation (Cleveland et al. 1998). (A standard deviation is a statistical measurement of the uncertainty of a measurement; a definite detection must be above three standard deviations and preferably above five.) In contrast, theoretical results using the Standard Solar Model predicted that the Homestake detector should have observed a flux of 8.5 ± 1.8 SNU. So the tank full of cleaning fluid captured almost one third of the expected number of neutrinos (Fig. 8.4).

The discrepancy between the observed and calculated values is known as the Solar Neutrino Problem. Its significance was confirmed in 2002, when Davis received the Nobel Prize in Physics; sharing it with Japanese scientist Masatoshi Koshiba (1926–), whose group used another giant, underground detector, named Kamiokande, to detect both solar and supernova neutrinos.

In 1987, the Kamiokande neutrino detector began to monitor solar neutrinos, confirming the neutrino deficit observed by Davis. This second experiment, located in a mine at Kamioka, Japan, consisted of a 4,500-ton, or 4.5 million-liter, tank of pure water. Nearly 1,000 light detectors were placed in the tank walls to measure signals emitted by electrons knocked free from water molecules by passing neutrinos.

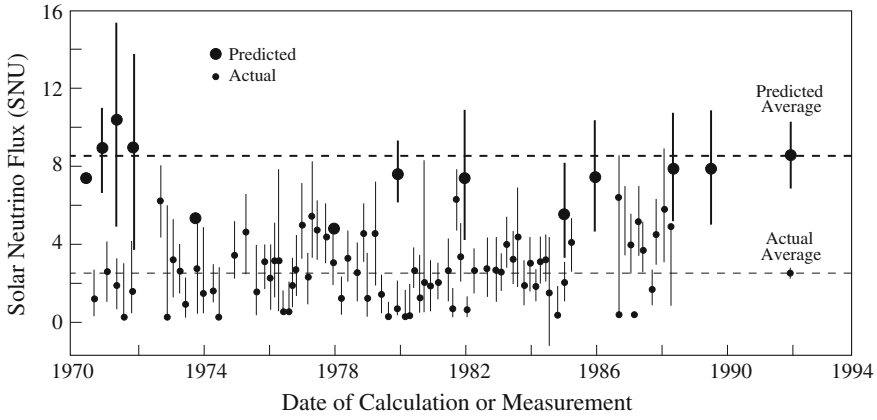


Fig. 8.4 Solar neutrino fluxes Calculated and measured solar neutrino fluxes have disagreed for several decades. The fluxes are measured in solar neutrino units (SNU) which are defined as 1 neutrino interaction per trillion trillion trillion, or 10^{36}, atoms per second. Measurements from the chlorine neutrino detector (*small dots*) give an average solar neutrino flux of 2.6 ± 0.2 SNU (*lower broken line*), well below theoretical calculations (*large dots*) that predict a flux of 8.5 ± 1.8 SNU (*upper broken line*) for the Standard Solar Model. Other experiments also have observed a deficit of solar neutrinos, suggesting that either some process prevents neutrinos from being detected or there is an incomplete understanding of the nuclear processes that make the Sun shine

When an energetic solar neutrino collides with an electron in the water, the neutrino knocks the electron out of its atomic orbit, pushes it forward in the direction of the incident neutrino, and accelerates the electron to nearly the speed of light. In water, the electron moves faster than the light it radiates, and as a result the electron produces a cone-shaped pulse of light about its path. The faint blue glow is technically known as Cherenkov radiation, named after the Soviet physicist Pavel A. Cherenkov (1904–1958) who discovered the effect (Cherenkov 1937).

The axis of the light cone gives the electron's direction, which is the direction from which the neutrino arrived. Because the observed electrons were preferentially scattered along the direction of an imaginary line joining the Earth to the Sun, the Kamiokande water experiment also confirmed that the neutrinos indeed are produced by nuclear reactions in the Sun's core. After 1000 days of observation, Yoji Totsuka (1942–2008), speaking for the Kamiokande collaboration led by Masatoshi Koshiba (1926–), could therefore report that neutrinos really are coming for the Sun, where nuclear fusion reactions are taking place, while also confirming the neutrino deficit observed by Davis (Totsuka 1991).

The deficit of solar neutrinos was subsequently confirmed in the 1990s by two massive underground detectors using gallium, a rare and expensive metal used in the red lights of hand calculators and other pieces of electronic equipment. When combined with the chlorine and the Kamiokande results, all four experiments seemed to have confirmed that solar neutrinos are missing, and that the Solar Neutrino Problem is real.

8.4.3 Solving the Solar Neutrino Problem

After almost 40 years of meticulous measurements and calculations, the neutrino count still came up short! Massive underground detectors always observed fewer neutrinos than theory states they should detect, and many scientists thought that the discrepancy was due to an incomplete understanding of neutrinos (Bahcall and Bethe 1990).

As it turned out, neutrinos are transforming into a different form during their journey from the center of the Sun, escaping detection by changing character. We haven't mentioned it yet, but scientists have learned that there are three separate types, or flavors, of neutrinos, each named after the fundamental, subatomic particle with which it is most likely to interact (Focus 8.5). All of the neutrinos generated inside the Sun are electron neutrinos, designated v_e; this is the type that interacts with electrons, denoted e^-. The other two types, the muon neutrino, v_μ, and the tau neutrino, v_τ, interact with muons, μ, and tau particles, τ, respectively.

Focus 8.5 Leptons

A lepton is an elementary, subatomic particle, whose name comes from the Greek word *lepton* meaning "fine, small, thin, subatomic or slender." Altogether there are six types of leptons, which are divided into two classes: the three charged leptons and the three neutral, or uncharged leptons, known as neutrinos.

The electron, denoted by e, is the best-known lepton and the first to be discovered. Unlike the stable electron, the other two charged leptons, the muon denoted by μ and the tau particle designated τ, are unstable subatomic particles. They are unfamiliar to most of us because they die shortly after birth. The muon decays into an electron, a muon neutrino and an electron anti-neutrino in just 2 millionths, or 2×10^{-6}, of a second, and the tau particle disappears just three-tenths of a million-millionth, or 3×10^{-13}, of a second after it is made.

Every charged lepton has a corresponding antiparticle of opposite charge but equal mass. The antiparticle of the electron is thus known as the positron, or positive electron. The lepton of negative charge is denoted with a $-$ superscript, as e^-, μ^-, and τ^-, and the corresponding antiparticle with the positive charge is designated by a $+$ subscript, with e^+, μ^+, and τ^+.

The three electrically neutral leptons, or neutrinos, have a small but non-zero mass, and rarely interact with anything. Only the weak subatomic force affects them, and this weak interaction enables them to travel great distances through matter without being affected by it. The three types, or flavors, of neutrinos are the electron neutrino, designated v_e, the muon neutrino, denoted by v_μ, and the tau neutrino, denoted v_τ. Each type also has a corresponding antiparticle, called an antineutrino and denoted by a bar above the symbol, or by $\bar{v}_e, \bar{v}_\mu, \bar{v}_\tau$.

Neutrinos apparently have an identity crisis! Each type of neutrino is not completely distinct, and the different types can be transformed into one another. In the language of quantum mechanics, neutrinos do not occupy a well-defined state; they instead consist of a combination or mixture of states. As neutrinos move through space, the states come in and out of phase with one another, so the neutrinos change form with time.

The effect is called *neutrino oscillation* because the probability of metamorphosis between neutrino types has a sinusoidal, in and out, oscillating dependence on path length. The change in identity is not one way, for a neutrino of one type can change into another kind of neutrino and back again as it moves along. The three possible types of neutrinos are called electron neutrinos, muon neutrinos, and tau neutrinos, each named for the type of particle it interacts with.

In 1967, the Italian atomic physicist Bruno Pontecorvo (1913–1993) proposed that one type of neutrino might transform, or oscillate, into another type in the vacuum of space, and two years later, Pontecorvo and Vladimir Gribov (1930–1997), proposed that the Solar Neutrino Problem could be explained if solar neutrinos switch from electron neutrinos to another type as they travel in the near vacuum of space from the Sun to Earth, thereby escaping detection (Gribov and Pontecorvo 1969). Almost a decade later, the American physicist Lincoln Wolfenstein (1923–) showed that the neutrinos could oscillate, or change type, more vigorously by interacting with matter, rather than in a vacuum (Wolfenstein 1978), and the Russian physicists, Stanislav P. Mikheyev (1940–) and Alexei Y. Smirnov (1951–) subsequently explained how the matter oscillations might explain the Solar Neutrino Problem (Mikheyev and Smirnov 1985).

The theory, named the MSW effect after the first letters of the last names of the scientists who developed it, proposed that the electron neutrinos generated in the solar core could change type on their way out of the Sun, and therefore remain invisible to the first solar neutrino detectors.

Such a transformation was suggested first by observations of nonsolar neutrinos using the Super-Kamiokande detector, which replaced the older, nearby Kamiokande instrument in 1996. Super-Kamiokande can observe both solar electron neutrinos and atmospheric muon neutrinos. The former are created by nuclear fusion at the center of the Sun, whereas the latter are created when fast-moving cosmic rays enter the Earth's atmosphere from outer space. Solar electron neutrinos are distinguished by their relatively low energy, near the 5 MeV lower threshold of the detector. A high energy of 1,000 MeV is typical of an atmospheric muon neutrino. Neutrinos of higher energy produce a tighter cone of light, so a solar electron neutrino makes a fuzzy, blurred and ragged light pattern, while an atmospheric muon neutrino produces a neat, sharp-edged ring of light.

After monitoring light patterns for more than 500 days, the Super-Kamiokande scientists reported that there were roughly twice as many muon neutrinos coming from the atmosphere directly over the Super-Kamiokande detector than those coming from the other side of the Earth (Fukada et al. 1998a, b). The muon neutrinos are produced in the atmosphere above every place on our planet, but

some of them apparently disappeared while traveling through the Earth to arrive at the detector from below.

Subsequent experiments using neutrinos generated by particle accelerators on the Earth confirmed the effect (Eguci et al. 2003). They suggest that although all the neutrinos produced by nuclear reactions in the Sun are electron neutrinos, they do not stay that way. Nevertheless, the terrestrial neutrinos did not come from the Sun and are not directly related to nuclear fusion reactions there. So the solution to the Solar Neutrino Problem was not known definitely until 2001, when a new underground solar neutrino detector in Canada, the Sudbury Neutrino Observatory, demonstrated that solar neutrinos are changing type when traveling to the Earth.

The Sudbury Neutrino Observatory, abbreviated SNO and pronounced "snow", is located 2 km underground in a working nickel mine near Sudbury Ontario. It is a water detector, but unlike Kamiokande or Super-Kamiokande, the SNO detector contains heavy water.

Heavy water is chemically similar to ordinary water, and it does not appear or taste any different. In fact, heavy water exists naturally as a constituent of ordinary tap or lake water in a ratio of about 1 part in 7,000; expensive chemical and physical processes can separate it.

The hydrogen in heavy water has a nucleus, called a deuteron, which consists of a proton and a neutron. For ordinary water, the hydrogen is about half as light, with a nucleus that contains only a proton and no neutron. It is the heavier deuteron that makes SNO sensitive to not just one type of neutrino but instead to all three known varieties.

One thousand tons, or 1 million liters, of heavy water, was placed in a central spherical cistern with transparent acrylic walls. A geodesic array of about 10,000 photo-multiplier tubes surrounds the vessel to detect the flash of light given off by heavy water when it is hit by a neutrino. Both the light sensors and the central tank are enveloped by a 7,800-ton jacket of ordinary water (Fig. 8.5), to shield the heavy water from emissions of the underground rocks. As with other neutrino detectors, the overlying rock blocks energetic particles generated by cosmic rays.

The Sudbury Neutrino Observatory can be operated in two modes: one sensitive only to electron neutrinos and the other equally sensitive to all three types of neutrinos. Observations with both modes have confirmed that the Solar Neutrino Problem is caused by changes in the neutrinos as they travel from the solar core. When this is taken into account, the total number of electron neutrinos produced in the Sun is as predicted (McDonald 2005). Haxton et al. (2013) have provided a recent review of solar neutrinos.

8.5 How the Energy Gets Out

All of the Sun's nuclear energy is created deep down inside its high-temperature core, and no energy is created in the cooler regions outside of it. The energy-generating core extends to about one quarter of the distance from the center of the

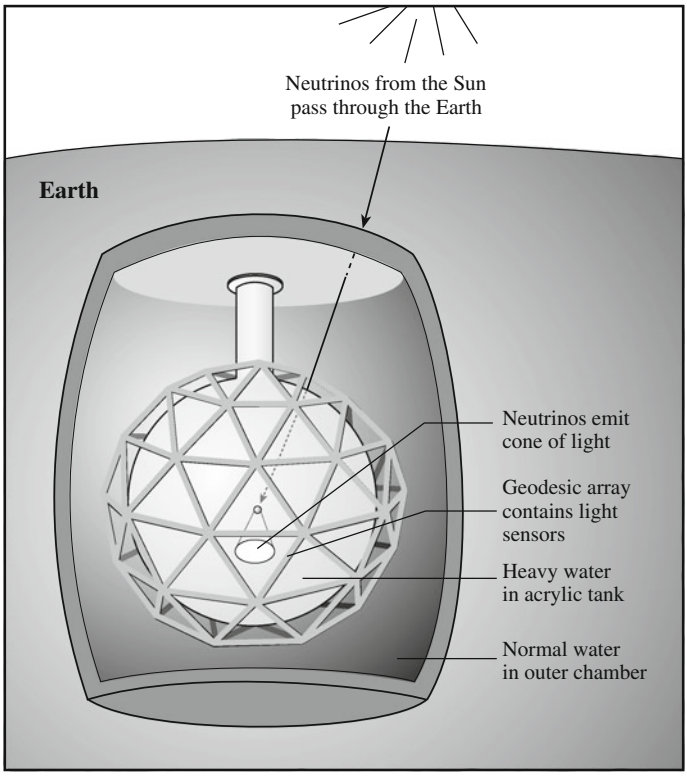


Fig. 8.5 How Sudbury works Neutrinos from the Sun travel through more than 2 km of rock, entering the acrylic tank of the Sudbury Neutrino Observatory, which contains 1,000 tons (1 million liters) of heavy water. When one of these neutrinos interacts with a water molecule, it produces a flash of light that is detected by a geodesic array of photo-multiplier tubes. Some 7,800 tons (7.8 million liters) of ordinary water surrounding the acrylic tank blocks radiation from the rock, and the overlying rock blocks energetic particles generated by cosmic rays in our atmosphere. The heavy water is sensitive to all three types of neutrinos

Sun to the visible solar disk, accounting for only 1.6 % of the Sun's volume. However, about half of the Sun's mass is packed into its dense core.

Because we cannot see inside the Sun, astronomers combine basic theoretical equations, such as those for equilibrium and energy generation or transport, with observed boundary conditions, such as the Sun's mass and luminous output, to create models of the Sun's internal structure. These models consist of two nested spherical shells that surround the hot, dense core (Fig. 8.6).

The innermost shell, called the radiative zone, extends from the core to 71.3 % of the Sun's radius. As the name implies, energy moves through this region by radiation. The outermost layer is known as the convective zone, where energy is transported in a churning, wheeling motion called convection.

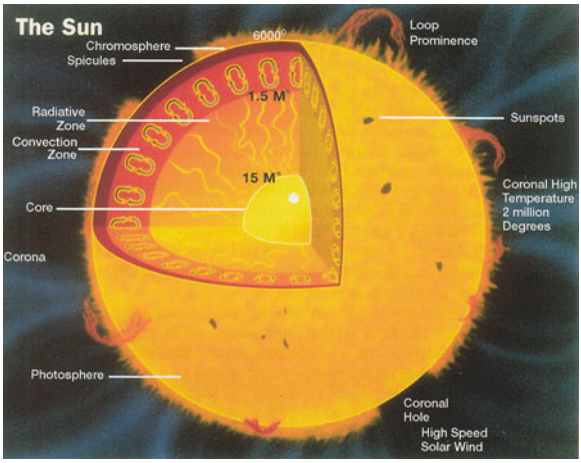


Fig. 8.6 Anatomy of the Sun The Sun is an incandescent ball of ionized gas powered by the fusion of hydrogen in its core. As shown in this interior cross-section, energy produced by nuclear fusion is transported outward, first by countless absorptions and emissions within the radiative zone and then by convection. The visible disk of the Sun, called the *photosphere*, contains dark sunspots, which are Earth-sized regions of intense magnetic fields. A transparent atmosphere envelops the photosphere, including the low-lying chromosphere with its jet-like spicules and the 1 million-degree corona that contains holes with open magnetic fields, the source of the high-speed solar wind. Loops of closed magnetic fields constrain and suspend the hot million-degree gas within coronal loops and cooler material in prominences

Radiation does not move quickly through the solar interior. A single gamma ray produced by nuclear fusion in the core of the Sun cannot move even a fraction of a millimeter before encountering a subatomic particle, where the radiation is scattered or absorbed and reemitted with less energy. This radiation quickly interacts with another particle in the radiative zone and is eventually reradiated at yet lower energy. The process continues again countless times as the radiation moves outward on a haphazard, zigzag path, steadily losing photon energy at each encounter.

Example: Scattering of radiation inside the Sun
Free electrons, which are not attached to atoms, scatter radiation with a Thomson scattering cross section, σ_T, given by Thomson (1903) (Sect. 2.7):

$$\sigma_T = \frac{8\pi}{3} r_e^2 = \frac{8\pi}{3}\left[\frac{e^2}{4\pi\varepsilon_0 m_e c^2}\right]^2 = 6.6525 \times 10^{-29}\ \mathrm{m}^2, \tag{8.58}$$

where $r_e = 2.8179 \times 10^{-15}$ m is the classical electron radius, $e = 1.6022 \times 10^{-19}$ C is the fundamental unit of charge, $\varepsilon_0 = 8.854 \times 10^{-12}$ F m^{-1} is the electric constant, and $c = 2.9989 \times 10^8$ m s^{-1} is the speed of light.

A radiation photon moving into plasma with an electron number density of N_e will travel a mean free path, l, before encountering an electron, where (Clausius 1858) (Sect. 5.2)

$$l = \frac{1}{N_e \sigma_T}. \tag{8.59}$$

Hydrogen is by far the most abundant element in the Sun, and throughout the solar core and radiation zone it is completely ionized into free protons and free electrons. In the solar core, the mass density $\rho \approx 1.5 \times 10^5$ kg m^{-3}, and since it is the protons that contribute the vast majority of this mass the proton number density is $N_P = \rho / m_P \approx 10^{32}$ m^{-3}, where the proton mass $m_P = 1.6726 \times 10^{-27}$ kg. (The proton is 1836 times more massive than the electron, so we can ignore the electrons' contribution to the mass.) That is, there are 100 million trillion trillion protons per cubic meter at the center of the Sun, so it is not surprising that they don't move very far before colliding with one another. Since each former hydrogen atom contained one proton and one electron $N_P = N_e$. Within the core, a photon will only travel a length $l = 1/(N_e \sigma_T) \approx 1.5 \times 10^{-4}$ m before encountering an electron. Further out, the mass density is lower, but even at the mean solar mass density of 1,409 kg m^{-3}, we still obtain average values of $N_e \approx 8.4 \times 10^{29}$ m^{-3} and $l = 0.018$ m. If the radiation was always headed straight out of the Sun, it would make about 2×10^{10}, or 20 billion, collisions in working its way across the radiation zone, whose thickness is about half the solar radius, or about 3.5×10^8 m.

Each photon will, however, be scattered in a random direction, which will generally be different from the outward direction. The decrease in temperature with increasing distance from the Sun center still assures that more radiation moves outward than inward, just as heat normally flows from a hotter region to a colder one. The radiation therefore follows the path of least resistance, heading for regions of lower density and temperature.

The total time for the radiation to diffuse through the radiation zone is a random walk problem, with a lengthening step, or path length, at larger distances from the solar center. The photon spends most of its time close to the core where the mean free path is shortest, and the average step length required to reach the inner edge of the convective zone is $l = 9.0 \times 10^{-4}$ m. The diffusion time for the radiation to move from the bottom to the top of the radiative zone is about 170,000 years (Mitalas and Sills 1992).

As a result of this continued ricocheting and innumerable collisions in the radiative zone, it takes about 170,000 years, on average, for radiation to work its way out from the Sun's core to the bottom of the convective zone (Mitalas and Sills 1992), where the temperature has become cool enough for heavy nuclei to capture electrons and form atoms that absorb radiation. These atoms block the

outward flow of radiation like dirt on a window, and the radiation heats the bottom of the convective zone.

This material becomes hotter than it otherwise would be, and it must find a way to release the pent-up energy. In response to heating from below, gases in the bottom layer of the convective zone expand, thereby becoming less dense than the gas in the overlying layers. Due to its low density, the heated material rises to the visible solar disk in about 10 days and then cools by radiation. The cooled gas then sinks because it is denser than the hotter gas, only to be reheated and rise again (Focus 8.6). Such convective motions can occur whenever a layer of fluid is heated from below (Jeffreys 1926), as in a kettle of boiling water or a simmering pot of oatmeal, with hot rising bubbles and cooler sinking material.

Focus 8.6 Convection

Convection is a method of transferring heat from hotter to cooler regions within a gas or liquid, and when it occurs in a star, energy is transported within the stellar interior by a wheeling gas motion. When a region of high density is displaced upward into a region of lower density and pressure, convection will take place if the displaced volume expands and becomes less dense than its surroundings. It will then continue to be buoyed up like a balloon or bubbles in a boiling pot of water.

For an adiabatic expansion in which no heat is exchanged with the new surroundings, convection will occur if the structural temperature gradient of the star is greater than the adiabatic gradient, or when (Schwarzschild 1906):

$$\left(\frac{dT(r)}{dr}\right) > \frac{\gamma - 1}{\gamma}\frac{T(r)}{P(r)}\frac{dP}{dr}, \tag{8.60}$$

where the adiabatic index $\gamma = 5/3$ for ionized hydrogen, $T(r)$ is the gas temperature at radius r, the dT/dr is the gas temperature gradient in the radial, r, direction, and $P(r)$ is the gas pressure at radius r given by the ideal gas law, in which $P =$ constant $\times\ \rho T$, for a mass density ρ.

The structural temperature gradient in a star is given by Eddington (1917)

$$\frac{dT(r)}{dr} = -\frac{3}{4ac}\frac{\kappa\rho}{T^3}\frac{L(r)}{4\pi r^2}, \tag{8.61}$$

where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}, the Rosseland mean opacity is κ, and $L(r)$ is the radiation luminosity at radius r.

If the temperature in a star falls fast enough with increasing radius, convection sets in. This happens in the outer, cooler layers of stars like the Sun, where the opacity from heavy elements becomes high enough to produce a steep temperature gradient.

In stars that are more massive than the Sun, the nuclear fusion of hydrogen into helium occurs by the CNO cycle rather than the proton–proton chain (see Sect. 10.3), and the CNO process is a very sensitive function of

temperature. As a result, the centers of these stars are very hot, but the temperature falls off rapidly with distance from the center and convection occurs in the stellar core.

Chandrasekhar (1961) has written a comprehensive book that includes convective instability of a layer of gas heated from below. The conditions for the instability can be expressed in terms of a Rayleigh number (Rayleigh 1916). Convective energy transport in stars is described by the mixing-length theory of convection (Böhm-Vitense 1953, 1958), Schwarzschild (1958), and standard textbooks of stellar astrophysics). Galloway and Weiss (1981), Wilson (1966, 1978), and Spiegel (1971) have also discussed convection in stars.

The convective zone is capped by the photosphere, the place where the gaseous material changes from being completely opaque to being transparent to radiation. In the photosphere, a process of absorption and reemission of radiation carries the Sun's energy out. Rupert Wildt (1905–1976) explained the detailed observations of sunlight by showing that both hydrogen atoms and negative hydrogen ions absorb radiation in the photosphere (Wildt 1939). Collisions between unionized, or neutral, hydrogen atoms and free electrons lead to the formation of the negative hydrogen ions. Despite their low concentration, they provide the absorption and extra opacity needed to account for the sunlight that escapes from the photosphere.

As first noticed by William Herschel (1738–1822), the photosphere contains a fine granular pattern (Herschel 1801). These closely packed granulation cells now can be examined using high-resolution images taken from ground-based telescopes under conditions of excellent observation (Fig. 8.7) or from spacecraft located outside of the Earth's obscuring atmosphere. The images reveal a host of granules with bright centers surrounded by dark lanes, exhibiting a non-stationary, overturning motion caused by the underlying convection.

The bright center of each granule, or convection cell, is the highest point of a rising column of hot gas. The dark edges of each granule are the cooled gas, which sinks because it is denser than the hotter gas. Each individual granule lasts only about 15 min before it is replaced by another one, never reappearing in precisely the same location.

The mean angular distance between the bright centers of adjacent granules is about 2.0 s of arc, corresponding to about 1,500 km at the Sun. That seems very large, but an individual granule is about the smallest thing you can see on the Sun when peering through our turbulent atmosphere.

There are at least a million granules on the visible solar disk at any moment. They are constantly evolving and changing, producing a honeycomb pattern of rising and falling gas that is in constant turmoil, bubbling away and completely changing on time-scales of minutes.

The granules are superimposed on a larger cellular pattern, called the supergranulation, studied at the California Institute of Technology by Robert B.

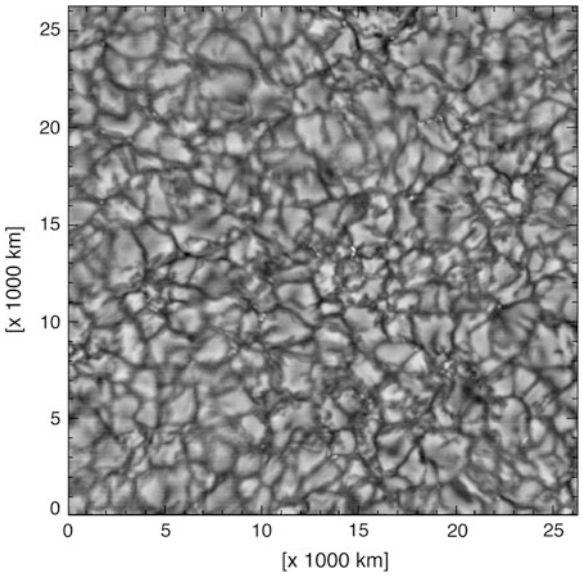


Fig. 8.7 The solar granulation Underlying convection shapes the photosphere, producing tiny, varying regions called *granules*. They are places where hot and therefore bright material reaches the visible solar disk. The largest granules are approximately 1,400 km across. They are not circular but rather angular in shape. This honeycomb pattern of rising (*bright*) and falling (*dark*) gas is in constant turmoil, completely changing on time-scales of minutes and never exactly repeating itself. This image was taken with exceptional angular resolution of 0.2 s of arc, or 150 km, at the Sun using the National Solar Observatory's Vacuum Tower Telescope at the Sacramento Peak Observatory. (Courtesy of Thomas R. Rimmele/AURA/NOAO/NSF.)

Leighton (1919–1997) and his collaborators in the early 1960s. They subtracted a long-wavelength image of the Sun from a short-wavelength one, revealing a supergranulation pattern of horizontal flow, each supergranulae is an estimated 16,000 km across, or almost three times as large as the Earth, with a lifetime of roughly 24 h. Because the motion is predominantly horizontal, the supergranules were not detected when looking directly at the center of the solar disk, but further out toward the sides of the round solar disk, where the horizontal motion is partially directed along the line of sight. Leighton (1963) has provided a review of the solar granulation.

Roughly 3,000 supergranules are seen on the visible solar disk at any moment. And like the ordinary granulation, the changing pattern of supergranulation is caused by convection. But unlike the granules, whose gases move up and down, the material in each supergranule cell rises in the center, and exhibits a sideways motion as it moves away from the center with a typical velocity of about 0.4 km s^{-1}. Only after this prolonged horizontal motion does the material eventually sink down again at the cell boundary. The supergranular flow carries the magnetic field across the photosphere, sweeping the magnetism to the edges of the

supergranulation cells where it collects, strengthens and forms a network of concentrated magnetic field.

When studying the supergranulation in the 1960s, Leighton and his co-workers unexpectedly discovered vertical up and down motions in the subtracted difference between long-wavelength and short-wavelength solar images. They exhibited a periodic oscillation with a period of about five minutes (Leighton 1961; Leighton et al. 1962; Noyes and Leighton 1963). These oscillations have subsequently been used to investigate the unseen depths of the Sun (Focus 8.7).

Focus 8.7 Helioseismology

Vigorous turbulent motion in the convective zone produces sound waves (Goldreich and Kumar 1990), which drive five-minute oscillations in the overlying photosphere (Ulrich 1970; Leibacher and Stein 1971; Deubner 1975). Each five-minute period is the time it takes for the localized motion to change from moving outward to moving inward and back outward again. Such five-minute oscillations are imperceptible to the unaided eye, for the photosphere moves a mere hundred-thousandth (0.00001) times the solar radius, but they can be detected using the Doppler effect of a single absorption line formed in the photosphere. Deubner and Gough (1984) provided a review of helioseismology at that stage of its development. More recent accomplishments of helioseismology are included with references to the relevant research papers in Lang (2009).

The information obtained from oscillations produced by sound waves that traveled to various levels within the Sun can be combined to create a picture of the Sun's large-scale internal structure (Fig. 8.8). The technique is known as helioseismology, a hybrid name combining the Greek words *Helios* for the "Sun" and *seismos* for "earthquake" or "tremor."

Observations from space, where night never falls, provide the best data for helioseismology. The *SOlar and Heliospheric Observatory*, abbreviated *SOHO*, has provided them. Instruments aboard this spacecraft have observed the solar oscillations 24 h a day, every day for more than ten years. By considering a sequence of waves with longer and longer wavelengths, that penetrate deeper and deeper, the radial profile of the sound speed has been determined and used to establish the lower boundary of the convective zone, at a radius of 71.3 % of the radius of the Sun.

Rotation imparts a clear signature to the oscillation periods, lengthening them in one direction and shortening them in the other. These opposite effects make the oscillation periods divide, and such rotational splitting depends on both depth and latitude within the Sun. The helioseismological observations indicate that differential rotation, in which the equator spins faster than the poles, is preserved throughout the convective zone, but it disappears in the radiative zone that rotates at one speed (Sect. 4.3).

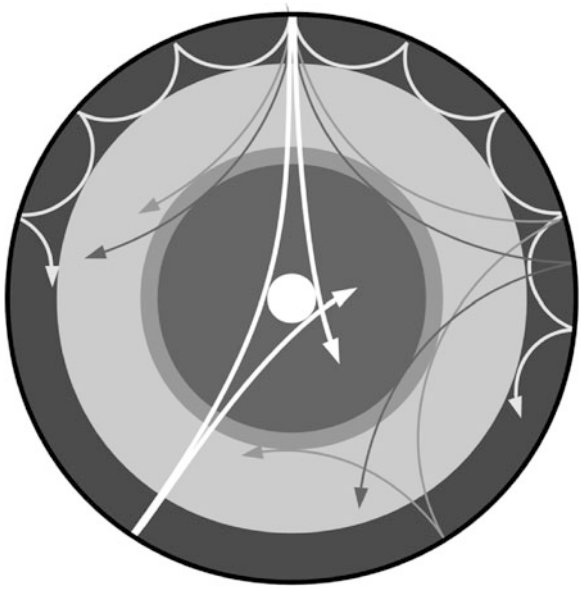


Fig. 8.8 Sound paths in the Sun The trajectories of sound waves are shown in a cross section of the solar interior. The rays are bent inside the Sun, like light within the lens of an eye. They circle the solar interior in spherical shells called *resonant cavities*. Each shell is bounded at the top by a large, rapid density drop near the photosphere and bounded at the bottom at an inner turning point where the bending rays undergo total internal refraction due to the increase in sound speed with depth inside the Sun. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

8.6 The Faint-Young-Sun Paradox

The Sun has grown slowly in luminous intensity since it formed; with a steady, inexorable brightening that is a consequence of the increasing amount of helium accumulating in the Sun's core. As the hydrogen in the Sun's center slowly depletes, and is steadily replaced by heavier helium, the core must continue producing enough pressure to prevent the Sun from collapsing. The only way to maintain the pressure and keep supporting the weight of a heavier material is to increase the central temperature. As a result of the slow rise in temperature, the rate of nuclear fusion gradually increases and so does the Sun's luminosity. The Sun is, for example, now 30 % more luminous than it was 4.6 billion years ago.

The Sun's luminosity increases as time goes on, so of course the Sun was significantly dimmer in the remote past. Therefore, the Earth should have been noticeably colder then. However, this does not agree with geological evidence. Assuming an unchanging terrestrial atmosphere, with the same composition and reflecting properties as today, the lower solar luminosity in the past would have caused the Earth's global surface temperature to be below the freezing point of

water during the planet's first 2.6 billion years. The oceans would have been frozen solid, there would have been no liquid water, and the entire planet would have been locked into a global ice age.

Yet, sedimentary rocks, which must have been deposited in liquid water, date back to a time when the Earth was less than 800 million years old. There is fossil evidence in those rocks of living things at about that time. Thus, for billions of years, the Earth's surface temperature was not very different from today; conditions have remained hospitable for life on the Earth throughout most of the planet's history.

There are several possible explanations for the discrepancy between the Earth's warm climatic record and an initially dimmer Sun, which is 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 (Sagan and Chyba 1997). Greater amounts of carbon dioxide would enable the early atmosphere to trap more solar heat near the Earth's surface, warming it by the greenhouse effect, which would prevent the oceans from freezing. Another possibility is that the Sun was more magnetically active in its youth, expelling strong winds, energetic particles, and radiation that might have kept the Earth warm (Schilling 2001; Sackmann and Boothroyd 2003; Minton and Malhotra 2007).

What about the future? In only 1 billion years the Sun will have brightened by another 10 %. Calculations suggest that the Earth's oceans could evaporate then at a rapid rate, resulting in a hot, dry, uninhabitable Earth. In about 3 billion years from now, the Sun will then be hot enough to boil the Earth's oceans away, leaving the planet a burned-out cinder, a dead and sterile place.

8.7 The Sun's Destiny

The Sun cannot shine forever, because eventually it will deplete the hydrogen fuel in its core. Although it has converted only a trivial part of its original mass into energy, the Sun has processed a substantial 37 % of its core hydrogen into helium in the past 4.6 billion years. There will be no hydrogen left in the solar core about 7 billion years from now. When that hydrogen is exhausted, the central part of the Sun will undergo a slow collapse, and the gradually increasing core temperature will cause the outer layers of the Sun to expand into a red giant star, with a dramatic increase in size and a powerful rise in luminosity (Sackmann et al. 1993). Eventually the Sun will become 170 times larger and 2,300 times more luminous than it is now (Fig. 8.9). This will result in a substantial rise in temperatures throughout the solar system, becoming hot enough to melt the Earth's surface.

Meanwhile, the core of the Sun will continue to contract until the central temperature is hot enough to ignite helium – which is at about 100 million K. However, this conversion of helium into carbon will not last long compared to the Sun's 12 billion years of hydrogen burning. In about 35 million years, the core helium will have been used up and there will be no heat left to hold up the Sun. In

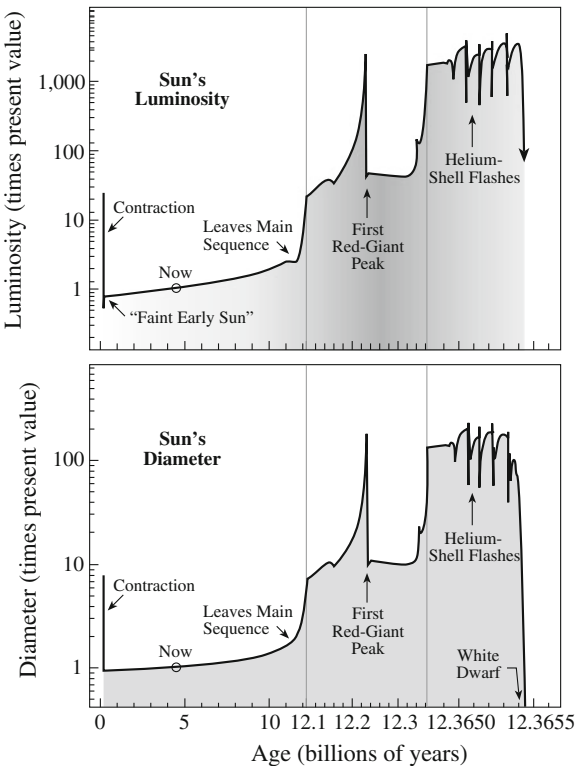


Fig. 8.9 The Sun's fate In about 7 billion years, the Sun will become much brighter (*top*) and larger (*bottom*). The time-scale is expanded near the end of the Sun's life to show relatively rapid changes. (Courtesy of I-Juliana Sackmann and Arnold I. Boothroyd.)

a last spurt of activity, the Sun will shed the outer layers of gas to produce an expanding "planetary" nebula around the star, and the core will collapse into a white dwarf star (see Sects. 13.1, 13.2).

By this time, the intense winds will have stripped the Sun down to about half of its original mass, and gravitational collapse will squeeze the remaining part to about the size of the present-day Earth. Nuclear reactions then will be a thing of the past, and there will be nothing left to warm the Sun or planets. The former Sun will gradually cool down and fade away, plunging all of the planets into a deep freeze.

Such events are in the very distant future, of course; but even now, the Sun threatens the Earth with its perpetually expanding atmosphere that envelops our planet and with explosive outbursts that can send energetic particles, intense radiation, and huge magnetic bubbles toward the Earth.

Chapter 9
The Extended Solar Atmosphere

This chapter discusses the Sun's outer atmosphere, the million-degree corona, which expands away from the Sun in fast, uniform winds and slow, gusty ones. Solar flares and coronal mass ejections are also reviewed together with the space-weather effects of these solar explosions on the Earth and nearby space. A complete, in-depth treatment of all of these topics, with numerous references, can be found in Lang, Kenneth R., *The Sun From Space*, *Second Edition*, Heidelberg: Springer-Verlag 2009.

9.1 Hot, Volatile, Magnetized Gas

9.1.1 The Million-Degree Solar Corona

The apparent edge of the visible solar disk, the photosphere, is illusory, for a hot, transparent atmosphere envelops it, extending all the way to the Earth and beyond. This unseen atmosphere is more rarefied than the best vacuum on Earth, and so tenuous that we see right through it.

The diaphanous outer atmosphere of the Sun includes – from its deepest part outward – the underlying photosphere, from the Greek word *photos* for "light"; the thin chromosphere, from the Greek word *chromos* for "color"; and the extended *corona* from the Latin word for "crown." We can observe the chromosphere and corona during a total solar eclipse, when the Moon blocks out the intense light of the underlying photosphere (Fig. 9.1).

Because of their very low densities and high temperatures, the chromosphere and corona produce bright spectral features called emission lines. Atoms and ions in a hot tenuous gas produce such emission features, heated to incandescence and shining at precisely the same wavelengths as the dark absorption lines produced by the same substance in the cooler photosphere. The corona's emission lines provided the initial evidence that it is hundreds of times hotter than the underlying photosphere.

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_9, © Springer-Verlag Berlin Heidelberg 2013

Fig. 9.1 Eclipse corona The million-degree solar atmosphere, known as the corona, was seen around the shadowed disk of the Moon during the solar eclipse on 11 July 1991. The electrically charged gas was concentrated by magnetic fields into numerous fine rays as well as larger helmet streamers. The expanding corona envelops the Earth and all the other planets. (Courtesy of HAO/NCAR.)

The corona's emission lines were first observed during total eclipses of the Sun (Table 9.1). The intense green line, at a wavelength of 530.3 nm, for example, was first observed during the solar eclipse of August 7, 1869, and for decades attributed to a previously unknown substance dubbed coronium (Young 1869). About 70 years later, it was attributed to emission of iron ions, denoted by Fe XIV, by Walter Grotrian (1890–1954) of Potsdam and Bengt Edlén (1906–1933), a Swedish astronomer who specialized in spectroscopy (Grotrian 1934, 1939; Edlén 1941, 1945). These ions are iron atoms missing 13 of their 26 electrons.

The reason it took so long to identify the coronal emission lines is that no one realized the corona was so hot and also because such spectral features can arise only in the very tenuous corona. They are "forbidden transitions" that do not occur in terrestrial circumstances where collisions between atoms keep them from happening even in the best vacuum.

Iron must be at a temperature of a few million K for atomic collisions to remove so many electrons from the atoms. Edlén provided additional evidence for this hot temperature from the observed widths of the emission lines (Edlén 1941). Elements move at a faster speed in a hotter gas, broadening the observed spectral features as well as producing them. The million-degree temperature of the corona was subsequently confirmed by observations of the Sun's radio radiation (Pawsey 1946) and intense x-ray radiation.

Table 9.1 Strong forbidden emission lines in the visible light of the Sun's low corona

Wavelength (nm)	Ion	Name	Wavelength (nm)	Ion
338.8	Fe XIII		670.2	Ni XV
423.2	Ni XII		789.2	Fe XI
530.3	Fe XIV	Green line	802.4	Ni XV
569.4	Ca XV	Yellow line	1074.7	Fe XIII
637.4	Fe X	Red line	1079.8	Fe XIII

[a] Adapted from Edlén (1941) and Swings (1943). The symbols Ca, Fe and Ni denote, respectively, Calcium, Iron and Nickel. Subtract one from the Roman numeral to obtain the number of missing electrons. Thus, the ion Fe XIII is an iron atom missing 12 electrons. The wavelength is in units of nanometers, or 1 nm $= 10^{-9}$ m. Astronomers have often used the Ångström unit of wavelength, where 1 Ångström = 1 Å = 0.1 nm $\times$ 10^{-10} m

Example: **The million-degree corona**

The identification of an emission line of the solar corona with Fe XIV indicated that the coronal gas would have to be very hot. We can estimate how hot that would be by equating the thermal energy of the gas, $3kT/2$, at temperature T, to the ionization potential of Fe XIV, which is 235.04 eV; this is the amount of energy needed to remove so many electrons from the iron atom. Using $k = 1.38065 \times 10^{-23}$ J K^{-1} for the Boltzmann constant and the conversion of 1 eV $= 1.602 \times 10^{-19}$ J, we obtain $T = 2 \times$ ionization potential$/(3k) \approx 1.82 \times 10^6$ K, or about 2 million K.

Withbroe and Noyes (1977) provided a review of mass and energy flow in the solar chromosphere and corona. Aschwanden (2006) discussed the physics of the solar corona, and Ashwanden, Poland and Rabin (2001) and Lang (2009) have provided reviews of modern observations of the corona.

Because of its high temperature, the corona emits most of its energy and its most intense radiation as x-rays. The x-rays can be used to image the hot corona all across the Sun's face with high spatial and temporal resolution. This is because the Sun's visible photosphere, being so much cooler, produces negligible x-ray radiation and appears dark under the million-degree corona. Since the Sun's x-ray radiation is absorbed totally in the Earth's atmosphere, it must be observed with telescopes lofted into space by rockets or in satellites.

Modern spacecraft obtain full-disk images of the corona at soft x-ray and extreme ultraviolet wavelengths in lines of ionized iron, Fe XVII, Fe IX, Fe X, Fe XIII, Fe XV, Fe XVI and one line of ionized helium, He II; these are the permitted lines emitted by ionized atoms, sensitive to temperatures from 60,000 to 4.0×10^6 K (Table 9.2), and not the forbidden lines detected at visible wavelengths.

Close inspection of the Sun's x-ray radiation shows that the star is in constant turmoil, driven by intense, variable magnetic fields. This magnetism is responsible for dark sunspots that temporarily mark the visible face of the Sun.

Table 9.2 Prominent soft x-ray and extreme ultraviolet emission lines from the Sun's low corona and transition region[a]

Wavelength (nanometers)	Emitting ion	Formation temperature (kelvin)
1.70	Iron, Fe XVII	4,000,000
1.90	Oxygen, O VIII	3,100,000
2.16	Oxygen, O VII	2,000,000
3.37	Carbon, C VI	1,300,000
17.11	Iron, Fe IX	630,000
17.45	Iron, Fe X	1,000,000
18.40	Oxygen, O VI	320,000
19.51	Iron, Fe XII	1,400,000
28.42	Iron, Fe XV	2,100,000
30.38	Helium, He II	60,000
33.54	Iron, Fe XVI	2,500,000
33.61	Iron, Fe XVI	2,500,000
46.52	Neon, Ne VII	630,000
60.98	Magnesium, Mg X	1,300,000
155.0	Carbon, C IV	126,000

[a] Subtract one from the Roman numeral to get the number of missing electrons. The wavelengths are in nanometers, abbreviated nm, where 1 nm $= 10^{-9}$ m. Astronomers sometimes use the Ångström unit of wavelength, abbreviated Å, where 1 Å $= 10^{-10}$ meters $= 0.1$ nm

9.1.2 Varying Sunspots and Ever-Changing Magnetic Fields

The solar corona is permeated by magnetic fields that are generated inside the Sun and rise up through the photosphere into the overlying atmosphere. The strongest magnetism protrudes to blemish the visible Sun with dark, Earth-sized sunspots (Fig. 9.2), which were seen by the unaided human eye up to 3,000 years ago – don't do it, staring at the Sun could burn your eyes.

In the early 20th century, the American astronomer George Ellery Hale (1868–1938) first used the Zeeman effect to show that sunspots are regions of intense magnetism, thousands of times stronger than the Earth's magnetic field (Hale 1908a, b). The intense sunspot magnetism acts as both a valve and a refrigerator, choking off the outward flow of heat and energy from the solar interior and keeping the sunspots cooler and darker than their surroundings.

The strong magnetism exerts a pressure that tends to push apart the magnetic fields; however, by using helioseismology to look under the photosphere astronomers have discovered that flowing material pushes against the magnetic fields of sunspots, holding them in place.

Because the Sun's magnetism is forever changing and is never still, the sunspots are temporary, with lifetimes ranging from hours to months. Moreover, the total number of sunspots varies periodically, from a maximum to a minimum and back to a maximum, in about 11 years (Fig. 9.3). Samuel Heinrich Schwabe (1789–1875), an amateur astronomer in Dessau, Germany, discovered this periodic variation in the mid-nineteenth century (Schwabe 1844). At the maximum in the

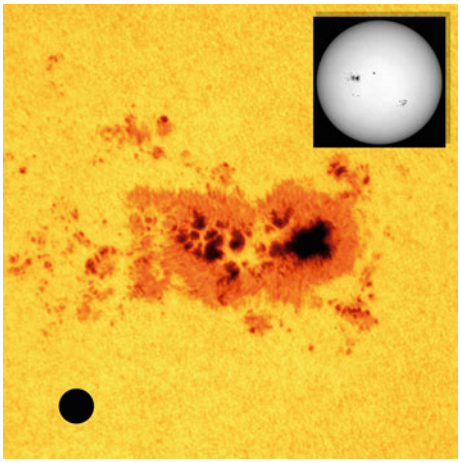


Fig. 9.2 Sunspot group Intense magnetic fields emerge from the interior of the Sun through the Sun's visible disk, the photosphere, producing groups of sunspots. The sunspots appear dark because they are slightly cooler than the surrounding photosphere gas. This composite image was taken in white light; that is, in all of the colors combined. The enlarged image shows the biggest sunspot group, which is about 12 times larger than the Earth, the size of which is denoted by the black spot (*lower left*). (Courtesy of *SOHO*/ESA/NASA.)

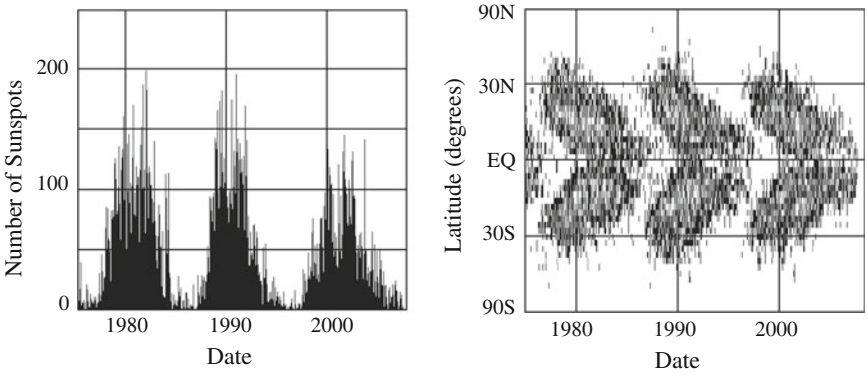


Fig. 9.3 Solar magnetic activity cycle The 11 year solar cycle of magnetic activity is plotted from 1975 to 2007. Both the numbers of sunspots (*left*) and the positions of sunspots (*right*) wax and wane in cycles that peak every 11 years. Similar 11 year cycles have been observed for more than a century. At the beginning of each cycle, the first sunspots appear at about 30° solar latitude and then migrate to 0° solar latitude, at the solar equator (EQ), when the cycle ends. This plot of the changing positions of sunspots resembles the wings of a butterfly, and therefore has been called the *butterfly diagram*. The cycles overlap with spots from a new cycle appearing at high latitudes while the spots from the old cycle persist in the equatorial regions. The solar latitude is the angular distance from the plane of the Sun's equator, which is very close to the plane of the Earth's orbit about the Sun, called the *ecliptic*. (Courtesy of David Hathaway/NASA/MSFC.)

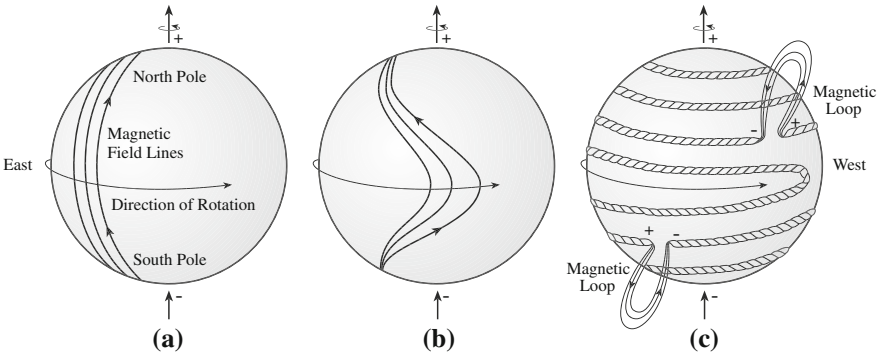


Fig. 9.4 Winding up the field A model for generating the changing location, orientation, and polarity of the sunspot magnetic fields. Initially, the magnetic field is supposed to be the dipolar field seen at the poles of the Sun (*left*). The internal magnetic fields then run just below the photosphere from the Sun's South to North Pole. As time proceeds, the highly conductive, rotating material inside the Sun carries the magnetic field along and winds it up. Because the equatorial regions rotate at a faster rate than the polar regions, the internal magnetic fields are stretched out and wrapped around the Sun's center, becoming concentrated and twisted together like ropes (*middle* and *right*). With increasing strength, the submerged magnetism becomes buoyant, rises and penetrates the visible solar disk – the photosphere – creating magnetic loops and bipolar sunspots that are formed in two belts, one each in the northern and southern hemisphere (*right*). The simplified model shows only two magnetic loops, but many of them are created at about the same time. [Adapted from Horace W. Babcock, "The topology of the Sun's magnetic field, and the 22 year cycle," *Astrophysical Journal* 133, 572–587 (1961).]

sunspot cycle, there may be 100 or more spots on the visible hemisphere of the Sun at one time; at sunspot minimum, very few are seen, and, for periods as long as a month or more, none can be found. The locations where sunspots emerge and disappear also vary over the 11 year sunspot cycle, from mid-latitudes on the Sun to the solar equator (see Fig. 9.3).

The American astronomer Horace W. Babcock (1912–2003) devised a conceptually simple model for the varying sunspots (Babcock 1961). Working with his son, Harold, he had shown that the Sun has a general dipolar magnetic field of about 10^{-4} tesla, usually limited to high solar latitudes near the solar poles (Babcock and Babcock 1955). His dynamo theory begins at sunspot minimum with a global, dipolar magnetic field that runs inside the Sun from south to north, or from pole to pole. Uneven, or differential, rotation – in which the equatorial regions rotate faster than the polar ones – shears the electrically conducting gases of the interior. As a result, the entrained magnetic fields are stretched out and squeezed together. The magnetism is coiled, bunched, and amplified as it is wrapped around the inside of the Sun. The surrounding gas buoys up the concentrated magnetism, and eventually the magnetic fields become strong enough to rise up to the photosphere and break through it in belts of bipolar sunspot pairs (Fig. 9.4).

The initial dipolar magnetic field is twisted into a submerged, ring-shaped field running parallel to the solar equator, or east to west. There are two buried magnetic

fields, one in the northern hemisphere and one in the southern hemisphere, but oppositely directed, which bubble up at mid-latitudes to spawn two belts of sunspots, symmetrically placed on each side of the equator.

As the 11 year cycle progresses toward maximum activity, the internal magnetic field is wound increasingly tighter by the shearing action of differential rotation. The two sunspot belts slowly migrate toward the solar equator, where the sunspots in the two hemispheres tend to merge.

Diffusion and poleward flows sweep the remnant magnetism into streams, each dominated by a single magnetic polarity, that slowly wind their way from the low- and mid-latitude belts to the Sun's poles. By sunspot minimum, the continued poleward transport of their debris may form a global dipole with reversed polarity. The north and south poles switch magnetic direction or polarity at the next sunspot minimum. When the Sun's magnetic flip is considered, we see that it takes two activity cycles, or about 22 years, for the overall magnetic polarity to return to where it began. The internal magnetism then has readjusted to its submerged dipolar form, and the magnetic cycle begins again.

9.1.3 Coronal Loops

Magnetic fields are described by lines of force, like those joining the opposite poles of a bar magnet. The direction of the lines of force and the orientation of the magnetic fields can be inferred from the polarization of the spectral lines that have been split by the Zeeman effect. Magnetic-field lines pointing out of the Sun have positive magnetic polarity, whereas inward-directed fields have negative polarity.

Sunspots usually appear in adjacent pairs or other close groupings of opposite magnetic polarity (Hale 1919). Invisible magnetic arches loop between these oppositely directed magnetic regions, often emerging from a sunspot with one polarity and reentering a neighboring sunspot of opposite polarity. Although they remain unseen in optically visible sunlight, these coronal loops shine brightly in x-ray and extreme ultraviolet images of the Sun taken with telescopes in space (Fig. 9.5). Because this radiation is absorbed in our atmosphere, such images cannot be obtained from the ground. Material is concentrated to higher densities and temperatures within these loops, so they emit this invisible radiation more intensely than their surroundings. This intense x-ray and extreme ultraviolet emission thus outlines the magnetic shape and structure of the Sun's outer atmosphere, indicating that the corona is stitched together by bright, thin magnetized loops.

The magnetized atmosphere in, around, and above bipolar sunspot groups is called a solar active region. Active regions are places of concentrated, enhanced magnetic fields, sufficiently large and strong to stand out from the magnetically weaker areas. These disturbed regions are prone to awesome explosions, marking a location of extreme unrest on the Sun.

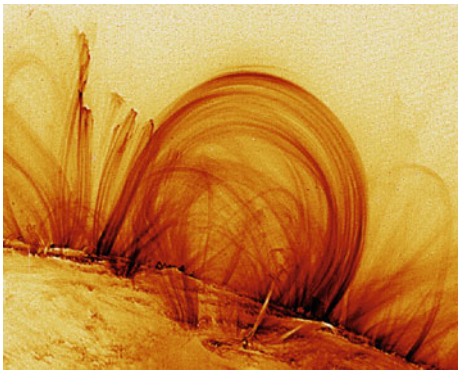


Fig. 9.5 Magnetic loops made visible An electrified, million-degree gas, known as *plasma*, is channeled by magnetic fields into bright thin loops. The magnetized loops stretch up to 500,000 km from the visible solar disk, spanning up to 40 times the diameter of planet Earth. The magnetic loops are seen in the extreme ultraviolet radiation of eight and nine times ionized iron, denoted Fe IX and Fe X, formed at a temperature of about 1.0 million K. The hot plasma is heated at the bases of loops near the place where their legs emerge from and return to the photosphere. Bright loops with a broad range of lengths all have a fine thread-like substructure with widths as small as the telescope resolution of 1 s of arc, or 725 km at the Sun. This image was taken with the *Transition Region And Coronal Explorer* (*TRACE*) spacecraft. [Courtesy of the *TRACE* consortium, LMSAL and NASA; *TRACE* is a mission of the Stanford-Lockheed Institute for Space Research, a joint program of the Lockheed-Martin Solar and Astrophysics Laboratory (LMSAL), and Stanford's Solar Observatories Group.]

Example: **Coronal loops in solar active regions**

Intense magnetic fields of strength $B \approx 0.03$ tesla confine a hot ionized gas in solar active regions within coronal loops. The gas pressure $P_g = N_e kT$ of the hot electrons, of number density N_e, is just equal to the magnetic pressure $P_B = B^2/(2\mu_0)$ required to confine the hot plasma when $N_e \approx 2.6 \times 10^{19}$ m^{-3}, assuming a temperature of $T = 1$ million or 10^6 K and using a Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and the permeability of free space $\mu_o = 4\pi \times 10^{-7}$ N A^{-2} $= 1.2566 \times 10^{-6}$ N A^{-2}. In other words, a million-degree gas with an electron density less than 10^{19} m^{-3} will be constrained by these coronal loops.

If these loops expel their electrons during a solar flare, the electrons will be accelerated to higher energies by the magnetic interaction that triggers and powers the flare (see Sect. 9.3). The energy of the electrons might be comparable to that of x-rays, with $E = 30$ keV $= 4.80 \times 10^{-15}$ J. For a coronal loop of radius R the volume can be approximated as $4\pi R^3/3$, and for a loop radius comparable to the size of the Earth, or $R = 6.378 \times 10^6$ m, the total flare energy released will be $E_f = 4\pi R^3 N_e E/3 \approx 1.25 \times 10^{26}$ J.

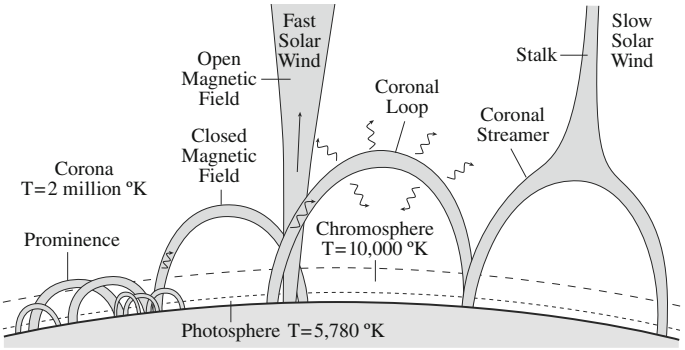


Fig. 9.6 Coronal loops The corona is stitched together with ubiquitous coronal loops that are created when upwelling magnetic fields generated inside the Sun push through the visible solar disk – the photosphere – into the overlying, invisible chromosphere and corona. These closed magnetic structures are anchored in the photosphere at foot points of opposite magnetic polarity. Coronal loops can be filled with hot gas that shines brightly at extreme ultraviolet and x-ray wavelengths. Driven by motions in the underlying photosphere and below, the coronal loops twist, rise, shear, and interact, releasing magnetic energy that can heat the solar corona and power intense solar flares or coronal mass ejections. Large coronal loops are found in the bulb-like base of coronal streamers, whose long, thin stalks extend out into space. Magnetic fields anchored in the photosphere at one end can also be carried by the solar wind into interplanetary space, resulting in open magnetic fields and a channel for the fast solar wind

The number of active regions, with their bipolar sunspots and coronal loops, varies in step with the sunspot cycle, peaking at sunspot maximum when they dominate the structure of the inner corona. At sunspot minimum, the active regions are largely absent and the strength of the extreme-ultraviolet and x-ray emission of the corona is greatly reduced. Because most forms of solar activity are magnetic in origin, the sunspot cycle also is called the solar cycle of magnetic activity.

Unlike the Earth, magnetism on the Sun does not consist of only one simple dipole; it contains numerous interlooped pairs of opposite magnetic polarity. Powerful magnetism, spawned deep inside the Sun, threads its way through the solar atmosphere, creating a dramatic, ubiquitous, and ever-changing panorama of coronal loops (Fig. 9.6).

Throughout the solar atmosphere, a dynamic tension is set up between the gas pressure of the charged particles and the pressure of the magnetic field (Focus 9.1). In the photosphere and convective zone, the gas pressure dominates the magnetic pressure, allowing the magnetic field to be carried around by the moving gas. Because the churning gases are ionized and hence electrically conductive, they sweep the magnetic field along (Fig. 9.7). The situation is reversed in the low corona within active regions, where hot ionized particles are confined within coronal loops (Fig. 9.8).

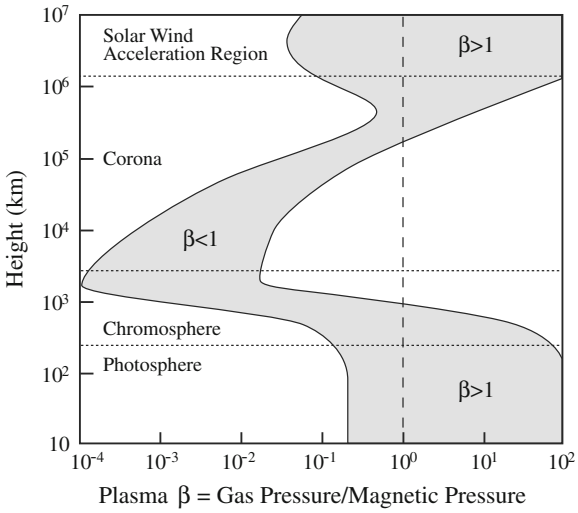


Fig. 9.7 Gas and magnetic pressure The ratio of gas to magnetic pressure, denoted by the symbol β, is plotted as a function of height above the photosphere. The magnetic pressure is greater than the gas pressure in the low corona, where β is less than 1, and magnetic fields determine the structure of the corona. Farther out, the gas pressure can exceed the magnetic pressure, which permits the solar wind to carry the Sun's magnetic field into interplanetary space. In the photosphere, below the corona and chromosphere, the gas pressure also exceeds the magnetic pressure, and the moving gas carries around magnetic fields. [Adapted from Allen (2001).]

Focus 9.1 Magnetic pressure and gas pressure

A magnetic field tends to restrain a collection of electrons and protons, called plasma, while the plasma exerts a pressure that opposes this field. The magnetic pressure is an energy density associated with the magnetic field. The magnetic pressure, P_B, produced by a magnetic field transverse to its direction is given by:

$$P_B = \frac{B^2}{2\mu_0}, \tag{9.1}$$

for a magnetic field of strength B in tesla, where the permeability of free space $\mu_0 = 4\pi \times 10^{-7}$ N A^{-2} $= 1.2566 \times 10^{-6}$ N A^{-2}. As expected, a stronger magnetic field applies a greater restraining pressure.

Hot plasma generates a gas pressure, P_G, owing to the motions of its particles. The ideal gas law describes it:

$$P_G = N\,k\,T, \tag{9.2}$$

where N is the particle number density, $k = 1.38065 \times 10^{-23}$ J K^{-1} is the Boltzmann constant, and T is the temperature. Hotter particles move faster

and create greater pressure to oppose the magnetic field, and denser plasma also results in greater gas pressure.

The two kinds of pressure compete for control of the solar atmosphere. In the low solar corona, strong magnetic fields in active regions hold the hot, dense electrified gas within coronal loops. The magnetic and gas pressures become equal for a magnetic field, B, given by:

$$B = [(2\mu_0 k)NT]^{1/2} = [3.48 \times 10^{-29} NT]^{1/2} \text{ tesla.} \quad (9.3)$$

If a coronal loop contains a hot, dense plasma with $N = 10^{17}$ electrons per cubic meter and $T = 10^6$ K, the magnetic field must be stronger than $B \approx 0.002$ tesla to restrain the plasma. By way of comparison, the magnetic field strength at the Earth's equator is 0.00003 tesla, or at least 100 times weaker than the magnetic field in some coronal loops.

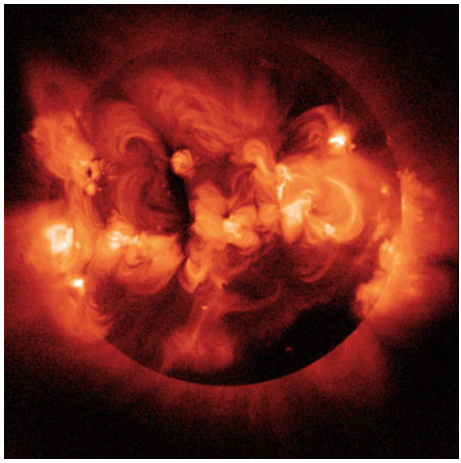


Fig. 9.8 The Sun in x-rays Ionized gases at a temperature of a few million K produce the bright glow seen in this x-ray image of the Sun. It shows magnetic coronal loops that thread the corona and hold the hot gases in place. The brightest features are called active regions and correspond to the sites 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 Feb 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 NASA /ISAS/LMSAL/NAO Japan, University of Tokyo.)

The coronal magnetic fields emerge from underneath the photosphere where they are rooted, and they are continually displaced and replaced by convective motions just below the photosphere. As a result, the corona has no permanent features and it is never still, quiet, or inactive. It is always in a continued state of metamorphosis.

9.1.4 What Heats the Corona?

The visible solar disk, the photosphere, is closer to the Sun's center than the million-degree corona, but the photosphere is several hundred times cooler, with a temperature of 5,780 K. This temperature difference is unexpected because energy should not flow from the cooler photosphere to the hotter corona any more than water should flow uphill. It violates the second law of thermodynamics, which states that heat cannot be continuously transferred from a cooler body to a warmer one without doing work.

We know that visible sunlight cannot resolve the heating paradox. Radiation from the photosphere does not go into the corona; it goes through the corona. There is so little material in the corona that it is transparent to almost all of the photosphere's radiation. Therefore sunlight passes right through the corona without depositing substantial quantities of energy into it, traveling out to warm the Earth and to also keep the photosphere cool.

So, radiation cannot resolve the heating paradox. We must look for alternate sources of energy, and they are related to either moving gases or the magnetic fields in the photosphere and below. Unlike radiation, either the kinetic energy of moving material or the magnetic energy released by magnetic fields can flow from cold to hot regions, keeping the corona hot.

In 1948–1949, astronomers in Germany, the United States, and France independently proposed that sound waves generated in the turbulent convective zone might heat the overlying atmosphere (Biermann 1948; Schwarzschild 1948; Schatzman 1949). The sound waves would accelerate and strengthen as they travel outward through the increasingly rarefied, overlying solar atmosphere, until supersonic shocks are created, dissipating energy and heating the gas.

Although observations from the eighth *Orbiting Solar Observatory*, abbreviated *OSO* 8, showed that sound waves do not transport significant amounts of energy into the corona, these measurements indicated that the sounds might warm the chromosphere to 10,000 K, or roughly twice the temperature of the underlying photosphere (Athay and White 1978, 1979; Bruner 1981). Modern observations indicate that even though the majority of sound waves generated in the convective zone are reflected back into the solar interior at the photosphere, a small percentage of them do manage to slip through the photosphere along inclined magnetic fields, forming shocks that

heat the low chromosphere and create numerous short-lived spicules there. This method of chromosphere heating is generally consistent with the fact that other stars with outer convective zones have chromospheres, while stars that have no convective zones do not exhibit a detectable chromosphere.

For coronal heating, other kinds of waves must be considered, and a likely candidate is magnetic waves that can propagate into the corona and carry energy into it. The Sun's ever-changing coronal magnetic fields are always being jostled, twisted, and stirred around by motions deep within the Sun where the magnetism originates. A tension acts to resist the motions and pull the disturbed magnetism back, generating waves that propagate along magnetic fields, somewhat like a vibrating string. These waves do not form shocks, and once generated they can propagate for large distances, directing their energy along open magnetic fields into the overlying corona.

Such waves are now called Alfvén waves after Hannes Alfvén (1908–1995) who first described them mathematically (Alfvén 1942a, b, Sect. 5.6), and argued that they might heat the corona (Alfvén 1947). Ronald G. Giovanelli (1915–1984), Jack H. Piddington (1910–1997), and Donald E. Osterbrock (1924–2007) subsequently discussed the heating of the chromosphere and corona by Alfvén waves (Giovanelli 1949; Piddington 1956; Osterbrock 1961). An important ongoing controversy is whether or not Alfvén waves propagating through the corona dissipate sufficient energy to heat it. The waves have been detected far from the Sun, suggesting that they might heat the distant corona by traveling along open magnetic fields (Belcher 1969; Cranmer and van Ballegooijen 2005).

Strong, interacting magnetic fields also play a role in heating closed magnetic regions closer to the Sun, within the coronal loops. After all, the hottest and densest material in the low corona is located where the magnetic field is strongest, usually within active regions above sunspots. Moreover, observations from solar spacecraft indicate that the entire magnetic flux in the so-called quiet solar atmosphere outside solar active regions is replenished every 15–40 h (Schrijver and Title 2003).

Motions down inside the convective zone twist and stretch the overlying magnetic fields, slowly building up their energy, and magnetic loops of all sizes always are being pushed up into the solar corona from below. When oppositely directed coronal loops are pressed together, they can merge and join at the place where they touch, releasing their energy to heat the corona. The magnetic fields then reform or reconnect in new magnetic orientations, so this method of coronal heating is termed magnetic reconnection.

Thus, magnetic fields seem to play a fundamental role in channeling, storing, and transforming energy into heat, supplying it on different timescales and sending it to various structures. When the magnetic geometry does not change, the magnetism plays a passive role, guiding the flow of charged particles, heat, and waves along the field lines. And when the magnetic configuration changes, the magnetism can play an active role by triggering instabilities and releasing stored magnetic energy through merging and reconnection of closed magnetic field lines.

9.1.5 Coronal Holes

In contrast to the dense, bright areas, the corona also contains less dense regions called coronal holes. These so-called holes have so little material in them that they appear as large dark areas on x-ray or extreme-ultraviolet images, seemingly devoid of radiation (also see Fig. 9.8).

The coronal holes are neither constant nor permanent; they appear, evolve, and die away in periods ranging from a few weeks to several months, continuously changing in content, shape, and form.

At times of low solar activity, near the minimum of the Sun's 11 year magnetic activity cycle, coronal holes cover the north and south polar caps of the Sun. During more active periods, closer to the cycle maximum, the large coronal holes at the poles shrink and even disappear, and smaller coronal holes appear at all solar latitudes – even at or near the solar equator.

The rarefied coronal holes are not completely empty. The normally constraining magnetic forces relax and open up in the coronal holes to allow an unencumbered outward flow of electrically charged particles and magnetic fields into interplanetary space, keeping the coronal hole's density low and expelling a relentless high-speed wind.

9.2 The Sun's Varying Winds

9.2.1 The Expanding Sun Envelops the Earth

The Sun's radiation is not all that passes through the space between the planets. It is filled with electrons, protons, and magnetic fields emanating from the Sun in a ceaseless flow. These unseen particles and fields form a perpetual solar wind that extends all the way to the Earth and far beyond. It was inferred from comet tails, suggested by theoretical considerations, and fully confirmed by direct measurements from spacecraft in the early 1960s (Focus 9.2). So the space between the planets is not completely empty; it contains the Sun's winds that stream out radially in all directions from the Sun.

Focus 9.2 Discovery of the solar wind

The notion that something is always being expelled from the Sun first arose from observations of comet tails. Comets can appear unexpectedly almost anywhere in the sky, moving in every possible direction, but with tails that always point away from the Sun. A comet therefore travels headfirst when approaching the Sun and tail first when departing from it. Ancient Chinese astronomers concluded that the Sun must have a *chi*, or "life force", that blows away the comet tails. In the early 1600s, the German astronomer

Johannes Kepler (1571–1630) proposed that solar radiation pushes the comet tails away from the Sun.

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 always point generally away from the Sun rather than toward it. One is a yellow tail of dust, which can litter the comet's curved path. The dust is pushed away from the Sun by the pressure of sunlight. The other tail is colored electric blue, shining in the light of ionized particles. The ions in comet tails always stream along straight paths away from the Sun.

The existence of the solar wind of charged particles was suggested from observations of comet ion tails in the mid 20th century. 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 continued 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 (Biermann 1951, 1957).

Eugene N. Parker (1927–) of the University of Chicago showed how such a relentless flow might work, dubbing it the *solar wind* (Parker 1958). 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 (Parker 1958, 1960, 1963). The expansion begins slowly near the Sun, where the solar gravity is the strongest. Then the expanding corona accelerates outward into space until the winds break away from the Sun, and eventually they cruise along at the roughly constant and supersonic velocities needed to account for the acceleration of comet tails. This creates a strong, persistent solar wind, forever blowing throughout the solar system.

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 September 12, 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 (Gringauz 1961). 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 August 27, 1962. Marcia Neugebauer (1932–) and Conway W. Snyder of the Jet Propulsion Laboratory used more than one hundred 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 (Neugebauer and Snyder 1966; Neugebauer 1997). It also unexpectedly indicated that the solar wind has a slow and a fast component. The slow one moves at a speed of 300–400 km s^{-1}; 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.

In the low corona, strong magnetic fields constrain the hot ionized gas within coronal loops. But further out in the corona, the magnetic fields decrease in strength and cannot restrain the outward flow of the million-degree gas; it also flows out unencumbered from the open magnetic fields in coronal holes.

Example: **Hot enough to break away from the Sun's gravity**

The corona is fully ionized, with a temperature of several million K and consisting of electrons and protons. We can determine if these particles are hot enough to escape from the Sun's gravity by equating their thermal velocity $(3kT/m)^{1/2}$ to the escape velocity of the Sun $(2GM_\odot/R_\odot)^{1/2}$, to obtain (Sects. 3.2, 5.2, 5.3):

$$T = \frac{2GmM_\odot}{3kR_\odot} \approx 9.2 \times 10^{33}\, m \text{ K}, \tag{9.4}$$

where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg and radius $R_\odot = 6.955 \times 10^8$ m, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and the constant of gravitation $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}. The temperature is denoted by T and the particle mass is designated by m.

For a proton of mass $m_P = 1.6726 \times 10^{-27}$ kg, the temperature required for escape is $T \approx 1.5 \times 10^7$ K, while for the electron of mass $m_e = 9.1094 \times 10^{-31}$ kg it is $T \approx 8.4 \times 10^3$ K. So the electrons in the million-degree solar corona would have no problem escaping from the Sun, and the protons would require just a little extra push to put them out where the solar gravity is somewhat diminished.

The solar gale brushes past the planets, wraps itself around the Earth, and carries the Sun's corona out to interstellar space. This radial, supersonic outflow creates a huge bubble of plasma, with the Sun at its center and the planets inside; this is called the *heliosphere*, from *Helios* the "God of the Sun" in Greek mythology.

9.2.2 Properties of the Solar Wind

The million-degree corona is so hot that it cannot stand still. Indeed, the solar wind consists of an overflow corona, which is too hot to be entirely constrained by the Sun's inward gravitational pull and perpetually moves out into surrounding space. The hot gas creates an outward pressure that tends to oppose the inward pull of the Sun's gravity. At great distances, where the solar gravity weakens, the hot protons and electrons overcome the Sun's gravity and continue to accelerate, like water overflowing a dam. So, the solar corona is really the visible, inner base of the solar wind, and the solar wind is simply the hot corona expanding into interplanetary space.

The Sun's continuous wind travels with two main velocities. There is a fast, uniform wind that blows at a speed of about 750 km s^{-1}, and a variable, gusty slow wind that moves about half as fast. Both winds are supersonic, moving at least 10 times faster than the speed of sound in the solar wind. The Sun's wind also rushes on with little reduction in speed because there is almost nothing out there to slow it down. Both the fast and slow winds from the Sun are much more tenuous, hotter, and faster than any wind on the Earth.

Example: **The supersonic solar wind**

The speed of sound, c_s, in the solar wind can be determined from the equation (Laplace 1816, Sect. 5.8):

$$c_S = \left(\frac{\gamma kT}{\bar{m}}\right)^{\frac{1}{2}}, \tag{9.5}$$

where the adiabatic index $\gamma = 5/3$ for a fully ionized gas like the solar wind, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, and the mean molecular mass $\bar{m} = \mu \times$ u where the mean molecular weight $\mu = 0.5$ for the solar wind (fully ionized almost entirely composed of protons), and the atomic mass unit u $= 1.66054 \times 10^{-27}$ kg. For a temperature of $T = 1.2 \times 10^5$ K for solar-wind protons near the Earth's orbit, this equation gives $c_s \approx 5.8 \times 10^4$ m $s^{-1} \approx 58$ km s^{-1}. By way of comparison, the slow and fast solar winds have respective velocities of about 375 and 750 km s^{-1}, which indicates that the solar wind is everywhere supersonic, or moving at speeds faster than sound.

Table 9.3 Mean values of solar-wind parameters at the Earth's orbit[a]

Parameter	Mean value
Particle density, N	$N \approx 10^7$ m^{-3} (5 electrons and 5 protons per cubic centimeter)
Velocity, V	Fast wind $V \approx 750$ km s^{-1}
	Slow wind $V \approx 375$ km s^{-1}
	Average $V \approx 600$ km s^{-1}
Mass density, ρ	$\rho = 10^{-20}$ kg m^{-3} (protons)
Flux, F	$F \approx 6 \times 10^{12}$ particles m^{-2} s^{-1}
Temperature, T	$T \approx 120{,}000$ K (protons) to 140,000 K (electrons)
Particle thermal energy, kT	$kT \approx 2 \times 10^{-18}$ J ≈ 12 eV
Proton kinetic energy, 0.5 m_pV^2	0.5 $m_pV^2 \approx 10^{-16}$ J $\approx 1{,}000$ eV $= 1$ keV
Particle thermal energy density	$NkT \approx 10^{-11}$J m^{-3}
Proton kinetic energy density	0.25 $N\, m_p\, V^2 \approx 10^{-9}$ J m^{-3}
Radial magnetic field, B_r	$B_r = 2.5 \times 10^{-9}$ T $= 2.5$ nT $= 2.5 \times 10^{-5}$ G
Alfvén velocity, V_A	$V_A = 32$ km s^{-1}
Sound speed, c_s	$c_s \approx 50$ km s^{-1}

[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 $= 1.496 \times 10^{11}$ m. The Boltzmann constant $k = 1.38 \times 10^{-23}$ J K^{-1} relates temperature and thermal energy. The proton mass $m_p = 1.67 \times 10^{-27}$ kg

Because the electrified wind material is an excellent conductor of heat, the temperature falls off only gradually with distance from the Sun, reaching between 120,000 K and 140,000 K at the Earth's distance. The tenuous wind has been diluted to a rarefied plasma by the time it reaches the Earth's distance from the Sun, where there are approximately 5 million electrons and 5 million protons per cubic meter of solar wind. The density of the solar wind is so low that there are not enough particles to heat astronauts that might venture outside their spacecraft.

Because it is wrapped into a spiral, the interplanetary magnetic field strength falls off linearly with distance from the Sun, in contrast to the solar wind number density that decreases more rapidly, as the inverse cube of the distance, as it fills a larger volume.

Physical properties of the solar wind at the Earth's distance from the Sun are listed in Table 9.3.

At a mean speed of about 600 km s^{-1}, the flux of solar wind particles is far greater than anything else in nearby space. Between one and ten million million (10^{12}–10^{13}) particles in the solar wind cross every square meter of space each second. That flux far surpasses the flux of more energetic cosmic rays that enter our atmosphere, with the abundant protons arriving with a flux of only 640 protons per square meter per second. The peak local energy density of cosmic rays is about one million electron volts per cubic meter, or about one ten thousandth (10^{-4}) the kinetic energy density of solar wind protons.

Moreover, interplanetary magnetic fields act as a barrier to electrically charged cosmic rays that are coming from the depths of space, preventing them from reaching the Earth. During the maximum in the solar cycle, stronger solar magnetic fields are carried out into interplanetary space by the Sun's wind, deflecting

more cosmic rays. Less extensive interplanetary magnetism, during a minimum in the 11 year cycle of magnetic activity, lowers the barrier to the cosmic particles and allows more of them to arrive at Earth. This unexpected anti-correlation between solar activity and cosmic rays arriving at Earth is often called the Forbush effect after its discovery by Scott Forbush (1904–1984) in the early 1950s (Forbush 1950). It was explained by Peter Meyer (1920–2002), Eugene Parker and John Simpson (1916–2000) who proposed that enhanced interplanetary magnetism near the maximum in the 11 year solar activity cycle deflects cosmic rays from their Earth-bound paths (Meyer et al. 1956).

Because the Sun is blowing itself away continuously, we might imagine that it would eventually vanish from view after expelling all of its substance into space. Every second, the solar wind carries about a billion kilograms, or 1 million tons, of the Sun into surrounding space. That seems significant, but in 4.6 billion years the solar wind has only carried away about 0.0001, or one ten thousandth, of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. Moreover, that is about three times less than the amount of mass turned into energy during this time by nuclear reactions near the center of the Sun (Focus 9.3).

Focus 9.3 Mass loss from the Sun

The Sun is relentlessly expelling mass into the solar wind, most of it in protons with a mass $m_P = 1.6726 \times 10^{-27}$ kg. Using measurements at the Earth, located at a mean distance from the Sun of 1 AU $= 1.496 \times 10^{11}$ m, we can estimate the total mass, ΔM_{SW} lost from the Sun in the solar wind every second:

$$\Delta M_{SW} = 4\pi m_P N_P V(\mathrm{AU})^2 = 1.41 \times 10^9 \text{ kg}, \tag{9.6}$$

where the number density of protons in the solar wind at the Earth's distance is $N_P = 5 \times 10^6$ m^{-3} and we have assumed an average wind velocity of $V = 600$ km s^{-1} $= 6 \times 10^5$ m s^{-1}. The total mass M_{SW} lost at this rate over the past 4.6 billion years, using 1 year $= 3.156 \times 10^7$ s, is:

$$M_{SW} = \Delta M_{SW} \times 4.6 \times 10^9 \times 3.156 \times 10^7 = 2.05 \times 10^{26} \text{ kg} \approx 0.0001 M_\odot, \tag{9.7}$$

where $M_\odot = 1.989 \times 10^{30}$ kg denotes the Sun's current mass.

About three times more mass is consumed every second by nuclear reactions that make the Sun shine. For just one fusion of four protons into one helium nucleus the mass lost is $\Delta m = 0.007\ (4m_P)$ with an energy release $\Delta E = \Delta m c^2$, for a velocity of light $c = 2.9989 \times 10^8$ m s^{-1}. Since these nuclear reactions are supplying the Sun's absolute luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, the mass loss ΔM_{NR} in one second is:

$$\Delta M_{NR} = \Delta m \left(\frac{L_\odot}{\Delta E} \right) = \frac{L_\odot}{c^2} \approx 4.26 \times 10^9 \text{ kg}, \tag{9.8}$$

and over the past 4.6 billion years, assuming the Sun has always been shining at the same rate, the total mass M_{NR} lost by nuclear reactions is:

$$M_{NR} = 4.6 \times 10^9 \times 3.156 \times 10^7 \times \Delta M_{NR} \approx 6.2 \times 10^{26} \text{ kg} \approx 0.0003\, M_\odot. \tag{9.9}$$

This mass is carried off by the radiation that makes the Sun shine.

9.2.3 Where Do the Two Solar Winds Come From?

Instruments aboard spacecraft have detected two solar winds with different physical properties. There is a fast wind that moves at a speed of about 750 km s^{-1} and a slow wind that blows at about half that speed. The high-speed wind is steady and uniform, whereas the slow-speed wind is variable and gusty.

The two solar winds do not blow uniformly from all places on the Sun; instead they depend on solar latitude. The spatial distribution of the two types of winds also depends on the Sun's magnetic field configuration, which varies dramatically with the 11 year solar-activity cycle.

As suggested by Sir William I. Axford (1933–2010), the steady, uniform, high-speed wind emanates from magnetically open configurations in the corona (Axford 1985). The open magnetic fields in coronal holes provide a conduit for the fast wind. In contrast the slow wind – which is filamentary and transient – involves the intermittent release of material from previously closed magnetic regions, so the slow wind may not be treated as an equilibrium flow in a steady state.

The distribution of the open and closed magnetic regions on the Sun, and therefore the places of origin for the two solar winds, depends on the 11 year cycle of solar magnetic activity.

Near activity minimum, the high-speed wind moves out of the open magnetic fields in large coronal holes located at the Sun's polar regions. A slow, gusty, and variable wind then moves away from closed magnetic regions near the Sun's equator (McComas et al. 2000).

The simple, bimodal distribution of fast and slow wind structures disappears near the maximum in the 11 year solar activity cycle. The large polar coronal holes then shrink and even disappear and smaller coronal holes appear at all solar latitudes. A chaotic and complex mixture of varying solar wind flows therefore is found at all solar latitudes near activity maximum (McComas et al. 2003). The slow winds still seem to be associated with closed magnetic structures, such as active regions, whereas the fast winds rush out of the interior of coronal holes all over the Sun. Solar active regions, with their explosive behavior, provide a noticeable third source for the solar winds near the activity-cycle maximum.

9.2.4 Where Does the Solar Wind End?

How far does the solar wind extend, and where does its influence end? The solar wind carves out a cavity in the interstellar medium known as the *heliosphere*. Zurbuchen (2007) has reviewed the coupling of the Sun and the heliosphere; also see Lang (2009).

Since the solar wind is weakened by expansion, thinning out as it moves into a greater volume, it eventually becomes too dispersed to repel interstellar forces. The winds are no longer dense or powerful enough to withstand the pressure of gas and magnetic fields coursing between the stars. The radius of this celestial standoff distance, in which the pressure of the solar wind falls to a value comparable to the interstellar pressure, has been estimated at about 100 AU, or one hundred times the mean distance between the Earth and the Sun.

The radius of the heliosphere can be estimated by determining the standoff distance, or stagnation point, in which the ram pressure, P_W, of the solar wind falls to a value comparable to the interstellar pressure, P_I. As the wind flows outward, its velocity remains nearly constant, while its density decreases as the inverse square of the distance. The dynamic pressure of the solar wind therefore also falls off as the square of the distance, and we can use the solar-wind properties at the Earth's distance of 1 AU to infer the pressure, P_{WS}, at the stagnation-point distance, R_S. Equating this to the interstellar pressure we have:

$$P_{WS} = P_{1AU} \times \left(\frac{1AU}{R_S}\right)^2 = \left(N_{1AU}\, V_{1AU}^2\right) \times \left(\frac{1AU}{R_S}\right)^2 = P_I, \tag{9.10}$$

where the number density of the solar wind near the Earth is about $N_{IAU} = 5$ million particles per cubic meter and the velocity there is about $V_{IAU} = 500$ km per second.

To determine the distance to the edge of the solar system, R_S, we also need to know the interstellar pressure, which is the sum of the thermal pressure, the dynamic pressure, and the magnetic pressure in the local interstellar medium. Its estimated value results in $R_S = 100$ AU or more, well beyond the orbits of the major planets.

The termination shock at the edge of the solar wind has now been measured from the *Voyager 1* and *2* spacecraft (Decker et al. 2005; Stone et al. 2005, 2008).

Instruments aboard the twin *Voyager 1* and *2* spacecraft, launched in 1977 and now cruising far beyond the outermost planets, have approached this edge of the solar system from different directions. *Voyager 1* is moving in the northern hemisphere of the heliosphere and *Voyager 2* in the southern hemisphere. *Voyager 1* crossed the termination shock of the supersonic flow of the solar wind on December 16, 2004 at a distance of 94 times the mean distance between the Earth and the Sun, or at 94 AU from the Sun. At this distance, the spacecraft's instruments recorded a sudden increase in the strength of the magnetic field carried by the solar wind, as expected when the solar wind slows down and its particles pile up at the termination shock.

Voyager 2 crossed the termination shock on August 30, 2007 at a distance of 84 AU from the Sun. It appears that there is a significant north/south asymmetry in the heliosphere, likely due to the direction of the local interstellar magnetic field.

Both *Voyager 1* and *2* have therefore now crossed into the vast, turbulent heliosheath, the region where the interstellar gas and solar wind interact, due to the reflection and deflection of the solar-wind ions by the magnetized wind beyond the heliosheath. In technical terms, the solar-wind ions in the heliosheath are deflected by magnetosonic waves reflecting off of the heliopause, causing the ions to flow parallel to the termination shock toward the heliotail.

Both *Voyager* spacecraft are equipped with plutonium power sources expected to last until at least 2020 and perhaps 2025. So they ought to eventually measure the heliopause at the outer edge of the heliosheath. It is the place where interstellar space begins.

In the meantime, the *Interstellar Boundary EXplorer*, abbreviated *IBEX*, was launched on 19 Oct 2009. Instruments on this spacecraft, which operates in Earth orbit, detect neutral, or unionized, atoms coming from the termination shock and the boundary between the solar wind and interstellar space.

9.3 Explosions on the Sun

9.3.1 Solar Flares

Suddenly, without warning, a part of the Sun explodes, creating a solar flare. Some of them are the biggest explosions in the solar system, releasing energy of ten million, billion, billion joule, or 10^{25} J, in just 100 s. This is comparable in strength to 20 million nuclear bombs exploding simultaneously, and each with energy of 100 Megatons of TNT.

A substantial fraction of the flare energy goes into accelerating electrons and protons to nearly the speed of light. Some of these high-energy particles are hurled into the Sun, briefly raising the temperature of Earth-sized regions of the Sun to more than 10 million K. Other accelerated particles are tossed out into interplanetary space and emit intense radio and x-ray radiation.

The short-lived solar flares unleash their energy in the vicinity of sunspots, covering just a few tenths of a percent of the solar disk. These incredible explosive outbursts become more frequent and violent when the number of sunspots is greatest; several solar flares can be observed on a busy day near the maximum of the sunspot cycle. However, they are not caused directly by sunspots; solar flares instead are powered by magnetic changes in the corona above sunspots.

Although it emits very large amounts of energy, a solar flare usually releases less than 1/1,000th of the total energy radiated by the Sun every second, so they are only minor perturbations in the combined colors, or white light, of the Sun. The first record of a solar flare detected on the visible solar disk therefore did not occur

until the mid-nineteenth century – on September 1, 1859, when the English astronomers Richard Christopher Carrington (1826–1875) and Richard Hodgson (1804–1872) independently noticed one (Carrington 1860; Hodgson 1860).

A new perspective, which demonstrated the frequent occurrence of solar explosions, was made possible when flares were observed in the red Balmer alpha emission line of hydrogen at 656.3 nm, originating in the chromosphere. This wasn't possible until the early 20th century, after the invention of instruments that isolated the red hydrogen-alpha emission from the Sun's intense visible light at adjacent wavelengths in the solar spectrum. Such observations showed that at the chromospheric level in the solar atmosphere a solar flare consists in simplest form as two extended, parallel flare ribbons. But a fundamental understanding of the physical processes responsible for solar flares had to wait until they were detected at invisible radio wavelengths from the ground and in x-rays from space.

During World War II (1939–1945) it was discovered that sudden, intense radio outbursts from the Sun, associated with solar flares, could interfere with radio communications and radar systems. Soon after the war ended, J. Paul Wild's (1923–2008) group of Australian radio astronomers used swept frequency radio receivers to show that some flares eject particles at about half the speed of light, or about 1.5×10^8 m s^{-1}, while others moved at about 1/100th this speed or at about 10^6 m s^{-1}, and were attributed to shock waves. Wild and Smerd (1972) and Wild et al. (1963) provided early summaries of observations of solar radio bursts. Bastian et al. (1998) provided a more recent review of the radio emission from solar flares.

Example: **Watching flare-accelerated electrons and shock waves**

When a flare-associated disturbance, such as an electron beam or a shock wave, moves through the coronal plasma, the local electrons are displaced with respect to the protons, which are more massive than the electrons. The electrical attraction between the electrons and protons pulls the electrons back in the opposite direction, and an oscillation is set up at the plasma frequency, ν_p, given by (Sect. 5.5):

$$\nu_P = \left[\frac{e^2 N_e}{4\pi^2 \varepsilon_0 m_e}\right]^{1/2} = 8.98 N_e^{1/2} \text{ Hz}, \tag{9.11}$$

for an electron density N_e, where the electron charge $e = 1.6022 \times 10^{-19}$ coulomb, the electron mass $m_e = 9.1094 \times 10^{-31}$ kg, the permittivity of free space is $\varepsilon_0 = 8.8542 \times 10^{-12}$ F m^{-1}, and $\pi = 3.14159$. Note that the wavelength, λ_p, of radiation at the plasma frequency is given by $\lambda_p \nu_p = c = 2.9979 \times 10^8$ m s^{-1}, the speed of light.

Low in the solar corona, where $N_e \approx 10^{14}$ m^{-3}, the plasma frequency is about 90 MHz, at radio frequencies, or radio wavelengths of about 3.3 m, which can be observed from the ground. Since the plasma frequency decreases with the diminishing coronal electron density at greater distances from the Sun,

where the corona occupies an increasing volume of space, beams or shocks sent out from a solar flare excite progressively lower plasma frequencies. Near the Earth, for example, $N_e \approx 5 \times 10^6\ \mathrm{m}^{-3}$ and the plasma frequency is about 20 kHz corresponding to a wavelength of 15 km.

When an electron density model of the solar atmosphere is used, the emission frequency can be related to height, and combined with the time delays observed at successively lower frequencies, to obtain the outward velocity of the moving disturbance. For a type III radio burst, an average of about 0.4 c is determined. The slower drift associated with type II radio bursts suggests an outward motion at about 1,000 km s^{-1} and has been attributed to shock waves.

Radiation from low-frequency plasma oscillations, between 100 kHz and 10 MHz, is absorbed in the Earth's atmosphere, but they have been observed for decades using instruments aboard spacecraft. These observations can be used to track the spiral magnetic field that guides the flare electrons as they move out through the increasingly rarefied coronal plasma.

The bulk of radiation from high-temperature solar flares is not emitted as radio waves but instead at extreme ultraviolet and x-ray wavelengths, where they can briefly outshine the entire Sun. This radiation is absorbed in the Earth's atmosphere; therefore, astronomers have observed it from outer space, beginning with primitive instruments aboard balloons or sounding rockets and continuing with increasingly sophisticated telescopes in many satellites, including the *Yohkoh*, *Ulysses*, *Wind*, *SOHO*, *ACE*, *TRACE*, *Hinode*, *RHESSI*, *STEREO*, and *SDO* spacecraft (Lang 2009).

Why does a solar flare occur? What triggers the instability and suddenly ignites an explosion from magnetic fields that remain unperturbed for long intervals of time? Answers to these questions were proposed in 1960 by the Austrian-born American Thomas Gold (1920–2004), then at Cornell University, and by the Englishman Fred Hoyle (1915–2001), at Cambridge University, when they showed how solar flares could be powered by stressed magnetic loops that interact, dissipating their energy in the corona (Gold and Hoyle 1960). The flares are triggered when magnetized coronal loops are pressed together, driven by motions beneath them, meeting to touch one another and merge.

Magnetic fields have a direction associated with them and if oppositely directed magnetic fields are pushed together, they can interact. When these merging magnetic fields are closed coronal loops, they will break open to release magnetic energy in the form of flare heating and particle acceleration. The magnetic fields are not broken permanently; they simply reconnect to their closed state. For this reason, this merging and coupling is known as *magnetic reconnection* (Sweet 1969; Hirajama 1974).

Benz and Güdel (2010) provided a review of magnetically driven flares on the Sun and other stars, whereas Zweibel and Yamada (2009) reviewed magnetic

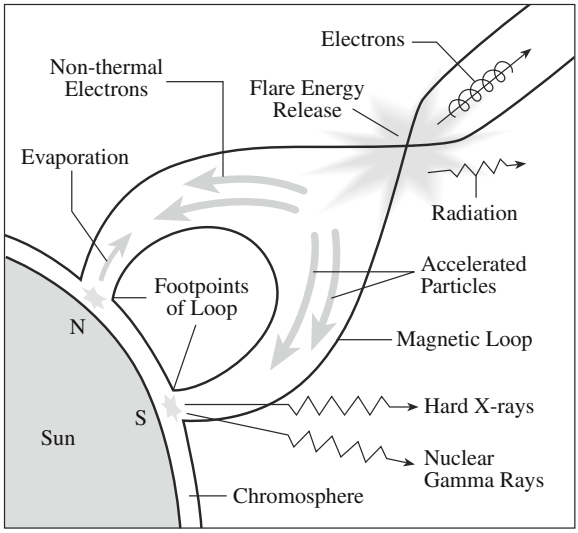


Fig. 9.9 Solar flare model A solar flare is powered by magnetic energy released from a magnetic interaction site above the top of a coronal loop. Electrons are accelerated to high speed during a solar flare, generating a burst of radio energy as well as impulsive loop-top hard X-ray emission. Some of these nonthermal electrons are channeled down the loop and strike the chromosphere at nearly the speed of light, emitting hard X-rays by electron−ion *bremsstrahlung* at the loop footpoints. When beams of accelerated protons enter the dense, lower atmosphere, they cause nuclear reactions that result in gamma-ray spectral lines and energetic neutrons. Material in the chromosphere is heated very quickly and rises into the coronal loop, accompanied by a slow, gradual increase in soft X-ray radiation. This upwelling of heated material is called *chromospheric evaporation* and it occurs in the decay phase of the flare

reconnection in astrophysical plasmas. Bhattacharjee (2004) discussed magnetic reconnection in the Earth's magnetotail and the solar corona.

So, powerful solar flares stem from the interaction of coronal loops. These loops are always moving about, and often are brought into contact by these movements. Magnetic fields coiled up in the solar interior, where the Sun's magnetism is produced, also can bob into the corona to interact with preexisting coronal loops. In either case, the coalescence leads to the rapid release of magnetic energy through magnetic reconnection.

Because flares apparently originate in the low corona and the ubiquitous coronal loops dominate their structure, it's not surprising that solar-flare models involve a coronal loop (Fig. 9.9). Magnetic reconnection triggers the release of magnetic energy just above the loop top, where electrons and protons are accelerated. In less than 1 s, electrons are accelerated to nearly the speed of light, producing intense radio signals. Protons likewise are accelerated to high speeds, and both the electrons and protons are hurled down into the Sun and out into space.

Instruments aboard spacecraft have observed nuclear reactions and the creation of anti-matter during solar flares (Chupp 1984; Share et al. 2004; Hurford et al.

2006). When protons and heavier ions are accelerated to high speed during solar flares, and beamed downward into the Sun, they slam into the dense, lower atmosphere, shattering the nuclei of atoms and sometimes tearing energetic neutrons out of them. Many of these neutrons are eventually captured by ambient, or non-flaring, hydrogen nuclei, the protons, in the photosphere, emitting one of the Sun's strongest gamma-ray lines at 2.223 MeV.

Another strong gamma-ray line emitted during solar flares is the 0.511 MeV pair-annihilation line. Positrons, the anti-matter counterpart of electrons, are released during the decay of radioactive nuclei produced when flare-accelerated protons and heavier nuclei are hurled down into the lower solar atmosphere during a solar flare. The positrons and the electrons annihilate, producing radiation at 0.511 MeV, which is the energy contained in the entire mass of a nonmoving electron.

Example: **Anti-matter created during solar flares**

Positrons, or positive electrons, are the anti-matter particles of electrons, and they are produced during solar flares. The positrons, denoted e^+, are inferred from observations of the pair-annihilation reaction during solar flares. This reaction is written as:

$$e^+ + e^- = \gamma + \gamma, \tag{9.12}$$

where e^- denotes an electron and γ is a gamma-ray photon. The energy, E, of this photon will be equal to the energy taken to completely destroy an electron of rest mass $m_e = 9.1094 \times 10^{-31}$ kg, or $E = m_e\, c^2 = 8.187 \times 10^{-14}$ J $= 0.511$ MeV, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, and 1 MeV $= 1.60217 \times 10^{-13}$ J. Since the positron has the same mass as an electron, both gamma-ray protons have this energy. The 0.511 MeV line has been observed during solar flares using instruments aboard spacecraft.

Because the chromosphere has been heated very rapidly by the accelerated particles that were hurled down into it, that part of the chromosphere explodes, or evaporates, up into the corona to release the excess energy. This process may include the gradual release of energy when the coronal loop relaxes into a more stable configuration during the decay phase of a solar flare.

Example: **Chromospheric evaporation**

Energetic particles are hurled down into the chromosphere during a solar flare, heating it to temperatures of $T = 20$ million, or 2×10^7, K, which is slightly hotter than the center of the Sun. The hot gas flows back up into coronal loops, emitting x-ray radiation during the decay phase of the solar flare. The photon energy of the x-rays is roughly equal to the thermal energy

of the gas, or $3kT/2 = 4.14 \times 10^{-16}$ J, where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. X-ray astronomers like to specify this energy in keV, where 1 keV $= 1.602 \times 10^{-16}$ J, and chromospheric evaporation is detected at x-ray photon energies of about 2.6 keV.

9.3.2 Coronal Mass Ejections

Coronal mass ejections are gigantic magnetic bubbles that can rush away from the Sun at supersonic speeds, expanding to become larger than the Sun in a few hours (Figs. 9.10, 9.11). They carry about 10^{13} kg or, 10 billion tons, of material out into space, produce intense shock waves, and accelerate vast quantities of energetic particles in interplanetary space. A coronal mass ejection moves through space at a speed of about 400 km s^{-1}, carrying a kinetic energy of about 10^{24} J, which is comparable to the explosive energy of a large solar flare. When directed at the Earth, a coronal mass ejection arrives at the planet about four days after being ejected from the Sun (Focus 9.4).

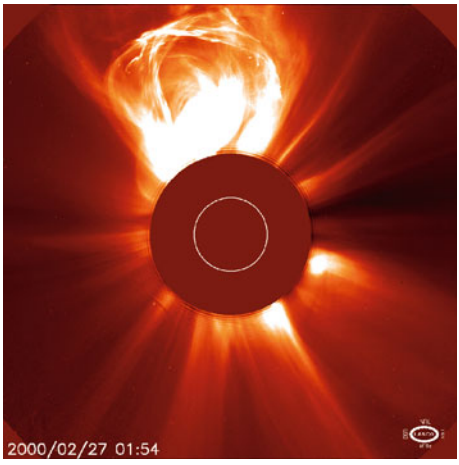


Fig. 9.10 Coronal mass ejection A huge coronal mass ejection is seen in this image, taken on 27 Feb 2000 with a coronagraph on the *SOHO* spacecraft. The white circle denotes the edge of the Sun's visible disk, so this mass ejection is about twice as large as the Sun. The dark area corresponds to the occulting disk of the coronagraph that blocks intense sunlight and permits the overlying solar atmosphere, or corona, to be seen. (Courtesy of the *SOHO* LASCO consortium, *SOHO* is a project of international cooperation between ESA and NASA.)

Focus 9.4 Physical properties of coronal mass ejections

Coronal mass ejections are detected as localized brightness increases in white-light coronagraph images. Integration of the brightness increase, that depends only on the electron density, N_e, permits evaluation of the total mass, M, of the ejection. For a sphere of radius, R, we have:

$$M = 4\pi R^3 N_e m_p/3, \tag{9.13}$$

where $\pi \approx 3.14159$ and the proton mass $m_p = 1.6726 \times 10^{-27}$ kg. The corona is a fully ionized, predominantly (90 %) hydrogen, plasma, so the number density of protons and electrons are equal, but since the protons are 1,836 times more massive than the electrons, the protons dominate the mass. For a mass ejection with an electron, or proton, density of $N_e = 10^{13}$ m^{-3}, that has grown as large as the Sun, with a radius of $R = 6.955 \times 10^8$ m, this expression gives

$$M \approx 2 \times 10^{13} \text{ kg}, \tag{9.14}$$

or about 20 billion tons.

At the rate of one ejection per day, and 10^{13} kg per ejection, this amounts to a mass flow rate of about 2×10^8 kg s^{-1}, since there are 86,400 s per day.

By way of comparison, the solar wind flux observed in the ecliptic at the orbit of the Earth is about 6×10^{12} protons m^{-2} s^{-1}, or about 10^{-14} kg m^{-2} s^{-1} (Table 9.3). If this flux is typical of that over the entire Sun-centered sphere, with an average Sun-Earth distance of $D = 1$ AU $= 1.496 \times 10^{11}$ m, we can multiply by the sphere's surface area, $4\pi D^2$, to obtain a solar wind mass flow rate of about 2.8×10^9 kg s^{-1}. That's roughly 15 times the mass flow rate from coronal mass ejections, or in other words the coronal mass ejection rate is about 5 % that of the steady, perpetual solar wind, and for just a relatively brief time during the ejection.

The kinetic energy, KE, of a coronal mass ejection with a speed of $V = 400$ km s^{-1} and a mass $M = 10^{13}$ kg is:

$$KE = MV^2/2 \approx 10^{24} \text{ J}. \tag{9.15}$$

This is comparable to the energies of large solar flares that lie between 10^{21} and 10^{25} J.

At a speed of $V = 400$ km s^{-1}, the time, T, to travel from the Sun to the Earth, at an average distance of $D =$ AU $= 1.496 \times 10^{11}$ m is:

$$T = D/V \approx 3.74 \times 10^5 \text{ s} \approx 4.3 \text{ days}, \tag{9.16}$$

where 1 day = 86,400 s. Zhang and Low (2005) have reviewed the hydromagnetic nature of solar coronal mass ejections.

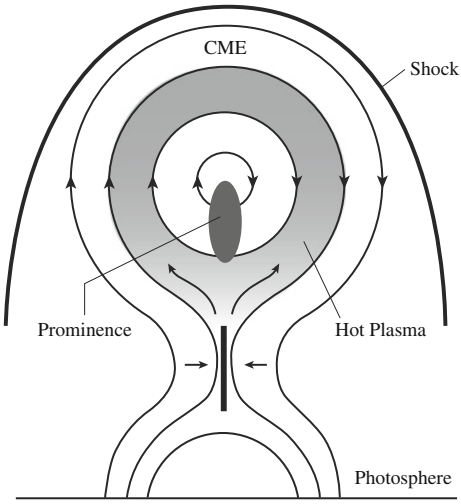


Fig. 9.11 Model of coronal mass ejection A magnetic reconnection takes place at a current sheet (*dark vertical line*) beneath a prominence and above closed magnetic field lines. The coronal mass ejection (CME) traps hot plasma below it (*shaded region*). The solid curve at the *top* is the bow shock driven by the CME. The closed field region above the prominence (*center*) is supposed to become a flux rope in the interplanetary medium. [Adapted from Martens and Kuin (1989).]

The physical size of the mass ejections dwarfs that of solar flares and even the active regions in which flares occur. However, like solar flares, the rate of occurrence of coronal mass ejections varies in step with the 11 year cycle of solar magnetic activity, ballooning out of the corona several times a day during activity maximum. Large coronal mass ejections can occur with or without a solar flare, but they both appear to be powered by the abrupt release of the corona's magnetic energy, with threatening effects for the Earth and nearby space.

9.4 Space Weather

9.4.1 Earth's Protective Magnetosphere

Our planet is immersed within the hot, electrically charged solar wind that blows out from the Sun in all directions and never stops, carrying with it a magnetic field rooted in the Sun. Solar flares and coronal mass ejections create powerful gusts in the Sun's winds, producing space weather – the cosmic equivalent of a terrestrial blizzard or hurricane. Fortunately, we are protected from the full force of this relentless, stormy gale by the Earth's magnetic field.

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 Magnete Tellure*, which translated into English is *Concerning Magnetism*, *Magnetic Bodies*, *and the Great Magnet Earth* (Gilbert 1600). In this work, which is still available in an English version, Gilbert showed that the Earth is itself a great magnet, which explains the orientation of compass needles. It is as if there was a colossal bar magnet at the center of the Earth.

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 stands upright, pointing into or out of the ground. In between, at intermediate latitudes, the compass needle points north or south with a downward dip of one end but not vertically as at a pole. Because 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; therefore, an arrowed compass points north.

We can describe the Earth's magnetism by invisible magnetic field lines, which orient compass needles. These lines of magnetic force emerge from the south magnetic pole, loop through nearby space and reenter at the north magnetic pole. The lines are close together near the magnetic poles where the magnetic force is strong, and spread out above the Earth's Equator where the magnetism is weaker than at the poles. We cannot see the invisible 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 0.0000305 tesla, or 0.305×10^{-4} tesla. Measurements of the surface magnetic fields of the Earth show stronger fields near the poles where the magnetic field lines congregate, at roughly twice the strength of the field at the Equator. Although these fields decrease in strength as the inverse cube of the distance, they remain strong enough to divert most of the solar wind around the Earth at a distance far above the atmosphere, thereby protecting humans on the ground from possibly lethal solar particles.

When any charged particle encounters a magnetic field, it must change direction, moving away from or around the magnetism. When the protons and electrons in the gusts or steady flow of the solar wind encounter the Earth's magnetic fields, they are deflected around it, like a rock in a stream or a windshield deflecting air around an automobile.

The Earth's magnetic fields hollow out a protective cavity in the solar wind, which is called the *magnetosphere*, a term coined in 1959 by Austrian-born American Thomas Gold (1920–2004), then at Cornell University (Gold 1959a, b). It is that region surrounding any planet in which its magnetic field dominates the behavior of electrically charged particles, such as electrons, protons, and other ions.

The dipolar (two poles) magnetic configuration applies near the surface of the Earth, but farther out, the magnetic field is distorted by the Sun's perpetual wind. Although it is exceedingly tenuous, the solar wind is powerful enough to mold the outer edges of the Earth's magnetosphere into a changing asymmetric shape (Fig. 9.12).

The solar wind usually bends around the Earth's magnetic field at a distance from the Earth's center of about 10 times the Earth's radius on the dayside that faces the Sun (Focus 9.5). Here, the solar wind pushes the Earth's magnetism in, compressing its outer magnetic boundary and forming a shock wave, shaped like waves that pile up ahead of the bow of a moving ship and resembling the flow of air around a supersonic aircraft. After forming this bow shock, the solar wind is

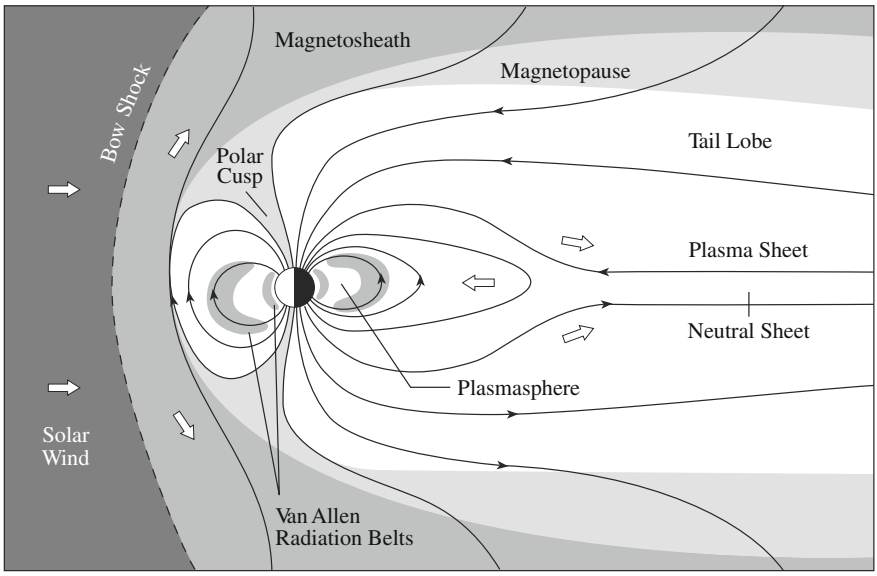


Fig. 9.12 Elements of the magnetosphere The Earth's magnetic field carves out a hollow in the solar wind, creating a protective cavity called the *magnetosphere*. A bow shock forms at about 10 Earth radii on the sunlit side of our planet. The location of the bow shock is highly variable because it is 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. Plasma in the solar wind is deflected at the bow shock (*left*), flows along the magnetopause into the magnetic tail (*right*), and then can be injected back toward the Earth within the plasma sheet (*center*). The Earth, its auroras, atmosphere, and ionosphere, and the two Van Allen radiation belts all lie within this magnetic cocoon

deflected around the Earth, pulling the terrestrial magnetic field into a long magnetotail on the night side. 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.

Focus 9.5 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 ram pressure, P_R, exerted on a stationary body by a gas, or fluid, of mass density, ρ, moving at a velocity, V, is given by:

$$P_R = \rho V^2. \tag{9.17}$$

This expression can also be used for the ram pressure exerted on a body that is moving at this velocity through a gaseous or fluid medium.

For the solar wind, we can use:

$$P_R = m_P N_P V^2, \tag{9.18}$$

where the proton mass $m_P = 1.6726 \times 10^{-27}$ kg, the N_P is the number density of protons at the distance from the Sun that is under consideration, and V is the solar wind velocity at that distance.

The magnetic pressure, P_B, at the surface of the planet is given by:

$$P_B = \frac{B_0^2}{2\mu_0}, \tag{9.19}$$

where B_0 is the equatorial magnetic field strength, $\mu_0 = 4\pi \times 10^{-7}$ N $A^{-2} = 1.2566 \times 10^{-6}$ N A^{-2} 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.

The standoff distance, R_{MP}, from the planet at which the two pressures are equal, or where $P_R = P_B$, therefore occurs when:

$$\frac{R_P^6 B_0^2}{2\mu_0 R_{MP}^6} = m_P N_P V^2, \tag{9.20}$$

where the planet's radius is R_P. Solving for R_{MP} we have:

$$R_{MP} = \left(\frac{B_0^2}{2\mu_0 m_P N_P V^2}\right)^{1/6} R_P. \tag{9.21}$$

Example: **Solar-wind bow shock distance for the Earth and Jupiter**

When the solar wind encounters the Earth the number density of protons is $N_P \approx 5 \times 10^6$ m^{-3}, and the average solar wind velocity is about $V = 600$ km s^{-1} (see Table 9.3). The equatorial magnetic field strength of the Earth is $B_{0E} = 3 \times 10^{-5}$ tesla. Substituting these numbers into the equation for the standoff point, where the solar wind ram pressure equals the Earth's magnetic pressure, gives $R_{ME} \approx 7R_E$, or seven times the Earth's radius.

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. The giant planets have stronger magnetic fields than the Earth, so

the bow shock distances for Jupiter and Saturn, for example, are 42 and 19 times the radius of these planets. The planet Venus has no global dipolar magnetic field, but the solar wind can be diverted around this planet by its thick atmosphere.

The planet Jupiter has an equatorial magnetic field strength of $B_{0J} = 4.28$ tesla, with a bow shock distance of $R_{MJ} = 42\ R_J \approx 3.0 \times 10^9$ m, where the radius of Jupiter is $R_J = 7.15 \times 10^7$ m. So, the magnetosphere of Jupiter is larger than the Sun, whose radius is $R_\odot = 6.955 \times 10^8$ m. The solar wind proton density is diminished as it spreads out to a greater volume between the Earth and Jupiter.

The Earth's magnetic shield is so perfect that only 0.1 % of the mass of the solar wind that hits it manages to penetrate inside. Yet, even that small fraction of wind particles has a profound influence on the Earth's nearby environment in space; they create an invisible world of energetic particles and electric currents that flow, swirl, and encircle the Earth.

9.4.2 Trapped Particles

One of the first scientific discoveries of the Space Age was the finding, by James A. Van Allen (1914–2006) and his students, of high-energy electrons and protons that girdle the Earth far above the atmosphere (Van Allen et al. 1959). They move within two belts that encircle the Earth's magnetic equator but do not touch it, like a gigantic, invisible, torus-shaped doughnut.

These regions sometimes are 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, and contain high-speed electrons and protons.

In 1907, about a half-century before the discovery of radiation belts, the Norwegian geophysicist Carl Størmer (1874–1957) showed how electrons and protons could be almost permanently confined and suspended in space by the Earth's dipolar magnetic field (Størmer 1907, 1955). 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 that it moves back toward the other pole.

Thus, the electrons and protons bounce back and forth between the north and south magnetic poles. It takes about 1 min 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 a half-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 one another or some external force distorts the magnetic fields.

The problem at the time that 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? The answer is the solar wind. They can 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. As postulated by the English physicist James Dungey (1923–), a solar magnetic field can open up the Earth's magnetic field when the two fields are pointing in opposite directions when they touch (Dungey 1961). With this orientation they can join one another and become linked, similar to how 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. Tons of high-energy particles may then flow into the magnetosphere through the opening before it closes again.

9.4.3 Earth's Magnetic Storms

The Earth's magnetic field can be compressed and distorted when a coronal mass ejection arrives at the Earth. The ejections have magnetic field strengths of about 30 nT, or 3×10^{-8} T, so the much stronger terrestrial magnetic field, which is about a thousand times more intense, at 3×10^{-5} T, usually provides good protection from them. However, if the magnetic fields of a coronal mass ejection and of the Earth are pointing in opposite directions when they meet, the two fields become linked, resulting in intense geomagnetic storms that cause compass needles to swing widely.

The flow of currents associated with these great magnetic storms can interfere with electrical power grids here on the Earth, creating voltage surges on long-distance power lines and overheating or melting the windings of transformers. They can send cities into complete darkness, especially in high-latitude regions where the currents are strongest, such as Canada, the northern United States and Scandinavia. This doesn't occur often, perhaps once a year; however, the potential consequences are serious enough to employ early warning systems.

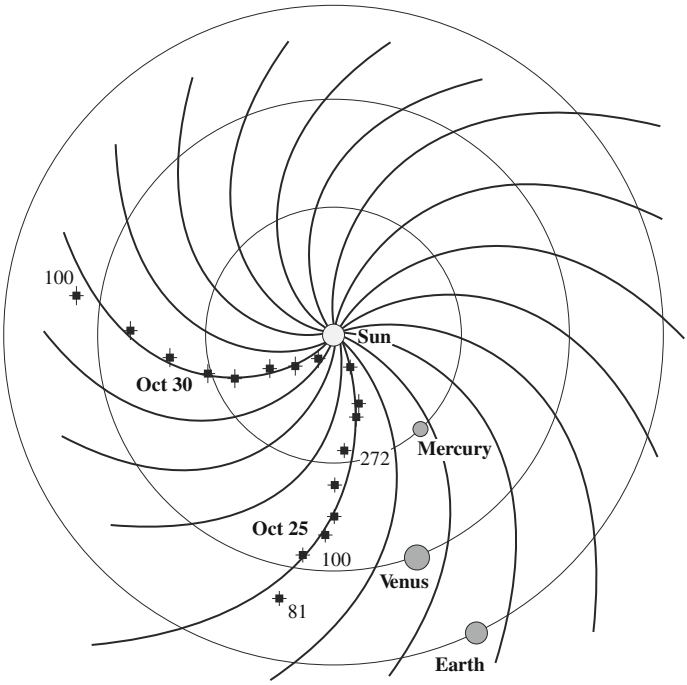


Fig. 9.13 Magnetic spiral The trajectory of flare electrons in interplanetary space as viewed from above the Sun's polar regions using the *Ulysses* spacecraft. The squares and crosses show *Ulysses* radio measurements of Type III radio bursts. As the high-speed electrons move out from the Sun into progressively more tenuous plasma, they excite radiation at successively lower plasma frequencies. The numbers denote the observed frequency in kilohertz (kHz). Because the flaring electrons are forced to follow the interplanetary magnetic field, they do not move in a straight line out from the Sun but instead travel along the spiral pattern of the interplanetary magnetic field, shown by the solid curved lines. The magnetic fields are drawn out into space by the radial solar wind and remain attached at one end to the rotating Sun. The locations of the orbits of Mercury, Venus, and the Earth are shown as circles. (Courtesy of Michael J. Reiner. *Ulysses* is a project of international collaboration between ESA and NASA.)

9.4.4 Solar Explosions Threaten Humans in Outer Space

When directed at our planet or at humans in deep space, both solar flares and coronal mass ejections produce dangerous gusts and squalls in the Sun's winds. Here on the ground, we are shielded from many of the effects by the Earth's atmosphere and magnetic fields, but out in space there can be no protection, and both humans and satellites are vulnerable.

Energetic charged particles generated during a solar flare threaten our planet only if the flare occurs at the right place on the Sun – that is, at one end of the spiral magnetic field that connects the Sun to the Earth. Given this circumstance, when a flare occurs near the west limb and the solar equator, the magnetic spiral

guides the high-speed charged particles that can threaten astronauts or satellites. The spiral magnetic pattern, produced when the solar wind carries the rotating Sun's magnetic field into surrounding space, has been confirmed by tracking the radio emission of charged particles thrown out during solar flares, as well as by spacecraft that have sampled the interplanetary magnetism near the Earth (Fig. 9.13).

As suggested by Cornell astrophysicist Thomas Gold (1920–2004), closed magnetic fields can be ejected from the Sun, generating shocks as they move into interplanetary space (Gold 1959a, b). And when a coronal mass ejection travels out into space, it can take the form of a magnetic cloud that moves behind an interplanetary shock (Fig. 9.14). The mass ejection plows into the slower-moving solar wind, driving huge shock waves that cross magnetic field lines and accelerate particles as they go. Following the shocks, the magnetic cloud can remain attached magnetically to the Sun, carrying its looping magnetic fields all the way to the Earth.

Because of their higher mass, it is solar-flare protons rather than electrons that provide the greatest threat to humans in outer space. Such high-speed protons, called solar energetic particles, can endanger the health and even the lives of astronauts when they are in outer space, unprotected by the Earth's magnetic field that deflects the charged particles.

High-energy protons from a solar flare or coronal mass ejection easily can pierce a spacesuit, causing damage to human cells and tissues, and even threaten the life of unprotected astronauts, who venture into space to unload spacecraft cargo, construct a space station or walk on the Moon, which has no global magnetic field or atmosphere, or Mars, which has no magnetosphere and only a thin atmosphere. So, solar astronomers keep careful watch over the Sun during space missions to warn of possible activity occurring in just the wrong place or time.

9.4.5 Disrupting Communication

Eight minutes after an energetic solar flare, a strong blast of x-rays and extreme ultraviolet radiation reaches the Earth and radically alters the structure of the planet's upper atmosphere, known as the ionosphere, altering its ability to reflect radio waves. During even moderately intense flares, long-distance radio communications can be silenced temporarily over the Earth's entire sunlit hemisphere. These radio blackouts are particularly troublesome for the commercial airline industry, which uses radio transmissions for weather, air traffic, and location information. The U.S. 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 to continuously monitor the Sun for intense flares that might severely disrupt military communications and satellite surveillance.

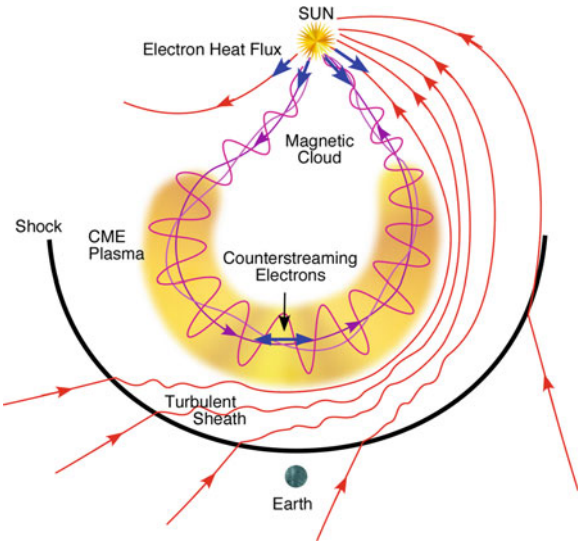


Fig. 9.14 Magnetic cloud When a coronal mass ejection (CME) travels into interplanetary space, it can create a huge magnetic cloud containing beams of electrons that flow in opposite directions within the magnetic loops that are rooted at both ends in the Sun. The magnetic cloud also drives a shock ahead of it. Magnetic clouds are present only in a subset of observed interplanetary coronal mass ejections. (Courtesy of Deborah Eddy and Thomas Zurbuchen.)

Although solar flares do not affect short-wavelength microwave signals that pass right through the ionosphere to communication satellites, solar explosions can destroy the satellites.

9.4.6 Satellites in Danger

More than 1,000 commercial, military, and scientific satellites are now in operation, affecting the lives of millions of people. And the performance and lifetime of all of these satellites are affected by Sun-driven space weather.

Geosynchronous satellites, which orbit the Earth at the same rate that the planet spins, stay above the same place on the Earth to relay and beam down signals used for cellular phones, global positioning systems, and Internet commerce and data transmission. They are endangered by the coronal mass ejections that cause intense geomagnetic storms. These satellites orbit our planet at about 6.6 Earth radii, or about 4,200 km, moving around the Earth once every 24 h. A coronal mass ejection can compress the Earth's protective magnetic fields from their usual location at about 10 Earth radii above the equator to below the satellites' synchronous orbits, exposing them to the full brunt of the gusty solar wind and its charged, energized ingredients.

Other satellites revolve around our planet in closer, low-Earth orbits at altitudes of 300 to 500 km, scanning the air, land, and sea for environmental change, weather forecasting and military reconnaissance. Space weather can increase the atmospheric friction exerted on these satellites, causing their orbits to decay more quickly than expected. The enhanced extreme ultraviolet and x-ray radiation from

solar flares heats the atmosphere and causes it to expand; similar or greater effects are caused by coronal mass ejections. The expansion of the terrestrial atmosphere brings higher gas densities to a given altitude, increasing the friction and drag exerted on a satellite, pulling it to a lower altitude, and sometimes causing ground controllers to lose contact with them. Space stations, for example, periodically must be boosted in altitude to a higher orbit to avoid a similar fate.

Infrequent, anomalously large eruptions on the Sun can hurl energetic protons toward the Earth and elsewhere in space. The solar protons can enter a spacecraft, producing erroneous commands and crippling their microelectronics. Such single-event upsets already have destroyed at least one weather satellite and disabled several communications satellites. However, to put the space-weather threat in perspective, only a few commercial satellites have been lost to storms from the Sun out of thousands deployed. The U.S. military builds satellites that can withstand the effects of a nuclear bomb exploded in space.

9.4.7 Forecasting Space Weather

Recognizing our vulnerability, astronomers use telescopes on the ground and in situ particle detectors or remote-sensing telescopes on satellites to carefully monitor the Sun, and government agencies post forecasts that warn of threatening solar activity. This enables evasive action that can reduce disruption or damage to communications, defense and weather satellites, as well as electrical power systems on the ground. Once it is known that a Sun storm is imminent, the launch of manned space flight missions can be postponed, and a walk outside a spacecraft or on the Moon or Mars can be delayed. Airplane pilots can be warned of potential radio communication failures. Operators can power down sensitive electronics on communication and navigation satellites until the danger passes. Utility companies can reduce load in anticipation of trouble on power lines, in that way trading a temporary "brown out" for a potentially disastrous "black out."

Everyone wants to know how strong a space storm is and when it is going to hit. Like winter storms on the Earth, some of the effects can be predicted days in advance. A coronal mass ejection, for example, arrives at the Earth one to four days after solar astronomers watch it leave the Sun. Solar flares are another matter; as soon as one is observed on the Sun, its radiation and fastest particles have already reached the Earth, taking just 8 min to travel from the Sun. One promising prediction technique is to observe when the magnetism on the Sun has become twisted into a stressed situation, because it then may be about to release a solar flare. Another technique employs helioseismology to look through the Sun and watch active regions develop before they rotate to face the Earth.

Chapter 10
The Sun Amongst the Stars

10.1 Comparisons of the Sun with Other Stars

10.1.1 How Far Away are the Stars?

To determine the distance of a nearby star (other than the Sun), astronomers measure its angular displacement when viewed from opposite sides of the Earth's orbit, or from a separation of twice the AU. The AU is the mean distance between the Earth and the Sun, with a value of about 149.6 million km. This angle is known as the *annual parallax*, from annual for the Earth's yearlong orbit and the Greek word *parallaxis* for the "value of an angle". Once the parallax is combined with the known value of the AU, the star's distance can be established by *triangulation*, the geometry of a triangle.

The measurement involves careful scrutiny of two stars that appear close together in the sky: a bright one relatively nearby and another fainter one much farther away (Fig. 10.1). The annual parallax of the nearer star can then be determined by comparing its position to that of the distant one for a year or more.

By definition, the annual parallax, denoted π_{A}, is half the apparent angular displacement of a nearby star observed against the more distant stars at intervals of six months from opposite sides of the Earth's orbit. That is:

$$\sin\pi_A = \mathrm{AU}/D \approx \pi_A \tag{10.1}$$

for a star at distance D and $1\ \mathrm{AU} = 1.496 \times 10^{11}$ m. As long as the stellar distances are much larger than the AU, which is always the case, the parallax angle π_A is small and $\sin\ \pi_A \approx\ \pi_A$. The distance D can therefore be given by:

$$D = \frac{1}{\pi_A}\ \mathrm{parsec}, \tag{10.2}$$

if π_A is given in seconds of arc, denoted by the symbol $''$. The name of this distance unit, the *parsec*, derives from the italicized part of the two words *par*allax and

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_10, © Springer-Verlag Berlin Heidelberg 2013

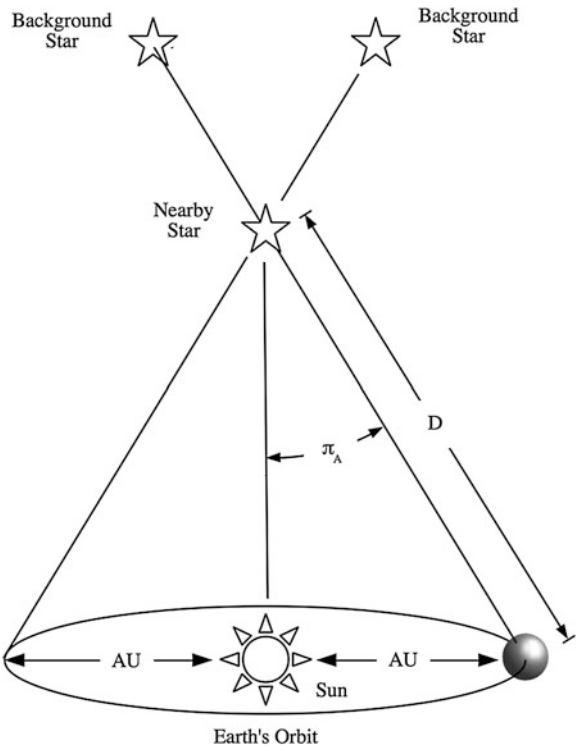


Fig. 10.1 Annual parallax When a distant and nearby star are observed at six-month intervals, from opposite sides of the Earth's orbit around the Sun, astronomers measure the angular displacement between the two stars. It is twice the annual parallax, designated by π_A, which can be used to determine the distance, D, of the nearby star. From trigonometry, $\sin \pi_A = AU/D \approx \pi_A$ for small angles, where 1 AU is the mean distance between the Earth and the Sun. The distance D to the star in units of parsecs is given by $1/\pi_A$, if the parallax angle is measured in seconds of arc. This angle is greatly exaggerated in the figure, for all stars have a parallax of less than 1 s of arc or less than 1/3,600th of a degree. The German astronomer Friedrich Wilhelm Bessel (1784–1846) announced the first reliable measurement of the annual parallax of a star in 1838

*sec*onds of arc. The parsec, abbreviated pc, is a convenient unit for measuring stellar distance since neighboring stars often are separated by about 1 parsec.

For conversion purposes, it is useful to know that:

$$1\,\text{parsec} = 1\,\text{pc} = 3.26\,\text{light years} = 3.0857 \times 10^{16}\,\text{m} = 206,265\,\text{AU}. \quad (10.3)$$

The light year is the distance light travels in one year, moving at the speed of light $c = 2.9979 \times 10^8\ \text{m s}^{-1}$ where 1 year $= 3.1557 \times 10^7$ s.

The fact that light travels through space at a constant speed provides another convenient unit of astronomical distance, by the time it takes light to move through space from the object to the Earth. This is known as the *light-travel time*. Light

from the Moon takes 1.5 s to reach the Earth, so we say that the Moon is 1.5 light-seconds away from the Earth. Sunlight takes 499 s or 8.3 min to cover the average distance between the Sun and the Earth. The nearest star (other than the Sun) is at a distance of a little more than 4 light-years.

Many of the brightest stars are hundreds of light-years away, and some stars in our Milky Way are many millions of light-years away. Starlight may reach us from stars that have now extinguished the internal nuclear fires that make them shine. The most distant objects in the universe are billions of light-years away, and the light we now detect from them was generated that long ago. So, radiation provides a method of looking back into time, to decipher the history of the universe – looking back to the time before the Sun and the Earth were formed, some 4.6 billion years ago.

The German astronomer Friedrich Wilhelm Bessel (1784–1846) made the first reliable determination of the annual parallax of a star, reporting a parallax of $0.31''$ for 61 Cygni, with an uncertainty of $\pm 0.02''$ (Bessel 1839). The modern measurement of $0.286''$ indicates that 61 Cygni is at a distance of 3.49 pc or 721,000 AU. Traveling at the speed of light, it takes 11.4 years to cross the distance from 61 Cygni to Earth.

Bessell was closely followed by the Scottish astronomer Thomas Henderson (1798–1844), who obtained a parallax of $1.19'' \pm 0.11''$ for the bright star Alpha Centauri (Henderson 1839). Bessell and Henderson were both observing stars that had exceptionally large motions across their line of sight, suggesting that they would be the nearest stars if all stars move at the same speed. Jackson (1956) and Hirshfeld (2001) describe attempts to obtain the first measurement for the distance of a nearby star (other than the Sun).

There is no known star, other than the Sun, whose annual parallax is greater than $1''$, or whose distance from Earth is less than 1 pc. The star with the largest parallax is Proxima Centauri, with a parallax of $0.769''$, making it the closest star to the Earth other than the Sun. The distance to Proxima is 1.30 pc, and light takes only 4.24 years to reach us from this star. It is about 268,000 times more distant than the Sun, indicating that there are vast, seemingly empty spaces between the stars.

Proxima Centauri is the nearest and dimmest companion of a triple star system that includes Alpha Centauri, the third brightest star in the sky. It takes about a million years for Proxima to orbit Alpha Centauri, and we only think it is associated with it because all three stars move together through space in about the same direction and with about the same speed. The large angular separation of Proxima Centauri from Alpha Centauri is a little more than 2.18°, or four times the angular diameter of the full Moon, which makes it possible to view the much dimmer companion outside the glare of the brighter star.

Because the Earth's atmosphere usually limits the angular resolution of a ground-based telescope to no more than $0.05''$, the annual parallax method can be used only for the very nearest stars, those that are closer than about 65 light-years or 20 pc. However, instruments aboard the ESA *HIPPARCOS* satellite, which orbited the Earth above its atmosphere in the 1990s, pinpointed the position of

more than 100,000 stars with an astonishing precision of $0.001''$ and obtained accurate measurements for the distances of stars up to 1,000 pc and 3,260 light-years away. This explains the spacecraft's name, which is an acronym for *HIgh Precision PARallax COllecting Satellite*; the name also alludes to the ancient Greek astronomer Hipparchus, who recorded accurate star positions more than 2,000 years ago. A successor to this mission is the ESA *GAIA* mission, short for *Global Astrometric Interferometer for Astrophysics*, currently scheduled for launch in March 2013. This mission is intended to measure 1 billion stellar distances to perhaps 10,000 pc and 32,600 light-years.

10.1.2 How Bright are the Stars?

The apparent brightness of a star is how bright it appears to us when its radiation reaches the Earth. The celestial positions and physical parameters of the ten brightest stars are given in Table 10.1, together with the brightest star, the Sun.

Because a human eye does not register directly the relative amount of radiation entering it, the Greek astronomer Hipparchus (c. 190 BC–c. 120 BC) divided the stars that he could see into six groups to better measure their relative brightness, relative to the eyes. This way of measuring brightness is called the *apparent visual magnitude* and is designated by the lowercase letter m or to be explicit about the visual aspect, by m_V with the subscript V denoting "visual". Hipparchus designated the brightest stars, such as Sirius or Rigel, with the first and most important magnitude, $m = 1$; Polaris and most of the stars in the Big Dipper were designated $m = 2$; and the faintest stars visible to the unaided eye received the sixth magnitude, or $m = 6$. Thus, in the magnitude system, brighter stars have lower magnitudes and fainter stars have higher ones.

About two millennia later, the British astronomer Sir Norman Pogson (1829–1891) noted that the stars of the first magnitude were 100 times as bright as stars of the sixth magnitude and that each magnitude unit is 2.512 times brighter than the next one down, where the number 2.152 is the fifth root of 100, or $100^{1/5}$ (Pogson 1856). The apparent magnitudes m_1 and m_2, of two objects of apparent brightness, or apparent radiation flux f_1 and f_2, are related by:

$$m_1 - m_2 = -2.512 \log\left(\frac{f_1}{f_2}\right) = 2.512 \log\left(\frac{f_2}{f_1}\right), \tag{10.4}$$

where the subscripts denote objects 1 and 2, and log denotes the logarithm to the base ten. An equivalent relation is:

$$\frac{f_1}{f_2} = 2.512^{(m_2 - m_1)} = 10^{0.4(m_2 - m_1)} = 2.512^{-(m_1 - m_2)} = 10^{-0.4(m_1 - m_2)}. \tag{10.5}$$

Table 10.1 The ten brightest stars as seen from Earth[a]

Star name	R.A. (2000)		Dec (2000)		m	Spectral Class[b]	D[c] (light-years)	L ($L_\odot$)	M	Mass ($M_\odot$)	R ($R_\odot$)
	h	m	°	′							
Sun					−26.74	G2 V	0.000016	1.0	+4.83	1.0	1.0
Sirius	06	45.2	−16	43.0	−1.46	A1 V	8.6	25.4	+1.42	2.02	1.71
Canopus	06	24.0	−52	41.8	−0.72	F0 Ib	310	13,600	−5.53	8.5	65.0
Alpha Centauri[d]	14	39.6	−60	50.0	−0.01	G2 V	4.3	1.52	+4.38	1.10	1.23
Arcturus	14	15.7	+19	11.0	−0.04	KI III	36.7	210	−0.29	1.5	25.7
Vega	18	36.9	+38	47.0	+0.03	A0 V	25.0	37	+0.58	2.14	2.5
Capella	05	16.7	+45	42.2	+0.08	G1 III	41	78	+0.20	2.6	9.2
Rigel	05	14.5	−08	12.1	+0.18	B8 Ia	772.5	66,000	−6.7	17.0	78.0
Procyon	07	39.3	+05	13.5	+0.34	F5 IV	11.46	7.73	+2.65	1.42	2.05
Achernar	01	37.7	−57	14.2	+0.50	B3 V	144	3,311	−2.77	6 to 8	10
Betelgeuse	05	55.2	+07	24.4	+0.42v	M2 Ia	643	140,000	−6.05	18 to 19	≈1,180

[a] The stars are listed in order of increasing apparent visual magnitude, m, or decreasing apparent brightness, for the brightest component if it is a binary system. The absolute magnitude is designated as M

[b] The luminosity classes are Ia = Supergiant of high luminosity, Ib = Supergiant of lower luminosity, II = bright giant, III = Normal giant, IV = Sub-giant, V = Main-sequence star, or dwarf star, VI = Subdwarf

[c] The luminosity, L, is in units of the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1} , the mass is units of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, and the radius, R, is units of the Sun's radius $R_\odot = 6.955 \times 10^{8}$ m

[d] Alpha Centauri is also known as Rigel Kentaurus

Table 10.2 Apparent visual magnitudes, m_V, of some astronomical objects

Object name	m_V
Sun	−26.74
Full moon	−12.7
Venus[a]	−4.5
Jupiter[a]	−2.5
Sirius	−1.44
Rigel	0.12
Saturn[a]	0.7
Polaris	1.97

[a] At maximum brightness when the planet is in the part of its orbit that brings it closest to the Earth

The apparent magnitude relation takes into account the nonlinear, roughly logarithmic response of the human eye to light intensity, and the apparent radiation flux is a technical name for a star's apparent brightness.

This logarithmic scale caused the very brightest stars to climb to negative apparent magnitudes. Sirius is $m = -1.44$, the planets Venus and Jupiter are a little brighter than Sirius, and the Sun is so close and bright that it is $m = -26.74$ (Table 10.2).

The number of stars increases dramatically with increasing apparent visual magnitude. There are 14 stars brighter than $m = 1$ and about 5,600 stars brighter than $m = 6$, which are all of the stars detectable by the unaided human eye. There are 335,000 stars brighter than $m = 10$, 1.5 million stars brighter than $m = 12$, and 4.8 billion stars brighter than $m = 25$. A backyard telescope can detect stars of apparent magnitude between 10 and 15; the *Hubble Space Telescope* can approach apparent magnitude 30. These stars are 4 billion times fainter than the human eye can see without a telescope, and there are 100 billion of them.

10.1.3 How Luminous are the Stars?

Luminosity is an intrinsic measure of a star, and it is not related to the star's distance from the observer. It is the amount of energy a star radiates per unit time in units of J s^{-1}, which also is the emitted power in watts, where 1 J s^{-1} = 1 watt. A star's luminosity usually is compared to the luminosity of the Sun, designated $L_\odot$, with a value of $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. There are stars that are 1 million times more luminous than the Sun (Table 10.3) and other stars that are 1 million times less luminous than the Sun. The exceptionally luminous beacons are rare, and the most common stars are not even as luminous as the Sun; they are so dim that telescopes are required to see them. The most luminous stars also are amongst the most massive, largest, and hottest stars, and the progressive decrease in stellar luminosity usually corresponds to a decrease in stellar mass and radius (see Table 10.3).

Table 10.3 The range in stellar luminosity[a]

Star name	Luminosity ($L_\odot$)	Mass ($M_\odot$)	Radius ($R_\odot$)	Temperature (K)
R 136a1[b]	8,700,000[c]	265	35.4	53,000
LBV 1806 ~ 20	5,000,000[c]	200	150	36,000
Pistol Star	1,700,000[c]	150	340	20,000
Betelgeuse	200,000[c]	19	1,180	3,500
Rigel	85,000[c]	17	78	11,000
Polaris	2,000	7.5	30	3,200
Aldebaren	425	1.7	44	4,010
Vega	37	2.1	2.5	9,600
Sirius A	25.4	2.0	1.7	9,940
Alpha Centauri A	1.5	1.1	1.2	5,790
Sun	1.0	1.0	1.0	5,780
Sirius B[d]	0.026	0.978	0.0084	25,200
Gliese 229B[e]	0.000006	0.03 to 0.05	0.1	950

[a] The luminosity is in units of the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, the mass is in units of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, the radius is in units of the Sun's radius $R_\odot = 6.955 \times 10^{8}$ m, and temperature is the effective temperature of the visible stellar disk in degrees kelvin, denoted K

[b] In the Large Magellanic Cloud, a nearby irregular galaxy and satellite of the Milky Way

[c] Bolometric luminosity

[d] Sirius B is an Earth-sized white dwarf star, which has a mass about equal to that of the Sun but has depleted its thermonuclear fuel

[e] Gliese 229B is a sub-stellar brown dwarf object, whose mass is below the lower limit, at about 0.08 solar masses, to sustain hydrogen fusion. This brown dwarf object has a radius about equal to that of Jupiter, which is very close to one tenth the Sun's radius, but a mass of 30–50 times the mass of Jupiter, or about 0.03–0.05 times the mass of the Sun

Astronomers observe an apparent brightness or *apparent radiation flux*, designated by the symbol f_S, which is the amount of radiant energy per unit time per unit area reaching the Earth. This quantity has units of J s^{-1} m^{-2}, and it depends on a star's distance from the Earth.

As stellar radiation travels out into space, it is distributed over an imaginary sphere of surface area $4\pi D_S^2$ at distance D_S from the star. The apparent radiation flux reaching a terrestrial observer from a star of luminosity L_S is:

$$f_S = \frac{L_S}{4\pi D_S^2} \tag{10.6}$$

This is sometimes called the inverse square law of light, since the apparent brightness, or radiation flux, falls off as the inverse square of the distance.

The apparent brightness of stars can be combined with measurements of their distances to determine their luminosity. Some of the stars that appear bright to the eye are relatively nearby and no more luminous than the Sun, but many are distant stars that are hundreds of thousands of times more luminous than the Sun. Thus, the exceptional brightness of the brightest stars, as seen from the Earth, can be due

to either the immense power of the radiation they emit or to their relative closeness, compared with other stars.

Example: The Sun's apparent brightness and intrinsic luminosity
Instruments aboard satellites have measured the Sun's apparent brightness, known as the solar constant, $f_\odot$. It is the total amount of radiant solar energy per unit area reaching the top of the Earth's atmosphere at the Earth's mean distance from the Sun, and the measurements indicate that $f_\odot = 1{,}361$ J s^{-1} m^{-2}. Using a mean distance of 1 AU $= 1.496 \times 10^{11}$ m, we obtain the Sun's luminosity $L_\odot = 4\pi\, f_\odot(\mathrm{AU})^2 = 3.828 \times 10^{26}$ J s^{-1}.

By extrapolating the Sun's radiation flux back to its visible disk, we can use that flux, designated $F_\odot$, to specify the disk's effective temperature, T_{eff}, by the relation (Sect. 3.4):

$$F_\odot = \frac{L_\odot}{4\pi R_\odot^2} = \sigma T_{eff}^4, \tag{10.7}$$

where the Sun's radius $R_\odot = 6.955 \times 10^8$ m and the Stefan-Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}. Given these quantities, we obtain $T_{eff} \approx 5{,}780$ K for the visible solar disk.

To compare stellar luminosities, distance has to be removed from the comparison, and a system of absolute magnitudes is used for that. The absolute magnitude, M, is defined as the apparent magnitude the star would have if it was at an arbitrary standard distance of 10 parsecs, 10 pc, or 32.6 light-years. So

$$M = m + 5 - 5\log D, \tag{10.8}$$

where D is the distance in parsecs. The absolute magnitude may also be derived from the apparent magnitude and the annual parallax by the formula

$$M = m + 5 + 5\log \pi_A, \tag{10.9}$$

where π_A is the annual parallax in seconds of arc denoted $''$.

Example: The Sun's absolute visual magnitude
The Sun has an apparent visual magnitude of $m_{v\odot} = -26.74$, and it lies at a mean distance of 1 AU $= 1.496 \times 10^{11}$ m $= 4.848 \times 10^{-6}$ pc, using the conversion of 1 parsec = 1 pc $= 3.0857 \times 10^{16}$ m. So the absolute magnitude of the Sun is given by $M_{v\odot} = -26.74 + 5 - 5 \log (4.848 \times 10^{-6}) = -26.74 + 5 - 5 \log (4.848) + 30 = 4.83$.

The luminosity, L_1 and L_2, of two objects of absolute magnitude, M_1 and M_2, are related by:

$$\frac{L_1}{L_2} = 2.512^{(M_2 - M_1)} = 10^{0.4(M_2 - M_1)} \tag{10.10}$$

where $2.512 = 100^{1/5} = 10^{0.4}$.

Absolute magnitudes range from about −12 to fainter than +20, but most stars are between −5 and +15 in absolute magnitude. It can have a negative value, so it is not absolute in a mathematical sense. The brightest star in the night sky, Sirius, has $M = +1.4$, but that is much less luminous than the seventh brightest star Rigel with $M = -8.1$. The apparent and absolute magnitudes of the ten brightest stars were included in Table 10.1. Our apparently brilliant Sun has an absolute magnitude of $M = +4.83$.

Since each step of 15 in absolute magnitude indicates a difference of a million in luminosity, there are stars with an absolute magnitude of −10 that are about a million times more luminous than the Sun and those with an absolute magnitude of 20 that are about a million times less luminous than the Sun.

You can convert absolute magnitude, M, to luminosity L through the relation:

$$\log\left(\frac{L}{L_\odot}\right) = 0.4(4.83 - M), \tag{10.11}$$

or equivalently:

$$L = 10^{0.4\,(4.83 - M)} L_\odot \tag{10.12}$$

where the absolute magnitude of the Sun is +4.83 and $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. The absolute magnitude, M, of a star of luminosity L is:

$$M = +4.83 - 2.5\log\left(\frac{L}{L_\odot}\right). \tag{10.13}$$

Example: Distance, luminosity, temperature, and size of the nearest stars

The Sun is the nearest star. It has a mean distance of 1 AU = 1.496×10^{11} m, an absolute luminosity of $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, an effective temperature of $T_{eff} = 5{,}780$ K, and a radius of $R_\odot = 6.955 \times 10^8$ m. The nearest star other than the Sun is Proxima Centauri, located at a distance of 4.243 light-years, where 1 light-year = 9.4605×10^{15} m. The ratio of the distance, D, of Proxima Centauri to the Sun's distance is $D/\mathrm{AU} = 2.68 \times 10^5$, so Proxima Centauri is 268,000 times farther away than the Sun and stars are separated by vast, seemingly empty space. Since 1 parsec = 1 pc = 3.26 light-years, the distance of Proxima Centauri is $D = 1.30$ pc, and its parallax is $\pi_A = 1/D = 0.769$ s of arc = $0.769''$. The spectral type of this star is M5.5, a cool red star with an effective temperature of $T_{eff} = 3{,}042$ K. The apparent visual magnitude of the star is $m_v = 11.05$, so its absolute visual magnitude is $M_v = m_v + 5 - 5 \log D = 11.05 + 5 - 5$

log (1.30) $\approx$ 15.5, where D is the distance in parsecs. In the visible range of wavelengths, the absolute magnitude of the Sun is $M_{v\odot} = 4.83$, so in visible light Proxima Centauri has a luminosity given by $\log(L/L_\odot) = 0.4(M_{v\odot} - M_v)$, or $L \approx 10^{-4.27} L_\odot = 0.000054\, L_\odot$ in visible light. At a temperature of 3,042 K, the most intense emission, from the Wien displacement law, is at a wavelength of $\lambda = 0.0029/T = 9.53 \times 10^{-7}$ m, or at infrared wavelengths. Most of Proxima Centauri's power is radiated at unseen infrared wavelengths, and the total luminosity over all wavelengths is $L \approx 0.0017\, L_\odot$. We can use the Stefan-Boltzmann law $L = 4\pi\sigma R^2 T_{eff}^4$, where $\pi = 3.14159$ and the Stefan-Boltzmann constant $\sigma = 5.670 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}, to infer the star's radius of $R \approx 10^8$ m $\approx 0.14\, R_\odot$ from the known values of L and T_{eff}. Assuming that the luminosity of a star varies as the fourth power of its mass, the mass of Proxima Centauri would be about a tenth of that of the Sun.

The difference between the apparent magnitude, m, and the absolute magnitude, M, of an object is related to its distance, D, by the distance modulus μ given by:

$$\mu = m - M = 5\log D - 5, \tag{10.14}$$

for D in parsecs, or

$$\mu = m - M = -5(1 + \log \pi_A), \tag{10.15}$$

where π_A is the annual parallax in seconds of arc, denoted by the symbol $''$.

The absorption, or extinction, of light by interstellar dust diminishes the light intensity and increases the apparent magnitude by an amount A. The distance modulus has to be corrected for this to yield the true distance modulus:

$$\mu = m - M - A = 5\log D - 5. \tag{10.16}$$

Example: The absolute magnitude and luminosity of Sirius A and Sirius B

The brightest star in the sky is Sirius, the Dog Star, often denoted Sirius A to distinguish it from its faint companion Sirius B. Sirius A has an apparent magnitude of $m = -1.47$ and a distance of $D = 8.60$ light-years. Divide by 3.26 to get the distance in parsecs, or $D = 2.64$ parsecs, and the absolute magnitude is $M = m + 5 - 5 \log D = -1.47 + 5 - 5 \log (2.64) = 1.42$. The luminosity of Sirius A is $L = 10^{0.4(4.83-M)} L_\odot \approx 23\, L_\odot$. The apparent magnitude of Sirius B is $m = 8.30$, which corresponds to an absolute magnitude of $M = 11.19$ and a luminosity of $L = 0.0029\, L_\odot$ for the star's visible light.

There is one more caveat to our magnitude system, for the visual magnitudes that are measured by the eye or a conventional optical telescope only sample the visible wavelengths, and a star can emit intense radiation outside our range of vision, including the ultraviolet or infrared part of the electromagnetic spectrum. Astronomers use the term *bolometric*, from the Greek word meaning "measure of rays" to indicate the total radiation output of a star, and it has both an apparent and absolute bolometric magnitude. Fortunately, stellar temperature, which can be determined by other methods, can be used to determine the amount of radiation that was not detected in the visible range, and apply a bolometric correction to convert visual magnitudes to bolometric ones. And this brings us to the temperature of the stars.

10.1.4 The Temperatures of Stars

An understanding of the physical properties of a star requires knowledge of its temperature as well as its luminosity. The effective temperature of the Sun's visible disk, the photosphere, is inferred from the solar radius and luminosity, with a value of $T_{eff} = 5{,}780$ K. However, we do not have direct knowledge of the radius of most stars. They are too far away and too small in angular size for a telescope to resolve them.

Fortunately, there are two methods to infer a star's temperature even when we do not know its size and luminosity. We can estimate the temperature from the color of the star or infer its temperature from the relative intensities of absorption lines observed in its spectrum. These are the effective temperatures of the visible stellar disks, or the stellar photospheres. The photospheres of the hottest stars that we can see have temperatures of more than 100 times those of the coolest stars, with a range between 2,000 and 50,000 K. Böhm-Vitense (1981) has provided a review of the stellar effective temperature scale.

These are the stars we look at, but our eyes do not see all of the radiation that a star produces. At extreme hot or cold temperatures, a star can become visibly dim, even invisible, because most of its radiation is produced outside the visible part of the radiation spectrum and often is absorbed in the Earth's atmosphere before reaching the ground.

A very hot star, with a temperature of more than 100,000 K, emits most of its light at ultraviolet wavelengths that are absorbed in the atmosphere. These hottest stars are exceptionally luminous and massive.

The coolest star-like objects emit most of their radiation at infrared wavelengths, also absorbed in the Earth's atmosphere and outside our range of vision. There are the substellar, brown dwarf objects, for example, that do not have enough mass to begin nuclear fusion of hydrogen in their core. These stellar disks emit heat associated with their formation or by burning deuterium that was already present in them. The brown dwarf objects are sometimes colder then room temperature, or below 300 K.

10.1.5 The Colors of Stars

There are reddish stars like Betelgeuse and Antares, yellowish stars like the Sun and Capella, and whitish stars like Vega and Sirius. These colors provide a rough indication of the temperature of a star's photosphere. As the temperature rises, the colors change from red – near 3,000 K, to yellow – around 6,000 K, to white – at about 10,000 K.

A star often has a certain color because most of its radiation is emitted at the wavelengths corresponding to that color. The wavelength of maximum starlight intensity varies inversely with temperature. A blue star, for example, is hotter than a red star. The coloring of a star, however, is very subtle, depending on the relative amount of light seen in different colors.

Blue-colored stars, for example, are not just bluer than red stars. For a star of the same radius, a blue star is more luminous than a red one. Exceptionally hot stars emit most of their radiation at invisible ultraviolet wavelengths, and such a star is even more luminous. There is enhanced radiation intensity at adjacent wavelengths, and this ultraviolet spillover produces more blue light than expected for a cooler star. In this case, the temperature of the star is much hotter than that inferred from blue light alone.

Astronomers therefore decided to quantify color by comparing the apparent magnitudes measured in different wavelength bands. They are denoted U, B, and V for ultraviolet, blue, and visual bands, and centered at wavelengths of 350, 450 and 550 nm. The apparent magnitude differences $m_B - m_V = B - V$ or $m_U - m_B = U - B$, are then determined. The apparent magnitude of a star, denoted m, usually refers to its visual apparent magnitude, also written m_V.

The difference between the amount of light received at one color and the amount at another is known as the *color index*, which is usually measured by the difference between blue, designated B, and visual, denoted V, bands with a color index denoted by $B - V$. It provides a reasonable estimate of photosphere temperature by using the ratio of luminosities at two wavelengths, which is better than a temperature estimated from observations at only one wavelength. The temperatures increase from about 3,150 K for $B - V = 2.0$ to 60,000 K at $B - V = -0.4$, reflecting the fact that hotter stars emit more blue light.

However, in addition to ultraviolet spillover into the blue colors, interstellar dust reddens starlight as it travels through space to arrive at the Earth, and the amount of reddening increases with a star's distance. Thus, the observed colors may not reliably reflect the emitted colors. A star's spectral lines provide a more accurate indication of the temperature of a star's photosphere.

10.1.6 The Spectral Sequence

More than a century ago, astronomers noticed that stars of different colors exhibit different spectral lines. Strong absorption lines of hydrogen, for example, dominate the spectra of white stars like Vega and Sirius, whereas some blue stars have noticeable helium absorption lines. Yellow stars like the Sun have strong absorption lines of calcium and heavier elements, called metals, in their spectra.

The different spectral lines that are emitted by stars depend on the physical conditions in the visible disk – the photosphere – and therefore the level of ionization of the emitting atoms (also see Sect. 6.4). Stars that display spectral lines of highly ionized elements must be relatively hot, because high temperatures are required to ionize atoms. These hot stars have relatively weak hydrogen lines because nearly all of the hydrogen is ionized and all of its electrons have been set free from their atomic bonds, no longer emitting or absorbing radiation. In other words, stars that display hydrogen lines have moderate photosphere temperatures. Those exhibiting molecular lines have even cooler temperatures because molecules break apart into their component atoms when the temperature increases.

A system of stellar classification based on spectra was developed in the early twentieth century and is still in use today. Working under the direction of Edward C. Pickering (1846–1914), astronomers at the Harvard College Observatory examined the spectra of hundreds of thousands of stars. The astronomers were mainly women who had studied physics or astronomy at nearby women's colleges, including Wellesley and Ratcliffe. Harvard did not educate women at that time and did not permit women on its faculty.

One of these faithful, stalwart workers was Annie Jump Cannon (1863–1941), who classified the spectra of roughly 400,000 stars in her lifetime (Cannon and Pickering, 1918–1924). She distinguished the stars on the basis of the absorption lines in their spectra and arranged most of them in a smooth and continuous spectral sequence. The hottest stars, with the bluest colors, were designated as spectral type O, followed in order of declining photosphere temperature by spectral types B, A, F, G, K, and M (Table 10.4).

Table 10.4 The spectral classification of stars[a]

Class	Dominant lines	Color	Color index	Effective temperature	Examples
O	He II	Blue	−0.3	28,000–50,000	χ Per, ε Ori
B	He I	Blue–White	−0.2	9,900–28,000	Rigel, Spica
A	H	White	0.0	7,400–9,900	Vega, Sirius
F	Metals; H	Yellow–White	0.3	6,000–7,400	Procyon
G	Ca II; Metals	Yellow	0.7	4,900–6,000	Sun, α Cen A
K	Ca II; Ca I	Orange	1.2	3,500–4,900	Arcturus
M	TiO; Ca I	Orange-Red	1.4	2,000–3,500	Betelgeuse

[a] An H denotes hydrogen, He is helium, Ca is calcium, and TiO is a molecule. The Roman numeral I denotes an electrically neutral, unionized atom, the number II describes an ionized atom missing one electron, and the temperatures are in degrees kelvin, denoted K

Cannon further refined each spectral class by adding numbers from 0 to 9, running from hot to cold; the larger the number, the cooler the star in that class. For example, the hottest F star is designated as F0 and the coolest as F9, followed by G0. In this system, our Sun is classified as G2.

10.1.7 Radius of the Stars

Large stars come in two varieties: the giants and the supergiants. A relatively common type of big star is the red giant star, which is about 100 times bigger than the Sun; the other, exceedingly rare kind, the supergiant, is about 1,000 times larger than the Sun. The benchmark size is the radius of the Sun, denoted $R_{\odot}$, with a value of $R_{\odot} = 6.955 \times 10^8$ m.

The red giants can be found almost anywhere in the night sky, whereas the supergiants are sparsely scattered within the Milky Way. Only one in a million stars is likely to be a supergiant. As the name implies, supergiants are simply extreme examples of the giant stars. They are the rare anomalies that stand out because of their size. They are exceptionally big, massive, luminous and often bright. Well-known examples of both types of large stars are given in Table 10.5.

We can measure the angular size of the largest stars using an interferometer that employs two or more connected mirrors. The radiation waves detected by any two of the mirrors are combined to produce an interference pattern – hence, the name *interferometer*, short for "interference-meter". If the waves of electromagnetic radiation detected by the two mirrors are in phase when combined, their wave

Table 10.5 Some well-known large stars[a]

Star name	Radius ($R_{\odot}$)	Luminosity ($L_{\odot}$)	Mass ($M_{\odot}$)	Temperature (K)
Supergiant stars				
VY Canis Majoris	≈2,000	≈450,000	≈40	≈3,000
VV Cephei A	≈1,900	≈300,000	≈30	≈3,300
Mu Cephei	1,650	60,000	15	3690
Betelgeuse	1,180	140,000	19	3500
Antares	800	65,000	15	3500
Red giant stars				
Mira A	400	9,000	1.2	3,000
R Doradus	370	6,500	≈1.0	2,740
Aldebaren	44.2	425	1.7	4,010
Polaris	30	2,200	7.5	7,200
Arcturus	25.7	210	1.1	4,300
Pollux	8.0	32	1.86	4,865

[a] The radius, R, is units of the Sun's radius $R_{\odot} = 6.955 \times 10^8$ m, the luminosity, L, is in units of the Sun's luminosity $L_{\odot} = 3.828 \times 10^{26}$ J s^{-1}, the mass, M, is units of the Sun's mass $M_{\odot} = 1.989 \times 10^{30}$ kg, and the temperature, T, is the effective temperature of the stellar disk in degrees kelvin, abbreviated K

crests combine and strong light is detected. When they are out of phase, one wave crest matches the trough of the other and they cancel one another. In the earliest applications, the two mirrors were separated gradually to produce a set of light and dark bands, or "fringes"; when the fringes disappeared, the star was resolved. The angular diameter of the source, in units of radians, is the ratio of the wavelength to that mirror separation in which the fringes disappear.

The American physicist Albert A. Michelson (1852–1931) was one of the first to describe the interferometer technique (Michelson 1890), and thirty years later, he teamed up with the American astronomer F.G. Pease (1881–1938) to use an interferometer to measure the size of Betelgeuse. They mounted two moveable mirrors and two fixed mirrors on a 20 foot (6 m) steel beam that was placed across the frame of the 2.5 m (100 inch) Hooker telescope on Mount Wilson. By measuring the mirror separation when the interference fringes disappeared, they concluded that Betelgeuse has an angular diameter of about $\theta = 0.05''$, where the symbol $''$ denotes seconds of arc (Michelson and Pease 1921). By way of comparison, if the Sun were placed at the distance of the next nearest star, Proxima Centauri, it would have an angular diameter of approximately $0.007''$.

Modern visible-light interferometry has been used to measure the angular diameters of about 100 stars, including both supergiant stars and relatively near red giant stars. Current observations of the angular diameter of the supergiant Betelgeuse, at $0.055''$, indicate that it has a radius of 1,180 solar radii, which is equivalent to 5.48 AU – where 1 AU $= 1.496 \times 10^{11}$ m is the mean distance between the Earth and the Sun. The orbital distance of Jupiter from the Sun is 5.2 AU; therefore Betelgeuse and other supergiant stars would fill much of our major planetary system. The *Hubble Space Telescope* (*HST*) was used to obtain an image of Betelgeuse, obtaining the first direct picture of the visible disk of any star other than the Sun. Interferometric measurements in a recent 15 year period suggest that Betelgeuse may be shrinking, even though its visible brightness showed no significant dimming during the same period – a perplexing result (Townes et al. 2009)

Example: Measuring the radius of stars

The radius of the supergiant star Betelgeuse is $R_B = 1180\ R_\odot = 8.207 \times 10^{11}$ m, where the solar radius is $R_\odot = 6.955 \times 10^8$ m. So the radius of Betelgeuse is equivalent to 5.486 AU, where 1 AU $= 1.496 \times 10^{11}$ m is the mean distance between the Earth and the Sun. The star's distance is $D_B = 643$ light-years, with 1 light-year $= 9.460 \times 10^{15}$ m. The angular diameter θ_B of Betelgeuse is therefore $\theta_B = 2R_B/D_B \approx 2.7 \times 10^{-7}$ radians $= 0.0556''$, where $''$ denotes seconds of arc and 1 radian $= 206{,}265''$. The angular resolution, θ_r, of an interferometer consisting of two mirrors separated by a distance D_I is $\theta_r = \lambda/D_I$ radians for radiation at a wavelength λ. The interference fringes of such an interferometer would disappear when $\theta_r = \theta_B$, or at a mirror spacing of $D_I = \lambda/\theta_B = 2.59$ m for red light from Betelgeuse at a wavelength $\lambda = 700$ nm $= 7 \times 10^{-7}$ m.

The angular diameter of the red giant star R Doradus is $\theta_R = 0.057''$, about the same as Betelgeuse because the red giant is much closer, at a distance of about $D_R = 178$ light-years. After converting this angle into radians, we obtain a radius for Doradus of $R_R = D_R\,\theta_R/2 \approx 2.3 \times 10^{11}$ m $\approx 330\,R_\odot$, where $\theta_R = 2.76 \times 10^{-7}$ radians and $D_R = 1.68 \times 10^{18}$ m.

The red giant star Arcturus is so close, at a distance of $D_A = 36$ light-years $= 3.40 \times 10^{17}$ m, that its angular size has also been measured, at $\theta_A = 0.021'' = 10^{-7}$ radians. We therefore obtain a radius for Arcturus of $R_A = D_A\theta_A/2 \approx 1.7 \times 10^{10}$ m $\approx 25\,R_\odot \approx 0.1$ AU.

Optical interferometry with the Very Large Telescope Interferometer, abbreviated VLTI, on the Cerro Paranal in Chile, found the angular diameter of Proxima Centauri, the nearest star other than the Sun, to be $\theta_{PC} = 0.0010'' = 5 \times 10^{-9}$ radians, which at its distance of 4.24 light-years $= 4.0 \times 10^{16}$ m corresponds to a radius of about 10^8 m, or one-seventh that of the Sun and 1.4 times the radius of Jupiter $R_J \approx 7 \times 10^7$ m. The star's estimated mass is $0.123\,M_\odot = 2.446 \times 10^{29}$ kg, where $M_\odot$ denotes the mass of the Sun, and 129 times Jupiter's mass $M_J = 1.90 \times 10^{27}$ kg.

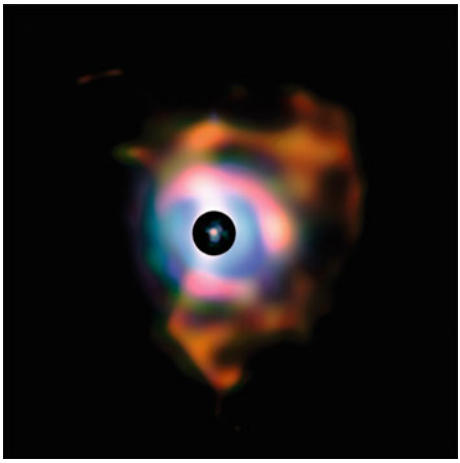


Fig. 10.2 The flames of Betelgeuse The red supergiant star Betelgeuse is slowly shedding its outer atmosphere, producing out-flowing gas that envelops the star and a much bigger nebula of gas and dust that surrounds it. The small circle in the middle of black disk denotes the edge of the supergiant's optically visible disk; it has a diameter of about 5.4 AU, where one AU is the mean distance between the Earth and the Sun. The black disk masks the bright central radiation of the star, in order to detect the infrared radiation of the outer plumes. They stretch to about 400 AU, or 60 million million, or 6×10^{13}, m from the supergiant Betelgeuse. (Courtesy of ESO/VISIR/VLT/Pierre Kervella.)

Supergiant stars are so large that they cannot hold onto their outer atmosphere. They are enveloped by dust and gas blown out by the stars' winds and have no well-defined apparent "edge" (Fig. 10.2).

Example: Mass loss from supergiant stars
The supergiant VY Canis Majoris has a mass, M, of about 40 times that of the Sun, or 40 $M_\odot$, but a radius, R, of about 2,000 solar radii, or 2,000 $R_\odot$, which is so far from the star's center that the stellar gravity GM/R^2 is but one hundred thousandth, or 10^{-5}, that of the Sun at its radius. The radius of the Sun is $R_\odot = 6.955 \times 10^8$ m and the Sun's mass is $M_\odot = 1.989 \times 10^{30}$ kg. The escape velocity (Sect. 3.2) required to overcome VY Canis Majoris' gravitational pull is $V_{esc} = (2GM/R)^{1/2} \approx 8.7 \times 10^4$ m s^{-1}, where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. The average thermal speed of a hydrogen atom in the visible disk of this star is (Sect. 5.2) is $V_{thermal} = (3kT/m_H)^{1/2} \approx 8.6 \times 10^3$ m s^{-1}, where the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1}, $m_H = 1.66 \times 10^{-27}$ kg, and the disk temperature of the star is $T = 3{,}000$ K. Since the average thermal speed is just 10 times less than the escape velocity, the atoms in the higher part of the Maxwell speed distribution should have little trouble overcoming the weak gravitational pull in the outer atmosphere of the star and breaking away from it. Images taken from the *Hubble Space Telescope* reveal arcs, filaments, and concentrations of material formed by the massive outflows from this supergiant star, some of them moving close to the star's escape velocity.

The next biggest known supergiant VV Cephei A is surrounded by opaque shells of a highly extended atmosphere, and is not entirely spherical in shape. The supergiant star Betelgeuse is enveloped by gas and dust that extends out to 400 AU from the star (Fig. 10.2).

For most stars, the radius is determined from the luminosity and temperature using the *Stefan-Boltzmann law*, which states that a star's luminosity increases with the square of its radius and the fourth power of its disk temperature. That is, the luminosity L_S of a star is intimately related to the star's radius R_S and effective temperature T_{eff}. If any two of these quantities are known, the third can be found using the Stefan-Boltzmann law

$$L_S = 4\pi\sigma R_S^2 T_{eff}^4 \tag{10.17}$$

where $\pi = 3.14159$ and the Stefan-Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J m^{-2} K^{-4} s^{-1} (Sect. 2.4). The Austrian physicist Joseph Stefan (1835–1893) first derived this law (Stefan 1879), on the basis of experimental measurements made by the English physicist John Tyndall (1820–1893), and Stefan's student Ludwig Boltzmann (1844–1906) derived it from theoretical considerations, using thermodynamics (Boltzmann 1872).

This means that

$$R_s = \left[\frac{L_S}{4\pi\sigma T_{eff}^4}\right]^{1/2}. \tag{10.18}$$

In this expression, the effective temperature, T_{eff}, is the temperature of a thermal (blackbody) gas emitting the observed luminosity, which is close to the temperature of the visible stellar disk, known as the photosphere. The radius is that of the photosphere, which is the level at which the stellar gases become opaque at visible wavelengths.

10.1.8 How Massive are the Stars?

The mass of a star usually is expressed in units of the mass of the Sun, $M_\odot = 1.989 \times 10^{30}$ kg, which is determined from Kepler's third law and the orbital period and distance of the Earth, or from the length of the year and the AU (Sect. 3.3). The range in the mass of most stars is relatively small, between about 0.1 and 100 $M_\odot$. The Sun is on the lower side of the stellar mass range, as are most stars. The most massive stars are relatively rare due to their relatively short lifetime.

Although there is not much variation between the masses of the stars, the mass of a star determines a star's luminosity, its effective disk temperature, the length of its life, and its ultimate fate. A small increase in a star's mass, for example, implies a big increase in its luminosity. Stars of lower mass have less weight pressing down on their core, so their core is cooler, the rate of their thermonuclear reactions is slower, and the stars are dimmer. The life span of stars also depends on their mass. The more massive a star is, the shorter its life span. A star of greater mass is more luminous, burns its nuclear fuel at a greater rate, and depletes its available energy in a shorter time.

When the measured masses of stars are combined with observations of the stars' luminosity, we find that stellar luminosity increases rapidly with increasing mass. The reason for this increase is the hotter temperature at the center of a high-mass star when compared to that of a low-mass star. The rate of nuclear reactions is greater at the higher temperature; therefore, the luminosity of the massive star is greater.

The English astronomer Arthur Stanley Eddington (1882–1944) first considered the theoretical aspects of such a mass-luminosity relation (Eddington 1924). For a star whose mass is supported by gas pressure, for example, the internal temperature scales directly with the mass, and the greater the mass, the hotter the central temperature and the greater the stellar luminosity.

Current data show that the luminosity, L_S, of most stars increases in rough proportion to the fourth power of the mass, M_S, with a stellar mass-luminosity relation (Fig. 10.3) given by:

$$L_S = \text{constant} \times M_s^{3.5}. \tag{10.19}$$

and

$$\log\left(\frac{L_S}{L_\odot}\right) = 3.5\log\left(\frac{M_S}{M_\odot}\right), \tag{10.20}$$

or

$$\left(\frac{L_S}{L_\odot}\right) = \left(\frac{M_S}{M_\odot}\right)^{3.5}. \tag{10.21}$$

where the subscript $\odot$ denotes the solar value, with $L_\odot = 3.828 \times 10^{26}$ J s^{-1} and $M_\odot = 1.989 \times 10^{30}$ kg. Notice that although the range of stellar masses is relatively small, usually between about 0.1 and 60 $M_\odot$ in mass, the stellar luminosity varies over about nine orders of magnitude from 10^{-3} to 10^6 $L_\odot$.

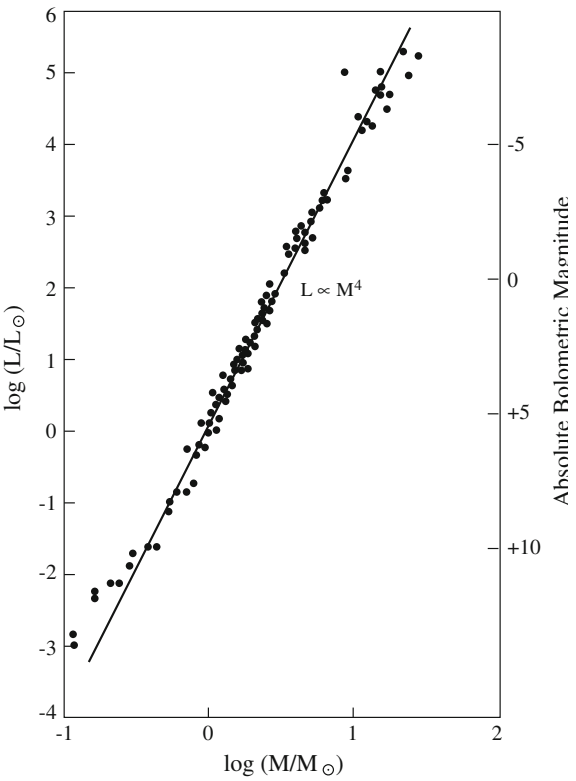


Fig. 10.3 Stellar mass–luminosity relation An empirical mass–luminosity relation for main-sequence stars of absolute luminosity, L, in units of the solar luminosity, $L_\odot$; and mass, M, in units of the Sun's mass, $M_\odot$. The straight line corresponds to a luminosity that is proportional to the fourth power of the mass. The English astronomer Arthur Eddington (1882–1944) proposed a theoretical explanation for this relation in 1924

At about 0.08 $M_\odot$ we reach the lower limit for a gaseous body to become a star. Its central regions are too rarefied and too cool to sustain the hydrogen-burning reactions that energize a Sun-like star and make it shine. Some of these nonstellar objects, known as brown dwarfs, can glow for a brief time as the result of heat generated during their formation by gravitational contraction. Although never hot enough for proton fusion, certain brown dwarf stars can shine for a time by burning deuterium that was present in the star at the time of its birth. The low-mass brown dwarfs eventually cool, compressing their near-stellar mass into the size of planets and disappearing from view.

As the mass increases, so does the central temperature. Nevertheless, the temperature inside a star cannot become too high, and therefore its mass cannot exceed an upper bound of 120 $M_\odot$. That is, the internal temperature and pressure of a very massive star can become so great that the star will be blown apart from inside (Focus 10.1). One of the most massive known stars is, for example, R136a1, with a mass of about 270 times that of the Sun, and it has been shedding a large fraction of its initial mass through a continuous stellar wind. It is estimated that, at its formation, the star held 320 solar masses and that it has lost 50 solar masses over the past million years.

Focus 10.1 The upper mass limit for a star

At formation, a star cannot keep on getting larger and more massive. There is a limit established when a star gets so big that the outward force of its internal radiation exceeds the inward gravitational force of the entire star.

Although the gas pressure of the hot, moving subatomic particles supports a star like the Sun, radiation pressure becomes important in more massive stars. As realized by the prolific Arthur Stanley Eddington (1882–1944), the outward pressure of radiation can exceed gas pressure in very massive stars (Eddington 1916), and a decade later he showed that this results in a maximum luminosity, now called the Eddington luminosity, at which the radiation blows away the outer atmosphere of a star (Eddington 1926a, b).

The temperatures within stars are high enough to ionize atoms, creating a plasma of free electrons and protons. As the radiation produced within the core of a star works its way out, the free electrons scatter the radiation with a Thomson scattering cross section $\sigma_T = 6.65246 \times 10^{-29}$ m^2. The outward force of the radiation, F_R, on the free electrons at distance r from the center of the star is

$$F_R = \frac{\sigma_T}{c} \frac{L_S}{4\pi r^2} \tag{10.22}$$

where L_S is the radiation luminosity, $\pi \approx 3.14159$, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The inward force of gravitation, F_G, of a star of mass M_S on a proton at distance r is:

$$F_G = \frac{GM_S m_P}{r^2}, \tag{10.23}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ N m^{2-} kg^{-2} and the proton mass $m_P = 1.6726 \times 10^{-27}$ kg. When the two forces are equal, at the maximum Eddington luminosity L_{Edd}, we have

$$L_S = L_{Edd} = \frac{4\pi G m_P c}{\sigma_T} M_S \approx 6.3\, M_S \approx 1.3 \times 10^{31} \frac{M_S}{M_\odot} \text{ J s}^{-1}, \tag{10.24}$$

or

$$L_S = 3.3 \times 10^4 \frac{M_S}{M_\odot} L_\odot \text{ J s}^{-1}, \tag{10.25}$$

for the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg and the solar luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}.

If the luminosity of a star reaches the Eddington luminosity, a significant proportion of the star's outer layers are ejected into space, and this therefore sets a limit to the mass the star can accumulate at formation. To a first approximation, we can use the mass-luminosity relation to obtain this limiting mass from:

$$L_S = \left(\frac{M_S}{M_\odot}\right)^{3.5} L_\odot \leq L_{Edd} \tag{10.26}$$

or

$$\left(\frac{M_S}{M_\odot}\right)^{2.5} \leq 3.28 \times 10^4 \tag{10.27}$$

to obtain the upper mass limit

$$M_S \leq 64\, M_\odot. \tag{10.28}$$

This is an approximate limit, and a more precise value can be obtained by setting the radiation pressure equal to the gas pressure, and assuming that the gravitational binding energy of the star, GM^2/R, is, from the virial theorem, 3 times the product of the gas pressure and the volume, resulting in an upper mass limit of $M \leq 110\, M_\odot$.

A direct measurement of stellar mass can be obtained from observations of the relative motion of two stars in a binary-star, or double-star, system. Popper (1980) gave us a review of stellar masses.

The members of a double-star system are in mutual orbit around one another, revolving about a common center of mass. If the orbital period and the distance

separating the two stars are measured, for example, the sum of their masses, $M_1 + M_2$, can be determined from Kepler's third law (Sect. 3.2):

$$M_1 + M_2 = \frac{4\pi^2 a^3}{GP^2} \tag{10.29}$$

for a linear star separation, *a,* and an orbital period, *P*, where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. Since orbits are always mutual (each star going about the other), the relative sizes of the orbits provide the ratio of the masses, and the combination of the sum and the ratio gives the individual masses of the stars can be found and compared to the Sun's mass.

Example: Measuring the mass of two stars in a binary system
Suppose the spectral lines of two stars shift back and forth with a period of $P = 2$ years $= 6.312 \times 10^7$ s, that the lines of star *1* shift twice as far as the lines of the other star 2, and Doppler shift observations of spectral lines indicates an orbital speed of $V = 100$ km s^{-1} for star *1* relative to star *2*. The semi-major axis *a* of the system can be determined from the circumference of one orbit $2\pi a = VP$, or $a \approx 10^{12}$ m. Then the sum of the masses can be calculated from Newton's expression of Kepler's third law, $M_1 + M_2 = 4\pi^2 a^3/(GP^2) \approx 1.5 \times 10^{32}$ kg, where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. Because the lines of star *1* move twice as far as those of star *2*, star *1* is half as massive as star *2*, with star *1* weighing in at about 0.5×10^{32} kg $= 25\ M_\odot$ and star *2* at about 1.0×10^{32} kg or $50\ M_\odot$, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg.

Some members of binary-star systems are tens of thousands of AU apart, whereas others touch one another. Moreover, the binary stars come in at least four varieties. There are visual binaries, the two components of which both can be resolved with a telescope and separately observed (Fig. 10.4); however, they are separated so widely that their orbital period about a common center of mass is often greater than a human lifetime.

The periodic motion of just one component of an astrometric binary is observed, whereas its companion is too faint to be seen. An eclipsing binary is a pair of stars whose orbital plane contains the Earth's line of sight, so we periodically observe the stars when they pass in front of or behind one another (Fig. 10.5).

A famous example of an eclipsing binary system is the two brightest stars in the Algol system: they have an orbital period of 2.87 days and a combined mass of about 4.5 solar masses. The stars are located at a distance of 28.5 pc, or 93 light-years, and are separated by only 0.062 AU, where 1 AU $= 1.496 \times 10^{11}$ m is the mean distance between the Earth and the Sun. The two stars are so close to one another that the more massive and bigger component has entered the gravitational sphere of influence of the other, transferring mass to it.

Fig. 10.4 Alpha Centauri Two of the most brilliant stars in the southern sky appear as a single star, named Alpha Centauri, to the unaided eye, but they can be resolved into two stars with the aid of binoculars or a small 5 cm (2 inch) telescope. The yellowish Alpha Centauri A (*lower left*), also known as Rigil Kentaurus, and the blue Alpha Centauri B (*upper right*) are locked together in a gravitational embrace, orbiting each other every 79.91 years. The two components of this binary-star system can approach one another within 11.2 AU and may recede as far as 35.6 AU, where the mean distance between the Earth and the Sun is 1 AU = 1.495×10^{11} m. Both stars have a mass comparable to that of the Sun, denoted $M_{\odot}$, of 1.1 $M_{\odot}$ and 0.90 $M_{\odot}$ for A and B, and a luminosity near that of the Sun, at 1.519 $L_{\odot}$ and 0.500 $L_{\odot}$. They appear bright because they are very nearby, at a distance of just 4.37 light-years. A third and faint companion Proxima Centauri has a luminosity of just 0.0017 $L_{\odot}$ and is located at about 15,000 AU or 2.2° from the two bright stars. At a distance of 4.24 light-years from the Earth, Proxima Centauri is the closest star other than the Sun. (Courtesy of ESO/Yuri Beletsky.)

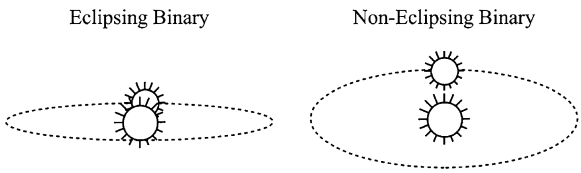


Fig. 10.5 Double stars Two close stars are joined in a gravitational embrace, orbiting each other and forming a binary star system (*left* and *right*). The orbital period and linear star separation can be used to determine the sum of their masses. If the orbital plane of the two companions is sufficiently inclined and within the line of sight, the star system becomes an eclipsing binary (*left*), in which one star is observed to pass behind the other and vice versa; this can provide additional information about the stars. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

Algol also is an example of the spectroscopic binary stars. The separation of this type of star pairs can be inferred from their orbital velocity and period. But this is an approximate determination, since the actual orbital velocities can be greater than the line-of-sight radial velocities measured from the Doppler shift of

the spectral lines. The spectral variations of a spectroscopic binary reveal the orbital motion of its unresolved components (Focus 10.2).

Focus 10.2 Determining the stellar mass in a spectroscopic binary system

The analysis of binary stars usually assumes circular motion about a center of mass located between two stars. We have:

$$r_1 M_1 = r_2 M_2 \tag{10.30}$$

with M_1 and M_2 being the masses and r_1 and r_2 their respective distances to the center of mass. Thus, if $a = r_1 + r_2$ is the separation between the masses,

$$r_1 = \frac{M_2}{M_1}(a - r_1) \tag{10.31}$$

or

$$r_1 = \frac{M_2}{M_1 + M_2} a \tag{10.32}$$

and

$$r_2 = \frac{M_1}{M_1 + M_2} a. \tag{10.33}$$

We can find the total mass of the binary system using Newton's version of Kepler's third law. The orbital period, P, is given by:

$$P^2 = \frac{4\pi^2}{G(M_1 + M_2)} a^3 \tag{10.34}$$

or

$$M_1 + M_2 = \frac{4\pi^2 a^3}{GP^2} \tag{10.35}$$

which provides the sum of the two stellar masses.

In spectroscopic binaries, the stars are usually not resolved, and their separations cannot be measured, but oscillations in the line-of-sight velocities are inferred from Doppler shifts of spectral lines. Because the perpendicular to the orbital plane is inclined to the line of sight by an angle i, the Doppler velocity amplitudes will be related to the true orbital velocity amplitudes by:

$$|V_{1obs}| = |V_1| \sin i, \tag{10.36}$$

and

$$|V_{2obs}| = |V_2| \sin i, \tag{10.37}$$

where the symbol $|x|$ denotes the absolute value of x. Since

$$|V_1| = \frac{2\pi r_1}{P}, \tag{10.38}$$

and

$$|V_2| = \frac{2\pi r_2}{P}, \tag{10.39}$$

then

$$\frac{|V_{1obs}|}{|V_{2obs}|} = \frac{r_1}{r_2} = \frac{M_2}{M_1}. \tag{10.40}$$

Replacing a with $r_1 + r_2 = P(|V_{1obs}| + |V_{2obs}|)/(2\pi \sin i)$ in Kepler's third law and using these expressions for r_1 and r_2 we can obtain:

$$(M_1 + M_2) \sin^3 i = \frac{P(|V_{1obs}| + |V_{2obs}|)^3}{2\pi G}. \tag{10.41}$$

If the spectrum of only one star, designated by the subscript *1*, is detected, due to the faintness of the second one, we can use:

$$(M_1 + M_2) \sin^3 i = \frac{P|V_{1obs}|^3 \left(1 + \frac{M_1}{M_2}\right)^3}{2\pi G} \tag{10.42}$$

or

$$\frac{M_2^3}{(M_1 + M_2)^2} \sin^3 i = \frac{P|V_{1obs}|^3}{2\pi G}. \tag{10.43}$$

When M_2 is much less than M_1, which would account for the faint luminosity of the second star, we obtain:

$$M_2 \sin i \approx \left(\frac{P}{2\pi G}\right)^{\frac{1}{3}} |V_{1obs}| M_1^{\frac{2}{3}}. \tag{10.44}$$

As might be expected, bigger stars are more massive, and there are fewer stars of high mass than those with low mass. The distribution of stars relative to mass is known as the *initial mass function*, with the term *initial* meaning the mass with

which the stars were formed before their subsequent evolution. The Austrian-born American astronomer Edwin E. Salpeter (1924–2008) derived the initial mass function for stars more massive than the Sun (Salpeter 1955), and found that the number of stars with masses in the range M to $M + dM$ is proportional to $M^{-2.35}$. The number falls off roughly as the inverse square of the mass and indicates that the star-formation process results in many more stars of low mass than high mass. When compared to the number of stars with a mass equal to that of the Sun, denoted $M_\odot$, there are roughly 100 times more stars with 1/10th of that mass, at 0.1 $M_\odot$, and about 1/100th fewer stars with a mass of 10 solar masses, or 10 $M_\odot$. Stars of higher mass are also bigger, whereas those of low mass are relatively small; the small stars outnumber the large stars.

10.2 Main-Sequence and Giant Stars

10.2.1 The Hertzsprung–Russell Diagram

Once the luminosity of stars was obtained from their brightness and measurements of their distance, astronomers were able to show that most stars exhibit a systematic decrease in luminosity as one progresses through the spectral sequence O, B, A, F, G, K, M. This progression is exactly what we would expect because the spectral sequence also denotes a scale of decreasing stellar temperatures, and the luminosity of a radiating body depends strongly on temperature.

The luminosity drop is illustrated in the famous *Hertzsprung–Russell diagram* (*H-R*) diagram, of luminosity or absolute magnitude plotted against the spectral class or effective temperature (Fig. 10.6). The diagram's name derives from the Danish astronomer Ejnar Hertzsprung (1873–1967), who plotted such diagrams for the Pleiades and Hyades star clusters, and the American astronomer Henry Norris Russell (1877–1957), who published an early version of this diagram for both noncluster and cluster stars (Hertzsprung 1911; Russell 1914).

Most stars, including the Sun, lie on the main sequence that extends diagonally from the upper left to the lower right, or from the high-luminosity, high-temperature blue stars to the low-luminosity, low-temperature red stars. The stars on the main sequence are the most common type in the Milky Way, constituting about 90 % of its stars.

The *Stefan-Boltzmann law* describes the general characteristics of the H-R diagram. It is given by:

$$L_S = 4\pi\sigma R_S^2 T_{eff}^4, \tag{10.45}$$

where L_S is the luminosity of the star, $\pi = 3.14159$, the Stefan-Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}, the radius of the star is R_S, and T_{eff} is the effective temperature of the visible stellar disk. This expression indicates that for a fixed radius, the luminosity of a star increases with the fourth power of the

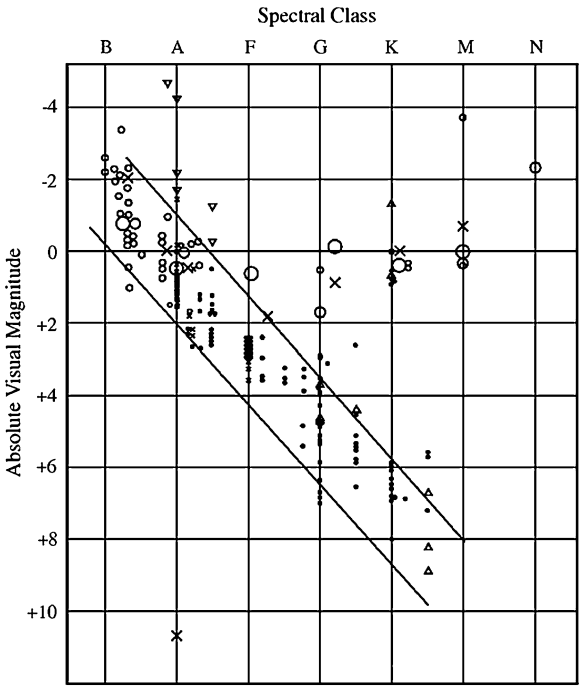


Fig. 10.6 Original Hertzsprung–Russell diagram The absolute luminosity, in magnitude units (*vertical axis*) plotted in 1914 by Henry Norris Russell (1877–1957) as a function of spectral class (*top horizontal axis*) for four moving star clusters: the Hyades (*black dots*), the Ursa Major group (*small crosses*), the large group in Scorpius (*small open circles*), and the 61 Cygni group (*triangles*). The *large circles* and *crosses* represent points calculated from the mean parallaxes and magnitudes of other groups of stars. The two *diagonal lines* mark the boundaries of Ejnar Hertzsprung's (1873–1967) observations of the Pleiades and Hyades open star clusters in 1911; this now is known as the main sequence along which most stars, including the Sun, are located. The giant stars are located at the *upper right*. In his publication, Russell included a similar diagram for individual bright stars, the distances of which had been established from stellar parallax measurements. It closely resembled the diagram shown here with an exceptional point in the lower left-hand corner, which is included here with an "x" mark. This star is the faint companion of a double-star system Omicron^2 Eridani, or 40 Eridani, now known to be a white dwarf star. [Adapted from Russell (1914).]

effective temperature; therefore, colder stars are less luminous. That is exactly what happens along the main sequence, for although the radius varies by a relatively small amount along the main sequence, the luminosity variation is due mainly to a change in temperature.

The observations also showed a different, unanticipated effect that gives the H-R diagram a peculiar shape. Some of the stars retain a high luminosity at decreasing temperature, in a band that extends to the upper right of the H-R diagram (see Fig. 10.6). This could be explained if the luminous cool stars were larger in radius than the less luminous ones, with the increase in size offsetting the

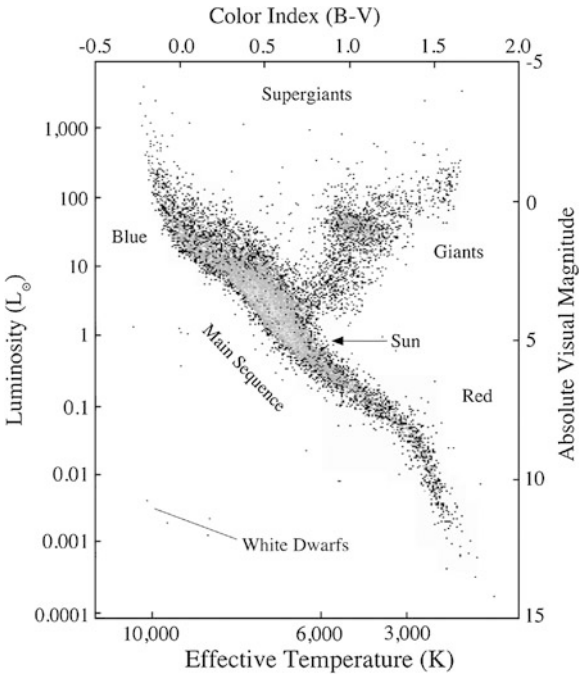


Fig. 10.7 Hertzsprung–Russell diagram for nearby stars A plot of the luminosity (*left vertical axis*) in units of the Sun's absolute luminosity, denoted $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, against the effective temperature of the star's disk in degrees kelvin, designated K (*bottom horizontal axis*) for 22,000 stars in the catalogue of the *HIPPARCOS* satellite. This plot is known as the Hertzsprung-Russell (H-R) diagram. The absolute visual magnitude (*right vertical axis*) and color index, B-V (*top horizontal axis*) are also designated. Most stars, including our Sun, lie along the main sequence, which extends from the high-temperature blue-white stars at the top left to the low-temperature red stars at the bottom right. The Sun is a main-sequence star with an absolute visual magnitude $M_V = 4.8$ and color index $B\text{-}V = 0.68$. The radiation from all main-sequence stars is sustained by hydrogen-burning reactions in their core. Stars of about the Sun's mass evolve into helium-burning red giant stars, located in the upper-right side of the diagram. Very rare bright giant stars and extremely scarce and luminous supergiants are found above the giant stars and along the top of the diagram. Faint and initially hot white dwarf stars are located in the lower left side. Due to their low luminosity, these endpoints of stellar evolution are relatively difficult to observe. (Data courtesy of the ESA/*HIPPARCOS* mission. From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

drop in temperature. The Stefan-Boltzmann law indicates that for a fixed temperature, the luminosity of a star increases with the square of the radius. If a star is 25,000 times more luminous than the Sun, with the same temperature, it follows from the Stephan-Boltzmann law that it will be 158 times bigger than the Sun.

Russell realized that he had found another type of star, which he named *giants* for their large size (Russell 1913). These stars have a radius as large as the mean

distance between the Earth and the Sun. Russell also used the name *dwarfs* for the more numerous main-sequence stars, because they are smaller than the giants, but the designation is confusing. There is no observable difference between the size and luminosity of the hottest dwarf and most giant stars, and the white dwarf stars are not even on the main sequence. In this book, therefore, we retain the designation *giant stars*, but use the term *main-sequence stars* for the other stars.

There are relatively few giant stars when compared to the number of stars on the main sequence. This is because stars spend the majority of their lifetime on the main sequence, and the giant stars belong to a subsequent and shorter-lived part of a star's evolution.

Nearly a century of increasingly accurate and extensive observations confirmed the initial characteristics of the H-R diagram (Fig. 10.7). To assist physical interpretations, the luminosities are displayed along the left vertical axis, and the color index, or equivalently, the effective temperature of the stellar disk is on the bottom horizontal axis.

Chiosi et al. (1992) have reviewed developments in our understanding of the H-R diagram. Reid (1999) reviewed the H-R diagram and the galactic distance scale after the *HIPPARCOS* mission and Lebreton (2000) has reviewed the mission's implications for stellar structure and evolution.

10.2.2 The Luminosity Class

There was an unresolved uncertainty in the H-R diagram, which created a dilemma for specifying the physical characteristics of a star. A star could be small or large as well as hot or cold. A red cool star, for example, might be either much more luminous than the Sun or much fainter. Once the spectral type establishes the temperature, the star could be on either the luminous giant or the dimmer main-sequence part of the H-R diagram. To resolve this ambiguity, astronomers found a way of classifying stars by their luminosity in addition to their spectral type.

Pioneering investigations by the American astronomer Walter S. Adams (1876–1956) and the German astronomer Arnold Kohlschütter (1883–1969) found that the relative intensities of certain neighboring spectral lines could be used to

Table 10.6 The Morgan–Keenan (M–K) luminosity classes

Ia	Bright supergiants
Ib	Supergiants
II	Bright giants
III	Giants
IV	Subgiants
V	Main sequence stars (or dwarfs)
VI (or SD)	Subdwarfs
D (or VII)	White dwarfs

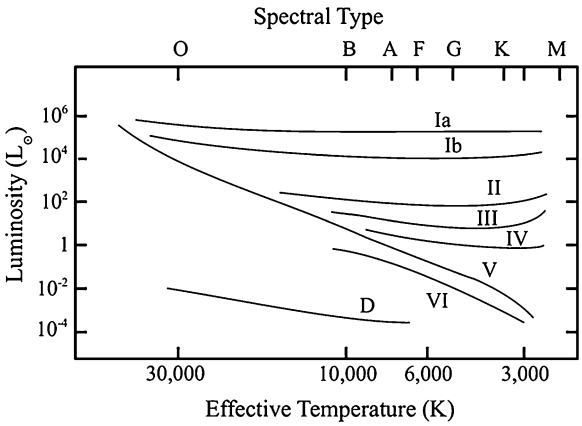


Fig. 10.8 Spectral type and luminosity class in the H–R diagram When both a star's spectral type (*top horizontal axis*) and luminosity class (*Roman numerals*) are known, the star's luminosity (*left vertical axis*) in units of the Sun's luminosity, denoted $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and effective disk temperature (*bottom horizontal axis*) can be obtained. The spectral types are shown at the coolest temperature for each type. A Roman numeral V designates the main-sequence stars: subgiants by IV, giants by III, bright giants by II, and supergiants by Ia and Ib. VI or SD denotes the subdwarfs, and D or VII designates the white dwarf stars. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

determine the luminosities of both main sequence and giant stars (Adams and Kohlschütter 1914).

In the mid-twentieth century, William W. Morgan (1906–1994) and Philip C. Keenan (1908–2000), of the Yerkes Observatory in Chicago, introduced the M–K system (Morgan et al. 1943), in which the most luminous and largest stars have the lowest numbers, given in Roman numerals (Table 10.6). In the M–K system, the Roman numeral III designates the giant stars, and V denotes the main-sequence stars; Class IV of subgiants is located between them. The most luminous, Class I stars are the supergiants, shown near the rarely occupied, upper edge of the H-R diagram. Both the spectral type and the M–K luminosity class can be specified in the H-R diagram (Fig. 10.8).

Because the spectral type O, B, F, G, K, or M depends solely on the physical properties of a star's outer atmosphere – the photosphere – it is not sufficient to determine the star's internal properties and evolutionary status. To solve this problem, both the spectral type and luminosity class are provided in a two-dimensional scheme; for example, the Sun is designated as G2 V.

If we know a star's luminosity class, we can find its luminosity, or absolute magnitude, and a star's distance can be inferred from the apparent magnitude and luminosity. This is known as the *spectroscopic distance*, or *spectroscopic parallax*.

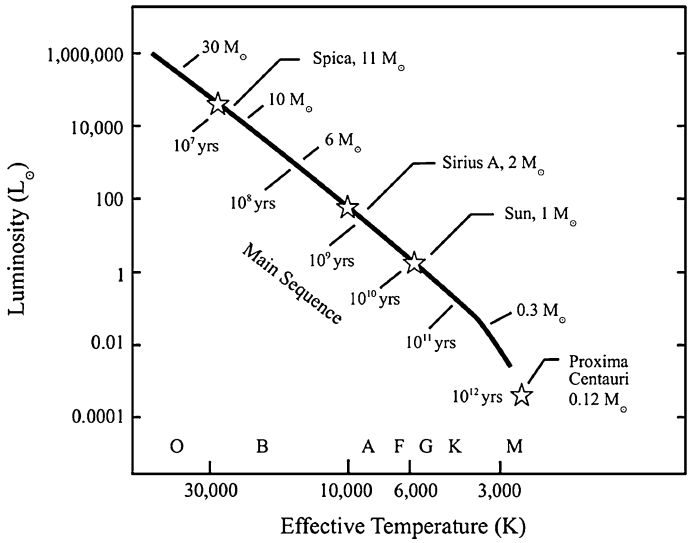


Fig. 10.9 Stellar mass and lifetime on the main sequence The relation between a star's luminosity (*left vertical axis*), in units of the Sun's luminosity, denoted $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and the star's effective disk temperature (*bottom horizontal axis*) in degrees kelvin, designated K, for the main-sequence stars in the Hertzsprung-Russell diagram. The stellar masses given along the main-sequence curve are in units of the Sun's mass denoted $M_\odot = 1.989 \times 10^{30}$ kg. Stars of higher mass are hotter and more luminous. All of these stars shine by hydrogen burning with a lifetime that also is denoted along the main-sequence curve. More massive stars burn their hydrogen fuel at a faster rate and have a shorter lifetime. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

10.2.3 Life on the Main Sequence

Today, the H-R diagram remains a primary tool for tracing the path of stellar evolution, but the routes are more complex than initially supposed. Russell, for example, thought that most stars began life as hot, blue-white stars and ended their life as cool red ones, moving from upper left to lower right along the main sequence, which is what might happen if the stars cool with time. But once scientists understood the ways that nuclear fusion makes a star shine, the early speculations proved to be wrong. The main sequence is not a singular evolutionary pathway, as once thought; it is simply a portrait of the sky at one moment, depicting different stars of varying mass. The giant stars represent later rather than earlier stages in a star's life cycle. As it turns out, a star begins its bright shining life on the main sequence.

Like our Sun, other stars on the main sequence generate energy by converting hydrogen into helium; as long as it shines in this way a star's position on the main sequence does not change substantially. It simply slowly becomes more luminous

Table 10.7 The main-sequence stars[a]

Spectral type	Effective temperature (K)	Mass ($M_\odot$)	Luminosity ($L_\odot$)	Radius ($R_\odot$)	Lifetime (years)
O5	44,500	60	7.9×10^5	12	3.7×10^6
B0	30,000	17.5	5.2×10^4	7.4	1.1×10^7
B5	15,400	5.9	8.3×10^2	3.9	6.5×10^7
A0	9,520	2.9	5.4×10	2.4	2.9×10^8
F0	7,200	1.6	6.5	1.5	1.5×10^9
G0	6,030	1.05	1.5	1.1	5.1×10^9
K0	5,250	0.79	0.42	0.85	1.4×10^{10}
M0	3,850	0.51	0.077	0.60	4.8×10^{10}
M5	3,240	0.21	0.011	0.27	1.4×10^{11}

[a] The mass, M_S, is in units of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, the absolute luminosity, L_S, is in units of the Sun's absolute luminosity, $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and the radius, R_S, is in units of the Sun's radius, $R_\odot = 6.955 \times 10^8$ m. The lifetimes are the amount of time required to exhaust the nuclear hydrogen fuel that supplies the energy of stars on the main sequence

and moves slightly to the upper right in the H-R diagram. Moreover, the different positions along the main sequence are closely related to only one property of the stars – their mass – and not to their different evolutionary status. The mass sets the central temperature and nuclear fusion rate at which the outward pressure and the inward gravitational force remain in balance.

The stellar masses decrease downward from upper left to lower right on the main sequence (Fig. 10.9). The high-mass stars are more luminous than the low-mass stars because the central temperatures of the former are higher, to support the greater mass, and their nuclear-reaction rates are faster, producing radiation of much greater luminosity. The hot, luminous O stars can have masses as high as 150 times that of the Sun, whereas the cool, dim, main-sequence M stars might have as little as 0.08 solar masses.

All of these main sequence stars shine by converting hydrogen into helium. The effective disk temperature, mass, luminosity, radius, and lifetime of main-sequence stars of different spectral types are listed in Table 10.7.

Because a star begins shining with a limited supply of hydrogen, it can remain on the main sequence for only a limited lifetime – that is, the time it takes to deplete all of the hydrogen fuel in its hot core. Although more massive stars certainly contain more hydrogen in their larger core, they are much hotter inside and fuse this hydrogen into helium at a faster rate, resulting in a shorter life. As indicated in Table 10.7, main-sequence lifetimes range from a few million to 100 billion years from spectral type O5 to M5.

A more massive star is hotter at its center than a less massive star, and it ought to be more luminous. It turns out that a star's nuclear energy supply is proportional to the mass, as indicated by Einstein's famous expression $E = Mc^2$, where E is the energy, M the mass, and c is the speed of light. The rate at which energy is being radiated away, the luminosity, L_S, also increases with the mass, but as the fourth power of the mass, M_S. So, the length of time, τ, that a star shines, is another

dramatic function of the mass; these times were also given in Table 10.7 for main-sequence stars.

How do we determine the length of time that a main-sequence star can continue to shine by converting protons into helium nuclei? These nuclear-fusion reactions are limited to the hot, dense stellar core. Outside of the core, where the overlying weight and compression are less, the gas is cooler and thinner so nuclear fusion cannot exist. For instance, the energy-generating core of the Sun extends about one quarter of the distance from the center to the visible solar disk. When all of the hydrogen within the core has been converted into helium, a star has exhausted its nuclear fuel supply and can no longer reside on the main sequence.

More than a half-century ago, the Brazilian astrophysicist Mario Schönberg (1914–1990) and the Indian-American astrophysicist Subrahmanyan Chandrasekhar (1910–1995) considered stellar models in which hydrogen is burned inside a star's core, or in a thin shell between the burned-out core and the overlying material (Schönberg and Chandrasekhar 1942). They found that it was impossible to construct models in which more than 12 % of the mass of the star is included in the exhausted core. This meant that the lifetime of a star on the main sequence is limited to the time it takes to convert 12 % of its hydrogen into helium.

The energy, ΔE, released in the conversion of four protons into one helium nucleus by hydrogen burning (Sect. 8.3) is $\Delta E = \Delta m c^2 = 0.007\ (4 m_P c^2)$, and the rest-mass-energy conversion process is just 0.007 or 0.7 % efficient. In this expression, Δm is the mass difference between the mass of four protons and the mass of the helium nucleus, m_P denotes the mass of a proton, and c is the speed of light. The mass difference is due to the binding energy liberated during nuclear fusion to make a star shine.

If we convert 12 %, or 0.12, of the Sun's mass into energy in this way, then the energy released is $E = 0.12 \times 0.007\, M_\odot c^2$ where the mass of the Sun $M_\odot = 1.989 \times 10^{30}$ kg and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The Sun's main-sequence lifetime, τ_{ms}, required to convert 12 % of its mass into helium is therefore:

$$\tau_{ms} = \left(\frac{E}{L_\odot}\right) = 0.12(0.007)\left(\frac{M_\odot c^2}{L_\odot}\right) \approx 3.92 \times 10^{17}\ \text{s} \approx 1.24 \times 10^{10}\ \text{years}, \tag{10.46}$$

where the luminosity of the Sun is $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and 1 year $= 3.156 \times 10^7$ s. Assuming that the luminosity of main-sequence stars increases with the 3.5 power of the mass, the main-sequence lifetime for a star of mass, M_S, is:

$$\tau_{ms} = 3.90 \times 10^{17} \left(\frac{M_\odot}{M_S}\right)^{2.5}\ \text{s}. \tag{10.47}$$

An exceptionally massive star, of say 100 times the mass of the Sun, will survive on the main sequence for only about 40 thousand years; in contrast, a star of moderate mass, say 10 times the mass of the Sun, will survive for 40 million years. That is why the more massive stars are so rare and hard to find.

Stars of intermediate mass, such as the Sun, will shine for 10–20 billion years. The Sun formed about 4.6 billion years ago; in another 7.8 billion years, it is expected to end its life on the main sequence.

So, the position of a star on the main sequence depends on its mass – the most massive stars being the most luminous and the more massive a star, the shorter it lives and the sooner it evolves off of the main sequence. Ninety percent of all main-sequence stars have a mass below 0.8 solar mass, and they have not yet had time to perish. They have been on the main sequence ever since they were born, thereby providing us with no information about stellar evolution. In contrast, some of the more massive stars, which were born long ago, have had enough time to burn up their available hydrogen fuel and advance to the next stage of stellar life. Thus, to understand *stellar evolution* we must examine the upper part of the main sequence in the *H-R diagram*, which applies to the more massive, shorter-lived stars.

10.2.4 The Red Giants and Supergiants

After the low-mass, main-sequence stars, the most common type of star is the red giant found in the upper right side of the H-R diagram (Fig. 10.10). These low-temperature stars are not exceptionally massive. They have an intermediate mass of roughly 1–10 times that of the Sun and are in a late state of stellar evolution from somewhat hotter main-sequence stars. Although cooler than the Sun, the red giants are about 100 times more luminous due to their much larger size, about 50 times the radius of the Sun. Prominent, bright-red giants include Aldebaran and Arcturus.

Because they are so luminous, we can see red giant stars that are relatively distant without using a telescope. However, they also are much less common than main-sequence stars because relatively few stars have entered this later stage of life. The red giants last only a few million years, which is a brief existence compared to the billions of years that stars of roughly solar mass spend on the main sequence.

The giant stars are enormously distended stars with a low mean mass density and a high luminosity. If we assume that the inner temperatures of giant stars are high enough to generate a gas pressure sufficient to balance gravitation, then their luminosity would greatly exceed that which is actually observed. This enigma was resolved almost a century ago when the great English astronomer Arthur Stanley Eddington (1882–1944) showed that radiation pressure must stand with gravitation and gas pressure as the third major factor in maintaining the equilibrium of a star

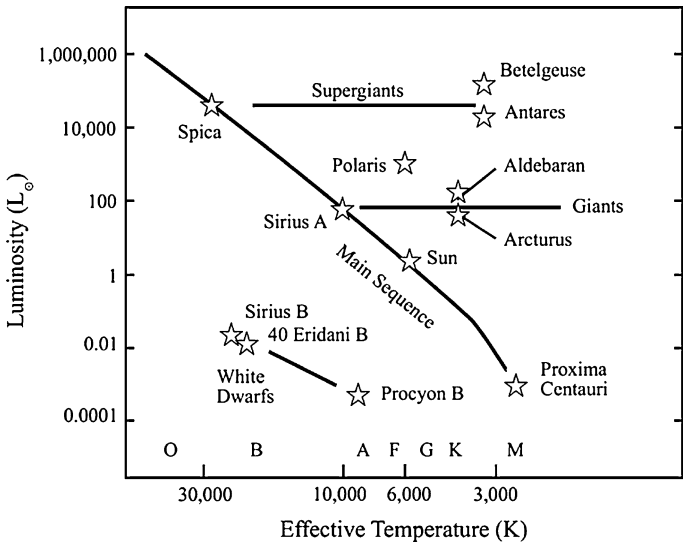


Fig. 10.10 Giants, supergiants, and white dwarfs The majority of stars occupy the main sequence in the H-R diagram. Stars with a mass comparable to that of the Sun will evolve into helium-burning giant stars, illustrated by Aldebaran and Arcturus in this diagram (*middle right*). These red giants are somewhat cooler than the Sun but about 100 times more luminous. The luminosity (*left vertical axis*) is in units of the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. More massive stars, which have a shorter lifetime on the main sequence, evolve into supergiant stars that are between 10 thousand and 1 million times as luminous as the Sun. Antares and Betelgeuse (*top right*) illustrate the supergiant stars on this diagram. After depleting all of their helium fuel, which is after the core hydrogen is exhausted, the less-luminous giant stars evolve into white dwarf stars of very low luminosity and initially hot disk temperatures. They are illustrated in the bottom left of this diagram by Sirius B, 40 Eridani B, and Procyon B. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

(Eddington 1917). The radiation pressure is the pressure exerted by electromagnetic radiation.

Although the outward pressure caused by the motion of gas particles, or the gas pressure, indeed does support the Sun and most other stars against the inward force of their immense gravity, it is insufficient for the much larger giant stars. They are also supported by radiation pressure, which increases with the fourth power of the temperature. In contrast, gas pressure is simply proportional to the temperature; so, if we sufficiently increase the central temperature, radiation pressure will become much larger than gas pressure.

Eddington also showed how some of the giant stars that are in radiative equilibrium could pulsate, with outer envelopes that move in and out, becoming alternately ionized and neutral during the course of pulsation (Sect. 14.1, Focus 14.1, Eddington 1918, 1919).

The supergiants are very massive, evolving from main-sequence stars of 10 to 100 solar masses, and exceptionally large, with radii of hundreds of times that of the Sun. They are also 10–100 times more luminous than the red giants. Antares and Betelgeuse are supergiant stars.

The supergiants are so exceedingly rare that we can only see them as a sparse sprinkling across the top edge of the H-R diagram. They are even less common than the O stars; rarely seen in any given part of the night sky, but so intrinsically luminous that we can see a few without a telescope.

Example: Properties of a large star

Suppose a star is ten thousand times more luminous than the Sun, with a stellar luminosity of $L_S = 10^4\ L_\odot$, where the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and a spectral type and effective temperature the same as that of the Sun, whose effective temperature is $T_{eff} \approx 5780$ K. We can use the Stefan-Boltzmann law $L_S = 4\pi\sigma R_S^2 T_{eff}^4$, where $\pi \approx 3.14159$ and the Stefan-Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J m^{-2} K^{-4} s^{-1}, to infer the star's radius $R_S \approx 6.9 \times 10^{10}$ m $\approx 100\ R_\odot$, where the Sun's radius $R_\odot = 6.955 \times 10^8$ m. The mass-luminosity relation indicates that the luminosity scales roughly as the 3.5 power of the mass, so the mass of the star is about $M_S = (L_S/L_\odot)^{1/3.5}\ M_\odot \approx 10\ M_\odot$ where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. The main-sequence lifetime of the star is about $\tau_{ms} \approx 10^{10}\ (M_\odot/M_S)^{2.5}$ years ≈ 14 million years.

Instruments aboard *HIPPARCOS* have obtained a parallax of $\pi_A = 3.78 \times 10^{-3}\ ''$ for the blue supergiant star Rigel, to give a distance of $D = 1/\pi_A \approx 264$ pc $= 8.15 \times 10^{18}$ m, where 1 pc $= 3.0857 \times 10^{16}$ m. Interferometer measurements indicate that Rigel has an angular diameter of $\theta = 2.75 \times 10^{-3}\ '' = 1.33 \times 10^{-8}$ radian, where 1 radian $= 2.06264 \times 10^5\ ''$, so its radius is $R = \theta D/2 \approx 5 \times 10^{10}$ m $= 78\ R_\odot$, where the Sun's radius $R_\odot = 6.955 \times 10^8$ m. Rigel's absolute bolometric magnitude $M = -7.84$, so its luminosity $L = 10^{0.4(4.83-M)} L_\odot \approx 10^5 L_\odot$, where the Sun's absolute magnitude is 4.83 and the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. We can determine the star's effective temperature, T_{eff}, from the Stefan-Boltzmann law $L = 4\pi\sigma R^2 T_{eff}^4$, where $\pi = 3.14159$ and the Stefan-Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J m^{-2} K^{-4} s^{-1}, obtaining $T_{eff} \approx 10^4$ K.

Both the giant and the supergiant stars are so large that their atmospheres are slowly blowing away with strong winds that carry their outer atmospheres into surrounding space. For example, the giant star Mira, "The Wonderful," is pulsating and also losing mass at about one millionth of a solar mass per year. That is more than 1 million times the mass loss rate of the Sun's wind.

Because giants and supergiants are former main-sequence stars that have exhausted their core supply of nuclear hydrogen, they must be undergoing other forms of nuclear fusion. We now consider these various nuclear burning reactions.

10.3 Nuclear Reactions Inside Stars

Clayton (1984) and Rolfs and Rodney (2005) have provided textbooks that include stellar nuclear reactions and nucleosynthesis.

10.3.1 The Internal Constitution of Stars

Although we cannot see the inside of a star, its internal structure can be explained by a few simple concepts, one of which is a star's equilibrium. Like the Sun, almost every star we see is neither collapsing nor expanding, and it remains the same size throughout most of its long life. At every point inside such a star, the inward pull of its gravity is balanced precisely by the outward push of its internal pressure.

As with the Sun, all of the other main-sequence stars are composed mainly of the lightest element, hydrogen, but it is too hot for whole hydrogen atoms to exist within them. These atoms are fragmented into their subatomic constituents by frequent collisions. The material in these stellar interiors therefore is in the plasma state, composed almost entirely of hydrogen nuclei, protons, and free electrons no longer attached to atoms. Compressed to high density, protons still occupy the vast empty spaces of former atoms, so plasma behaves like a perfect gas with a pressure that increases with the temperature.

The central temperature of any main-sequence star can be estimated by assuming that a proton at the center is hot enough and moving fast enough to counteract the gravitational compression on the proton from the rest of the star. For the Sun, this balance is achieved at a central temperature of 15.6 million K (Sect. 8.2). A more massive star produces greater compression at its center, so a higher central temperature is required to hold it up.

The temperature, T_C, at the center of a star like the Sun can be estimated by assuming that each proton down there is hot enough and moving fast enough that the thermal energy $3kT_C/2$ counteracts the gravitational compression it experiences from all the rest of the star. That is:

$$\frac{3}{2}kT_C = \frac{Gm_P M_S}{R_S}, \tag{10.48}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the mass of the proton is $m_P = 1.6726 \times 10^{-27}$ kg, and M_S and R_S respectively denote the mass and radius of the star.

Solving for the central temperature we obtain:

$$T_C = \frac{2Gm_P M_S}{3kR_S} = 1.54 \times 10^7 \left(\frac{M_S}{M_\odot}\right)\left(\frac{R_\odot}{R_S}\right) \text{K}. \tag{10.49}$$

In the numerical approximation, the mass and radius of the star are given in solar units, denoted by the subscript $\odot$, where $M_\odot = 1.989 \times 10^{30}$ kg and $R_\odot = 6.955 \times 10^8$ m. Thus, the temperature at the center of the Sun is about 15 million K.

The temperature at the center of a main-sequence star is proportional to its mass, so a star that is 10 times more massive than the Sun is 10 times hotter in its center. At a large enough mass, the star becomes so hot that it is blown apart; this explains why there are no known stars with a mass greater than about 120 times the mass of the Sun.

We have assumed that the outward gas pressure of the moving protons supports the inward pull of gravity, a condition known as hydrostatic equilibrium. The ideal gas law (Sect. 5.4) gives this gas pressure, P_g:

$$P_g = NkT, \tag{10.50}$$

which is equivalent to:

$$P_g = \frac{\rho kT}{\bar{m}} = \frac{\rho kT}{\mu m_H} \tag{10.51}$$

for a gas of mass density ρ and mean mass per particle given by:

$$\bar{m} = \rho/N = \mu m_H \tag{10.52}$$

for mean molecular weight μ. The mass of the hydrogen atom $m_H = 1.00794$ u $\approx 1.67 \times 10^{-27}$ kg, which is roughly equal to the atomic mass unit u $= 1.6605 \times 10^{-27}$ kg and good enough for order of magnitude estimates. For a fully ionized hydrogen gas, the mass per particle is $m_P/2$ or half the proton mass.

The gas pressure is the outward pressure caused by the motion of gas particles and increases with their temperature. Gas pressure does indeed support the Sun and most other stars against the inward force of their immense gravity, but this does not apply for much larger, and relatively rare, giant stars.

Eddington (1917) showed that in addition to the kinetic gas pressure, P_g, the radiation photons in a giant star exert an additional radiation pressure, P_r, given by:

$$P_r = \frac{aT^4}{3} \tag{10.53}$$

where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}. For radiative equilibrium at the center of star of radius, R_S, and mass, M_S:

$$\frac{4}{3}\pi R_S^3 P_r = \frac{GM_S^2}{R_S}, \tag{10.54}$$

where the gravitation constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

The total pressure, P, inside a star is given by the sum of the gas pressure and the radiation pressure. The radiation pressure is much less than the gas pressure at the center of the Sun, but it can compete with gas pressure in supporting giant stars. The complete equation of state for the pressure, P, is then given by:

$$P = P_g + P_r = \frac{\rho kT}{\bar{m}} + \frac{1}{3}aT^4. \tag{10.55}$$

Example: Supporting a star by gas pressure or radiation pressure
At the center of a star of mass density ρ_c and temperature T_c, the gas pressure is $P_g = NkT \approx \rho_c kT/m_p$, where the proton number density $N \approx \rho_c/m_p$, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and the proton mass $m_p = 1.6726 \times 10^{-27}$ kg. The radiation pressure, P_r, is given by $P_r = aT_c^4/3$, where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}.

At the center of the Sun $\rho_c \approx 1.5 \times 10^5$ kg m^{-3} and $T_c \approx 1.5 \times 10^7$ K, so the gas pressure $P_g \approx 2 \times 10^{16}$ Pa and the radiation pressure $P_r \approx 1.3 \times 10^{13}$ Pa, which is 1 thousand times less than the gas pressure. In other words, the gas pressure is about 1 thousand times greater than the radiation pressure at the center of the Sun.

Suppose a main-sequence star is one hundred times as massive as the Sun, with a mass $M_S = 100\ M_\odot$, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. The mass-luminosity relation indicates that the luminosity scales roughly as the 3.5 power of the mass, so the star's luminosity would be $L_S \approx 10^7\ L_\odot$, an exceptionally luminous star, where the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. If the effective temperature was about the same as that of the Sun, and because the luminosity varies as the square of the radius, the star's radius will be $R_S \approx 10^{3.75}\ R_\odot$, a large star, where the Sun's radius $R_\odot = 6.955 \times 10^8$ m. Assuming that the star is entirely composed of protons, then the central temperature, T_{CS}, of the star, which scales as M_S/R_S, will be $10^{-1.75}$ that of the Sun. The gas pressure, P_{gS}, varies as $M_S T_{CS}/R_S^3$ or as M_S^2/R_S^4, so it will be 10^{-11} that of the Sun. The central radiation pressure, P_{rS}, varies as T_C^4, or as $(M_S/R_S)^4$, which will be 10^{-7} that of the Sun, and the ratio of radiation pressure to gas pressure in the star center will be ten thousand times greater than that of the Sun. For this more massive and luminous star the central radiation pressure is estimated to be 10 times the gas pressure at the star's center.

Since the radiation pressure increases with the fourth power of the temperature, and the temperature has to increase with the mass, the radiation pressure can

overcome the gravity of an exceptionally massive star. That is, a giant star cannot remain in equilibrium if the central temperature and mass become too high, and this occurs for masses greater than about 120 solar masses (see Focus 10.1, Sect. 10.1).

But where does a star's heat come from? The energy released by nuclear fusion in the stellar core heats the gas and generates its pressure. That is, nuclear reactions that transform a light element into a heavier one liberate subatomic energy that sustains the high temperatures within a star. This energy also makes its way out of the star to provide its luminosity and keep it shining.

Thus, two other fundamental concepts in understanding a star's interior are: (1) the way energy is generated by nuclear reactions near its center, and (2) the methods in which the radiation produced by these reactions works its way out to the observed stellar disk, its photosphere. The energy generation depends on the nuclear fuel, as well as the mass density and temperature in a star's core. The radiation-energy transfer depends on a star's internal opacity to radiation, which prevents some of the radiation from escaping.

After arrival on the main sequence, which is designated the *zero age*, the internal structure of a star can be determined by only four equations, which describe the equilibrium, energy transport, conservation of mass, and conservation of energy within the star. The crucial equations, given in Focus 10.3, can be solved without any knowledge of the properties of the star before arrival on the main sequence. Kippenhahn et al. (2012) provide a good textbook of stellar structure and evolution.

Focus 10.3 The equations of stellar structure

To obtain information on the interior constitution of the stars, astrophysicists have to integrate basic equations. Pioneering work in this field can be found in the books of Eddington (1926a, b) and Chandrasekhar (1939). The four differential equations that determine a stars initial position on the main sequence of the Hertzsprung–Russell diagram and its subsequent evolutionary history are:

The equation of hydrostatic equilibrium. This equation states that the inward force of gravity caused by the mass, $M(r)$, within a distance, r, from the stellar center is just balanced by the outward gas pressure, $P(r)$, at radius, r, so that:

$$\frac{dP(r)}{dr} = -\rho(r)\frac{GM(r)}{r^2}, \tag{10.56}$$

or equivalently

$$\frac{dP(r)}{dM(r)} = -\frac{GM(r)}{4\pi r^4}, \tag{10.57}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the mass density is denoted by $\rho(r)$, and the gas pressure is given by the ideal gas law (Sect. 5.4).

The equation of mass continuity or *the equation of mass conservation*. This equation specifies the mass, $M(r)$, contained within radius, r, in terms of the mass density, $\rho(r)$, by:

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho(r) \tag{10.58}$$

or equivalently

$$\frac{dr}{dM(r)} = \frac{1}{4\pi r^2 \rho(r)}. \tag{10.59}$$

This equation is subject to the boundary conditions of zero mass at zero radius, or $M(r) = 0$ at $r = 0$, and a mass that is now equal to the total mass of the star, M_S, at the visible stellar radius, R_S, or $M(R_S) = M_S$. For the Sun, $M_S = M_\odot = 1.989 \times 10^{30}$ kg and the radius $R_S = R_\odot = 6.955 \times 10^8$ m.

The equation of energy conservation. This equation states that the energy generated per unit mass per unit time in the star's core, denoted by $\varepsilon(r)$, supplies the energy flux, $L(r)$, carried across radius, r, or that:

$$\frac{dL(r)}{dr} = 4\pi r^2 \rho(r) \varepsilon(r), \tag{10.60}$$

or equivalently

$$\frac{dL(r)}{dM(r)} = \varepsilon(r). \tag{10.61}$$

The energy generation, $\varepsilon(r)$, is a function of the initial composition, mass, density, and temperature. This equation has the boundary condition provided by the current luminosity, L_S, for a star of total mass, M_S, and radius, R_S. For the Sun, we have $L_S = L_\odot = 3.828 \times 10^{26}$ J s^{-1}.

The equation for radiative energy transfer. This equation relates the temperature, $T(r)$, at radius, r, to the amount of energy being transferred by radiation to that distance. It is related to the opacity to radiation, $\kappa(r)$, which measures the resistance of the material to energy transport by radiation. The equation is:

$$\frac{dT(r)}{dr} = -\frac{3\kappa(r)\rho(r)L(r)}{16\pi r^2 ac[T(r)]^3}, \tag{10.62}$$

or equivalently

$$\frac{dT(r)}{dM(r)} = -\frac{3\kappa(r)L(r)}{64\pi^2 acr^4[T(r)]^3}, \tag{10.63}$$

where the radiation density constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The total luminosity of a star, L_S, with a radius, R_S, is given by the Stefan-Boltzmann law $L_S = 4\pi\sigma R_S^2 T_{eff}^4$ where the Stefan-Boltzmann constant $\sigma = ac/4 = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4} and T_{eff} is the effective temperature of the visible stellar disk.

The chemical composition of a zero-age main-sequence star is assumed to be homogenous, and is often specified by $X = \rho_H/\rho$ the fraction by mass of material in the form of hydrogen, $Y = \rho_{He}/\rho$ the fraction by mass of material in the form of helium, and $Z = 1 - X - Y = \rho_{metals}/\rho$ the fraction by mass of material heavier than helium. The mass density for fully ionized plasma of total number density, N, is:

$$\rho = \left(2X + \frac{3}{4}Y + \frac{1}{2}Z\right)^{-1} Nm_H, \tag{10.64}$$

where the mass of the hydrogen atom is $m_H = 1.6726 \times 10^{-27}$ kg.

The approximate abundances observed in the disk of the Sun are $X = 0.71$, $Y = 0.27$, and $Z = 0.02$, and $\bar{m} = 0.61\, m_H$; in the solar core nuclear fusion reactions have converted about half the hydrogen into helium and $X = 0.34$, $Y = 0.64$ and $Z = 0.02$, and $\bar{m} = 0.85\, m_H$.

At any given time, stars of the same composition have radii, luminosities, effective temperatures, and mean densities determined solely by the star's mass. The German astronomer Heinrich Vogt (1875–1936) demonstrated this concept in 1926, and the American astronomer Henry Norris Russell (1877–1957) derived it independently the following year in his textbook; therefore, it is known as the Vogt–Russell theorem (Vogt 1926; Russell et al. 1927). It implies that a star of a given mass, age, and chemical composition occupies a unique position, related to the star's evolutionary history, on the H-R diagram. The mass, age, and composition are all we need to know to understand the life history of a star.

A star will continue shining with a luminosity and temperature determined by its mass, remaining stable and fundamentally unchanged for millions to billions of years. The only caveat to this understanding of stellar life is that the core of a star is the only place hot enough for nuclear reactions to occur. The composition of the core slowly changes as the result of these reactions; eventually, there is no more nuclear energy in the stellar core so it loses its equilibrium.

It appears to be simple, but the theory is complex – with detailed applications that are found in advanced texts. Moreover, our understanding of the internal constitution of stars includes explanations of how their energy is generated, by thermonuclear reactions in the stellar cores, beginning with the fusion of hydrogen into helium in main-sequence stars.

10.3.2 Two Ways to Burn Hydrogen in Main-Sequence Stars

All main-sequence stars generate energy by the thermonuclear fusion of hydrogen nuclei, the protons, into helium nuclei. Because the hydrogen is "burned up" or consumed to fuel the nuclear fires, we call this process *hydrogen burning*, although it is a chain of nuclear-fusion reactions rather than the combustion of an ordinary fire.

There are two methods of burning, or fusing, hydrogen into the heavier element helium; the dominant mechanism, which produces the most power, depends on the mass of a star. The main source of energy for main-sequence stars with a mass less than 1.5 times the Sun's mass is the *proton–proton chain* of nuclear reactions, abbreviated as the *p–p chain*. A different sequence of nuclear reactions converts protons into helium nuclei inside main-sequence stars more massive than 1.5 times the mass of the Sun. This is known as the *carbon–nitrogen–oxygen* (*CNO*) *cycle*. As the name suggests, this is a cyclic set of nuclear reactions.

The thermonuclear process responsible for the energy production in any star is not limited to a single nuclear transformation, but rather consists of a sequence of linked transformations that together form a nuclear chain reaction. The p–p chain is linear, with only one direction, whereas the CNO cycle occurs in a closed circular chain. For both types of hydrogen burning, four protons combine to make one helium nucleus, thereby releasing energy.

In the lower right-hand, low-mass side of the main sequence, where the great majority of stars are located, nuclear energy is generated as a result of the proton–proton chain that makes the Sun shine. In 1939, the German-born American physicist Hans A. Bethe (1906–2005) first delineated this complete nuclear transformation of four protons into one helium nucleus (Bethe 1939).

Because the proton–proton chain was discussed in detail in Sect. 8.3, we now limit the discussion to the general result. Four protons, each designated by ^{1}H or p, combine to form a helium nucleus, denoted by ^{4}He, releasing powerful gamma ray radiation, designated γ, positrons with the symbol e^+, electron neutrinos, denoted by the symbol ν_e, and an energy of about 4×10^{-12} J per reaction chain. We can use the notation of nuclear reactions to describe the p–p chain, with nuclei on the left side of a reaction designated by an arrow, $\rightarrow$, fusing to make the nuclei and other particles or radiation on the right side of the arrow. The proton–proton chain is described using this notation in Focus 10.4, with the net result for each reaction chain:

$$4\,^1\mathrm{H} \rightarrow {}^4\mathrm{He} + 6\gamma + 2\nu_e. \tag{10.65}$$

Focus 10.4 The proton–proton chain

The proton–proton chain of reactions begins with the merger of two protons, each designated by ^{1}H, in the reaction:

$$^1\mathrm{H} + {}^1\mathrm{H} \rightarrow {}^2\mathrm{D} + e^+ + \nu_e. \tag{10.66}$$

It is followed by

$$^{2}\mathrm{D} + {}^{1}\mathrm{H} \rightarrow {}^{3}\mathrm{He} + \gamma, \tag{10.67}$$

with a last step

$$^{3}\mathrm{He} + {}^{3}\mathrm{He} \rightarrow {}^{4}\mathrm{He} + {}^{1}\mathrm{H} + {}^{1}\mathrm{H}. \tag{10.68}$$

In the process, additional gamma rays are released when the positrons combine with electrons, denoted e^{-}, during pair annihilation, denoted by $e^{+} + e^{-} \rightarrow \gamma + \gamma$, and the net result of the p–p chain is:

$$4\,{}^{1}\mathrm{H} \rightarrow {}^{4}\mathrm{He} + 6\gamma + 2\nu_{e}. \tag{10.69}$$

At about the same time that Bethe delineated the proton–proton chain, the German physicist Carl Friedrich von Weizsäcker (1912–2007) examined the thermonuclear reactions that might occur in stars, publishing his results in two papers that appeared in the *Physikalische Zeitschrift* (Weizsäcker 1937, 1938). In the first paper, he reasoned that the merger of two protons must have started hydrogen burning in the Sun. In the second paper, Weizsäcker proposed that elements that are heavier than hydrogen already were created before the formation of stars, as we know them now. He no longer was limited to reactions that began with the lightest element, hydrogen, and this led him to the important discovery of the cyclic CNO chain of reactions in which carbon acts as a catalyst for the synthesis of helium from hydrogen (Focus 10.5). The overall result for each CNO reaction chain is as follows:

$$^{12}\mathrm{C} + 4\,{}^{1}\mathrm{H} \rightarrow {}^{12}\mathrm{C} + {}^{4}\mathrm{He} + 5\gamma + 2\nu_{e}, \tag{10.70}$$

where the carbon nucleus, denoted ^{12}C, is forever being regenerated and acts like a catalyst. That is, the CNO cycle is a circular reaction chain in which carbon is destroyed and then re-created. Therefore it is available for the sequence of reactions to occur repeatedly. Like the non-circular, linear proton–proton chain, each CNO reaction chain releases about 4×10^{-12} J of energy.

Focus 10.5 The CNO cycle

The cyclic CNO chain of reactions starts when a carbon nucleus, ^{12}C, fuses with a proton, ^{1}H, to produce a nucleus of nitrogen, ^{13}N, and gamma ray radiation, γ. The nitrogen decays to form a nucleus of heavier carbon, ^{13}C; a positron, e^{+}; and an electron neutrino, ν_{e}. These beginning nuclear-fusion reactions are as follows:

$$^{12}\mathrm{C} + {}^{1}\mathrm{H} \rightarrow {}^{13}\mathrm{N} + \gamma, \tag{10.71}$$

and

$$^{13}\mathrm{N} \rightarrow {}^{13}\mathrm{C} + e^{+} + \nu_e. \tag{10.72}$$

The cycle then continues when the heavy carbon combines with a proton to form heavier nitrogen, ^{14}N, which then fuses with another proton to form oxygen, ^{15}O. The oxygen decays to make ^{15}N, which then combines with a proton to make the original carbon, ^{12}C, together with a helium nucleus, ^{4}He. The nuclear-fusion reactions are as follows:

$$^{13}\mathrm{C} + {}^{1}\mathrm{H} \rightarrow {}^{14}\mathrm{N} + \gamma \tag{10.73}$$

$$^{14}\mathrm{N} + {}^{1}\mathrm{H} \rightarrow {}^{15}\mathrm{O} + \gamma \tag{10.74}$$

$$^{15}\mathrm{O} \rightarrow {}^{15}\mathrm{N} + e^{+} + \nu_e \tag{10.75}$$

and

$$^{15}\mathrm{N} + {}^{1}\mathrm{H} \rightarrow {}^{12}\mathrm{C} + {}^{4}\mathrm{He}. \tag{10.76}$$

The positrons, e^{+}, annihilate with the electrons, e^{-}, to produce energetic gamma radiation by the following reaction:

$$e^{+} + e^{-} \rightarrow 2\gamma. \tag{10.77}$$

This occurs for both positrons generated during the each CNO reaction chain.

Like the proton–proton chain, the net result of the CNO cycle is that four protons are fused together to form one helium nucleus, gamma rays and electron neutrinos. By summing the left and right sides of all the participating reactions, we obtain the following:

$$^{12}\mathrm{C} + 4\,{}^{1}\mathrm{H} \rightarrow {}^{12}\mathrm{C} + {}^{4}\mathrm{He} + 5\gamma + 2\nu_e. \tag{10.78}$$

The circular CNO reaction chain is induced by high temperatures in the cores of massive main-sequence stars and becomes self-sustaining by the catalytic action of carbon. This cycle also could begin at the intermediate stages with nitrogen or oxygen, so the entire cycle is called the carbon–nitrogen–oxygen, or CNO, cycle.

It was Bethe who realized that the proton–proton reaction, which explains the luminous output of the Sun, fell short of the much greater luminosity of the hotter and more massive stars. So, he systematically examined a great number of nuclear reactions that would not operate within stars and eliminated them. He independently found that the CNO cycle would generate about the same energy as the proton–proton process for each nuclear reaction chain. Bethe also showed that the greater rate and temperature dependence of the CNO cycle could account for the

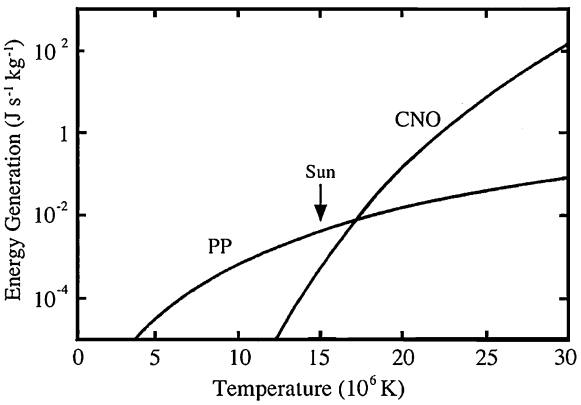


Fig. 10.11 Energy generation by two hydrogen-burning processes The energy output (*left vertical axis*), in units of power per kilogram, or J s^{-1} kg^{-1}, as a function of core temperature (*bottom horizontal axis*) in millions, or 10^6, degrees kelvin, designated K. The proton–proton chain, denoted PP, dominates the hydrogen-burning energy production for the Sun and less massive stars that have lower core temperatures. At the center of the Sun, where the temperature is 15.6×10^6 K, the PP chain is the dominant nuclear-reaction chain for converting hydrogen nuclei into helium nuclei, with an energy output of 0.016 J s^{-1} kg^{-1}, or 51 million MeV g^{-1} s^{-1} in the units used by nuclear astrophysicists. In more massive main-sequence stars, the central temperature is higher and the CNO cycle of hydrogen burning is the most efficient process. Main-sequence stars of mass less than 1.5 solar masses shine by the PP chain of nuclear reactions, whereas the main-sequence stars with mass greater that 1.5 solar masses burn hydrogen by the CNO set of nuclear reactions. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

high luminosity of the massive stars. Subsequently, this conclusion was placed on a firm basis when William A. "Willy" Fowler (1911–1995) and his colleagues at the California Institute of Technology made laboratory measurements of the nuclear reaction cross sections for every reaction in the chain; the results were published in a series of papers spanning several decades.

Given today's knowledge, it is the mass and therefore the central temperature of a main-sequence star that determine which hydrogen burning reaction supplies most of its power. The relevant formula for the energy production rate, ε_{pp}, of the proton–proton chain is given by

$$\varepsilon_{pp} = 0.24\rho X^2 \left[\frac{10^6}{T}\right]^{2/3} \exp\left[-33.8\left(\frac{10^6}{T}\right)^{1/3}\right] \text{J s}^{-1}\ \text{kg}^{-1}, \tag{10.79}$$

where ρ is the mass density and X is the mass fraction of hydrogen. The energy production rate for the CNO cycle is:

$$\varepsilon_{CNO} = 8.7 \times 10^{20} \rho X_{CNO} X \left(\frac{10^6}{T}\right)^{2/3} \exp\left[-152.3\left(\frac{10^6}{T}\right)^{1/3}\right] \text{J s}^{-1}\ \text{kg}^{-1}, \tag{10.80}$$

where X_{CNO} is the sum of the mass fractions for carbon, nitrogen, and oxygen. The variations of the two energy-producing reactions as a function of the core temperature T is illustrated in Fig. 10.11 for a typical stellar composition.

Both the proton–proton chain and the CNO cycle operate within main-sequence stars, and release about the same amount of energy during each reaction chain; however, the total amount of energy generation differs depending on the mass and central temperature of a star (Fig. 10.11). The CNO cycle contributes little energy in low-mass stars with low central temperatures, and the proton–proton chain produces almost all of the energy radiated by main-sequence stars with a mass less than or equal to the Sun's mass (Bahcall et al. 2003). However, the CNO cycle is faster and generates more energy in massive stars that have high central temperatures. Stars with a mass of about 2 solar masses or above generate almost their entire energy output by the CNO cycle.

In the Sun, with a central temperature of 15.6 million K, only 1.5 % of its energy is generated by the CNO cycle. That is why the proton–proton chain explains the Sun so well. However, with increasingly more mass and an increasingly hotter stellar core, the CNO cycle becomes the dominant energy source for a main-sequence star. For a star with a mass of 1.5 times the mass of the Sun, where the central temperature reaches 18 million K, each method of burning hydrogen produces the same amount of total energy and half the luminosity of a star.

The role of radiation and convection in transporting energy out of the stellar core also depends on the mass of a main-sequence star (Fig. 10.12). When the stellar mass is comparable to that of the Sun, the energy-generating core is surrounded by a radiative zone and topped by a convective zone (see Sect. 8.5). Main-sequence stars with a mass of more than 2 solar masses have a convective core. In these stars, the rate of energy generation by the CNO cycle is sensitive to temperature, so the fusion is highly concentrated in the core. Consequently, there is a high-temperature gradient in the core region, which results in a central convection zone. The outer regions of such a massive star transport energy by radiation with little or no convection.

All stars that begin their relatively long and placid life on the main sequence are composed mainly of hydrogen. These stars are initially uniform balls of plasma with the same composition throughout. As time passes, the central stellar core is changed slowly from hydrogen to helium, so the inside of a main-sequence star eventually becomes different from the outside. Eventually, the core is all used up, exhausting its supply of hydrogen by converting it into helium. The star has to leave the main sequence and become hot enough inside to burn helium – the ash of its former hydrogen-burning fires.

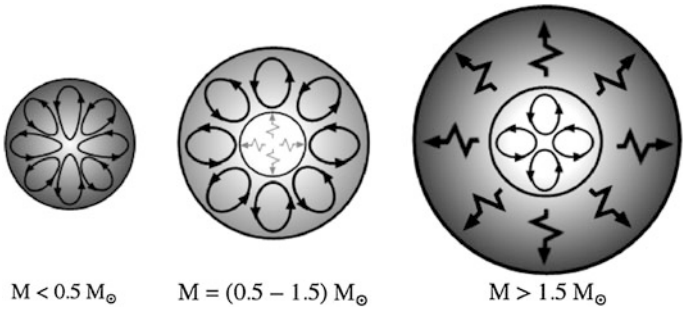


Fig. 10.12 Convection inside stars of different mass Most stars have convective zones in which energy is transported by the wheeling motion of convection, denoted here by closed curves with arrows for stars of different mass, designated by M, and compared to the Sun's mass denoted $M_{\odot}$. The symbol $<$ means less than and the symbol $>$ denotes greater than. Low-mass stars, with less than half a solar mass, are fully convective from core to visible disk and therefore of uniform composition. Their low temperatures result in a high opacity to radiation. In intermediate-mass stars, such as the Sun, radiation transport dominates convection in the hot central regions, which are enveloped by a cooler convective region. The visible disks of these stars do not include the nuclear-fusion products from their core but rather retain the same composition as the interstellar medium from which these stars were formed. High-mass stars, with more than 1.5 times the mass of the Sun, have a large radiative zone that is not enveloped by a convective zone. The temperature-sensitive hydrogen-burning reactions of the CNO cycle cause the development of a convective core in these stars. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

10.3.3 Helium Burning in Giant Stars

The fact that giant stars are connected to the main sequence of the H-R diagram suggested that the giants are the next stage of stellar evolution. However, because giant stars have larger luminosities at lower disk temperatures than main-sequence stars, they seemed to shine by a different and unknown process. The enigma was resolved partially when the Estonian astronomer Ernst Öpik (1893–1985), working at the Armagh Observatory in Northern Ireland, argued that the inside of a giant star can become very hot and dense at the same time that its outer parts become cool and rarefied (Öpik 1938).

The hydrogen-burning process of a main-sequence star is confined to the central stellar core, which is surrounded by an inert, nonburning envelope in which no nuclear reactions take place. When the core hydrogen is expended, the core is forced to contract, for it can no longer support itself under the crush of gravity. The central temperature will rise to about 100 million, or 10^8, K once gravitational forces compress the core to a smaller volume and increase the mass density 1,000 fold, to about 10 million kg m^{-3}. The rapid increase in core temperature causes the surrounding hydrogen envelope to expand, producing a vast, cool envelope of low mass density (Fig. 10.13). These spectacular changes in both the inside and

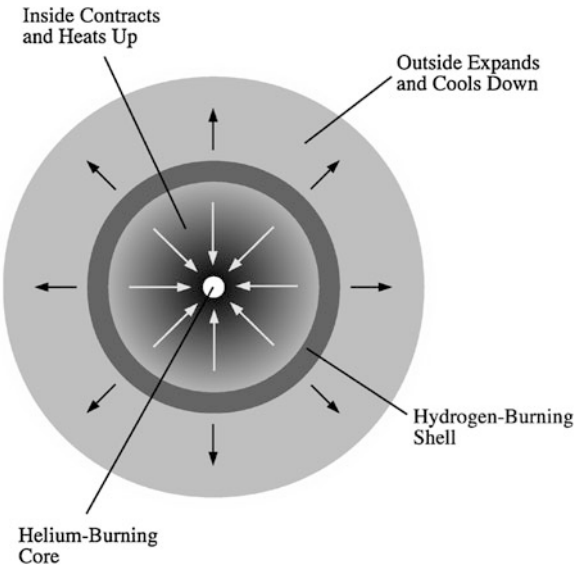


Fig. 10.13 Formation of a giant star When a main-sequence star consumes the hydrogen in its core, the inside of the star contracts and heats up, causing the outside to expand and cool down. Hydrogen burning resumes in a shell that envelops the collapsing core. The center of the star eventually heats up to about 100 million K, which is hot enough to burn helium and stop the core collapse. A giant star then has been created with a luminosity of about 100 times that of the Sun and a radius of approximately 50 times the radius of the Sun. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

outside of a dying main-sequence star account for the observed characteristics of red giant stars.

Öpik realized that the large increase in central temperatures and densities of giant stars open up a new source of energy not available to the main-sequence stars, and he proposed that the helium ash produced by hydrogen burning would serve as the nuclear fuel for giant stars.

The main difficulty with this scenario was that there is no stable nucleus of atomic weight 5, and this gap seemed to provide an impenetrable barrier for the synthesis of heavier elements from helium of weight 4 using protons of weight 1. A proton could not be attached to a helium nucleus to make the next heavier substance.

It took more than a decade to resolve the difficulty, which was explained almost simultaneously and independently by Öpik and the American astronomer Edwin E. Salpeter (1924–2008), at Cornell University (Öpik 1951; Salpeter 1952). When the core of a star reaches a sufficiently high temperature, of about 100 million, or 10^8, K, helium nuclei can be converted to carbon nuclei by triple collisions of helium nuclei, thus circumventing the mass 5 difficulties.

This release of energy by fusing helium into carbon within a star is known as *helium burning*. It also is called the *triple alpha process* because the helium nucleus is an alpha particle and a triple collision is required to make a nucleus of the carbon atom.

The net result of the helium burning, triple alpha chain reaction is

$$^4\text{He} + {}^4\text{He} + {}^4\text{He} \rightarrow {}^{12}\text{C} + \gamma, \tag{10.81}$$

where ^{4}He is a helium nucleus or alpha particle, ^{12}C is a carbon nucleus, the γ denotes gamma radiation. About 7.275 MeV or 1.165×10^{-12} J of energy is released each time it occurs. The triple alpha process happens so frequently at the high central temperatures of giant stars that it can power their intense luminosity.

The difficulty in pushing together even two helium nuclei is exacerbated by their electrical charge. Each helium nucleus contains two protons; therefore, the electrical repulsion between two helium nuclei is four times that between two protons.

Example: Hot enough to burn helium

To overcome the electrical, or Coulomb, repulsion between two helium ions, of charge $2e = 2 \times 1.6022 \times 10^{-19}$ C each, the kinetic energy $m_{He}V^2/2$ of a colliding helium ion, of mass $m_{He} = 6.6445 \times 10^{-27}$ kg and velocity V must be equal to the electrical potential energy $4e^2/(4\pi\varepsilon_0 R_{He})$ when the two ions touch, where $\pi = 3.14159$, the permittivity of free space is $\varepsilon_0 = 10^{-9}/(36\pi) = 8.854 \times 10^{-12}$ F m^{-1}, and the radius of the helium ion is $R_{He} \approx 10^{-15}$ m. Solving for the velocity $V \approx 10^7$ m s^{-1}, and setting it equal to the thermal velocity $V = V_{thermal} = (3kT/m_{He})^{1/2}$, with the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, we obtain a temperature of $T \approx 1.6 \times 10^{10}$ K to overcome the electrical repulsion.

This is an impossibly high temperature, about 4 times hotter than that required for the direct fusion of two protons, but quantum–mechanical tunneling lowers the central star temperature needed to about five billion, or $T_C \approx 5 \times 10^9$ K. The increase in temperature is created by core collapse to a radius R_c. By equating the thermal energy $3kT_c/2$ to the gravitational energy $Gm_{He}M_c/R_c$, where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and assuming a core mass about equal to that of the Sun with $M_\odot = 1.989 \times 10^{30}$ kg, we can infer a core radius of $R_c \approx 10^7$ m $\approx 0.014\ R_\odot$ for the Sun's radius $R_\odot = 6.955 \times 10^8$ m and a mass density of $\rho = 3\ M_\odot/(4\pi R_c^3) \approx 0.5 \times 10^9$ kg m^{-3}.

It turns out that certain nuclear resonances also enhance the reaction rate of the triple helium burning process, and modern calculations indicate that a core temperature of 100 million, or 10^8 K might suffice to initiate the helium fusion, but we have the basic idea. The central core must collapse to increase the temperature and enable helium fusion.

The quantum–mechanical tunneling effect will help, as it does in permitting two protons to fuse together in the Sun (see Sect. 8.3). The rise in the central temperature of a giant star is needed to increase the number of helium nuclei moving fast enough to penetrate the larger electrical barrier with the aid of tunneling, which incidentally also explains why helium burning doesn't occur in low-temperature, main-sequence stars. Only the helium nuclei in the high-velocity tail of the Maxwellian speed distribution can merge together in a giant star, which means that most of the helium nuclei are not moving fast enough to merge and that the helium-burning reactions occur relatively slowly.

Under most circumstances, helium burning still would be exceedingly unlikely because it involves the nearly simultaneous collision of not two but rather three helium nuclei. Such a triple collision is favored by two exceptionally large collision cross sections, termed *resonance reactions* in the parlance of nuclear physics. The bigger the cross section for a collision, the more likely it will occur.

In 1952, Salpeter, who was unaware of Öpik's work the previous year, presented some of these more detailed explanations for the formation of carbon in the triple alpha, helium burning process. Two helium nuclei can combine to form a beryllium 8 nucleus, but because the beryllium is unstable only a tiny fraction remains at any instant. As Salpeter noticed, a beryllium nucleus can occasionally combine with a third helium nucleus to form carbon 12. Nevertheless because beryllium is so extremely rare, it must have a large cross section for helium capture if any substantial amount of carbon is to be produced.

Two years later, Fred Hoyle (1915–2001) showed that the triple alpha process occurs at a rapid enough rate to provide the luminosity of a red giant star if the carbon goes through an excited state and the giant core temperature reaches about 10^8 K at a mass density of about 10^7 kg per cubic meter (Hoyle 1954). William Fowler and his colleagues subsequently showed that the required excited state of carbon does in fact exist. The formation of carbon from helium is thus enhanced enormously by two facts: the existence of beryllium 8 (itself a kind of resonance) and the existence of the excited state of carbon.

The details are somewhat complicated but suffice it to say that the mass 5 barrier can be overcome; Fynbo et al. (2004) provide relatively recent estimates for the rates of the stellar triple-alpha process. In retrospect, it seems almost miraculous that nature has conspired in this way to make helium burning and giant stars exist. Our understanding of the evolution of main-sequence stars into these larger, more luminous counterparts was stimulated by investigations of star clusters.

10.4 Using Star Clusters to Watch How Stars Evolve

Entire stars have had a beginning, followed by a long period of growth and inevitable decay, eventually turning into something else, and we can use the H-R diagrams of star clusters to map out the stages of stellar transfiguration. As time elapses, the more massive stars evolve into the next phase of stellar life and the

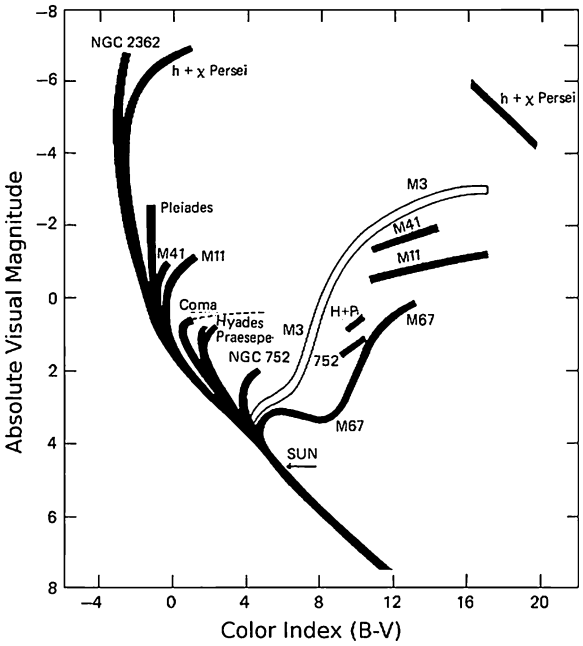


Fig. 10.14 Open star clusters in the H–R diagram Open star clusters are relatively young, and most of their stars have not yet left the main sequence in the H-R diagram. The youngest clusters, such as the Pleiades, retain all but the topmost part of the main sequence. The turnoff point of the Pleiades star cluster from the main sequence indicates an age of roughly 100 million years. The Hyades star cluster, which turns off about halfway down the main sequence, is about 600 million years old. The lowest open cluster in this diagram, M 67, is an estimated 5 billion years old, with a main sequence that stops just above the Sun. One globular star cluster, M 3, is shown for comparison. [Adapted from Allan Sandage (1957).]

main sequence disappears from the top down. Very massive stars at the upper left of the main sequence become supergiants; those with intermediate masses comparable to the Sun become red giants.

Benacquista (2013) gives a fine introduction to the evolution of both single and binary stars. Gallart et al. (2005) have provided a review of stellar evolution models and color magnitude diagrams. Iben and Renzini (1983) have reviewed the asymptotic giant branch (AGB) evolution and beyond, and Winckel (2003) has reviewed the post AGB stars. Chiosi and Maeder (1986) have reviewed the evolution of massive stars with mass loss; Gilles and Baraffe (2000) have reviewed the theory of low-mass stars and substellar objects; and Maeder and Meynet (2000) have discussed the evolution of rotating stars.

Stars within a star cluster are all of the same approximate age, within a few million years, dating back to the formation of the cluster. They also began with the same initial composition of material and exhibit a full range of stellar mass. Because the stars in a given cluster are all at the same distance from the Earth, we

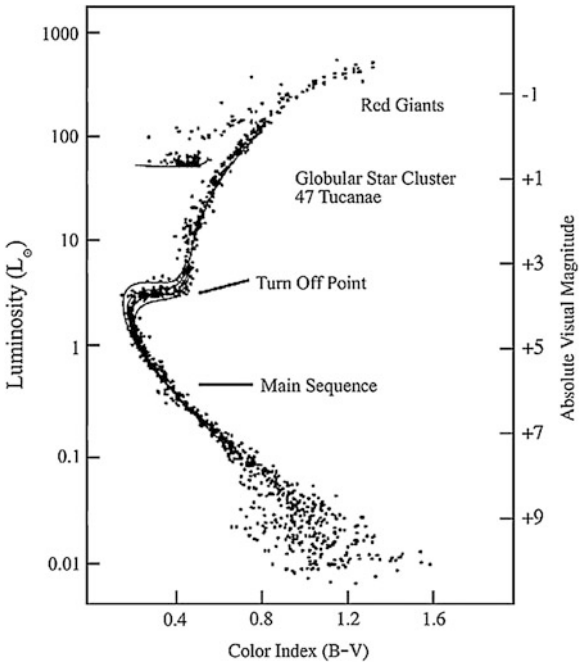


Fig. 10.15 How old is a globular star cluster? A plot of the luminosity (*left vertical axis*), in units of the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, against the color index, B-V (*bottom horizontal axis*), for the stars in the southern globular star cluster 47 Tucanae, also designated NGC 104. The absolute visual magnitudes, M_V, of the stars also are shown (*right vertical axis*). Although low-mass, relatively faint stars are still on the main sequence (*diagonal line from middle left to bottom right*), the massive, bright stars in the cluster have left the main sequence and are evolving into giant stars (*top right*). Theoretical tracks, called *isochrones*, show the evolutionary distributions at different ages of 10 billion, 12 billion, 14 billion and 16 billion years from top to bottom and left to right. The best fit to the observed data corresponds to an age between 12 billion and 14 billion years for this star cluster. (Courtesy of James E. Hesser.)

can obtain direct observations of their relative luminosity without knowing the distance.

A cluster H-R diagram can be used as a clock, dating the age of the cluster and the stars in it by the place of their turnoff from the main sequence to become supergiants or giants. Stars with a luminosity and temperature greater than the turnoff value all have evolved away from the hydrogen-burning state of stellar life, and the age of the cluster is equal to the main-sequence lifetime of stars at this turnoff point. The lower its luminosity or temperature, the older is the star cluster.

The main-sequence turnoffs of the loosely bound open star clusters (Fig. 10.14) indicate that they are approximately 100 million years old (Mermilliod 1981; Sandage 1957). Such relatively young clusters are identified by the membership of O and B stars, which would leave the main sequence in a relatively short time.

The main sequence of the H-R diagram for globular star clusters does not contain these hot, luminous stars (Fig. 10.15), and these star clusters have ages between 10 billion and 14 billion years – or about two to three times the age of the Sun. This indicates that the masses of the stars at the main-sequence turnoff point of globular clusters are less than the Sun. Soderblom (2010) provided a review of the age of stars, whereas Van den Berg et al. (1996) reviewed the age of the galactic globular cluster system.

Due to their great age and numerous stars, the H-R diagrams of these dense stellar concentrations help us watch how stars evolve to the later stages of stellar life. Such investigations involve theoretical calculations of precisely how long a main-sequence star's central fuel supply can last and models of what happens when its fuel is used up. Martin Schwarzschild (1912–1997), the son of German astronomer Karl Schwarzschild (1873–1916), was one of the first to examine this phase of stellar evolution. After emigrating to the United States, Martin Schwarzschild used theoretical models and primitive computers, developed by his Princeton colleague John von Neumann (1903–1957), to chart the evolutionary trajectory of a star and compare it to the various kinks, bends, and gaps of missing stars on the H-R diagrams of globular star clusters. His theoretical evolutionary models were facilitated by the fact that all of the stars in globular star clusters have the same initial chemical composition.

Martin teamed with Allan Sandage (1926–2010), then a graduate student at the California Institute of Technology, whose H-R diagrams included faint stars that connected the main sequence to the red-giant branch (Fig. 10.16). When Sandage's data were compared to the model results, it improved our understanding of evolution away from the main sequence and this provided a sound observational basis for stellar aging (Sandage and Schwarzschild 1952; Sandage 1957; Schwarzschild 1958).

When the hydrogen-burning fires are quenched in a star of roughly solar mass, the shrinking stellar core heats up and causes the star as a whole to swell into a bloated red giant. The gravitational energy released by the collapsing, nonburning core is then spread over a much larger area, resulting in a lower disk temperature and a shift of the visible starlight into the red part of the spectrum. This accounts for the red-giant branch in the H-R diagram of globular star clusters.

The rising heat from the collapsing core ignites hydrogen burning in an internal shell that envelops a core of inactive helium. As realized by Schwarzschild and the English astronomer Fred Hoyle (1915–2001), the giant core is compressed into a degenerate state and eventually heats up to a temperature of about 10^8 K, when core helium burning begins (Hoyle and Schwarzschild 1955). Once the star fuses helium into carbon within its core, it enters the horizontal branch of the cluster H-R diagram and is technically no longer considered a red giant. As then demonstrated by Schwarzschild and the Richard Härm (1909–1996), helium begins to

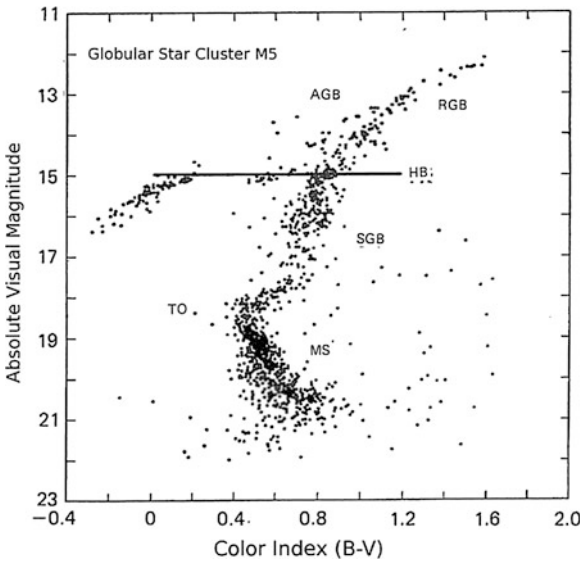


Fig. 10.16 Globular star cluster in the H–R diagram The Hertzsprung-Russell diagram for the globular star cluster M 5, where the absolute visual magnitude (*left vertical axis*) is plotted as a function of color index (*bottom horizontal axis*). It is very different from the H-R diagrams for open star clusters shown in Fig. 10.14. The high-mass stars in this globular star cluster have left the main sequence (*MS*) at a relatively low turnoff point, denoted TO, indicating a greater age than the open star clusters. This diagram illustrates the evolutionary tracks of these stars into the red-giant branch, designated RGB (*top right*), as well as other evolutionary stages such as the subgiant branch, denoted SGB; the asymptotic giant branch, designated AGB; and the horizontal branch, denoted HB, that extends to the left. The gap of missing stars in the horizontal branch for the globular star cluster M 5 shows the instability strip of pulsating stars, known as RR Lyrae stars. [Adapted from Arp (1962).]

burn abruptly in the core, releasing a flash of intense energy that rejuvenates a star's luminous output and initiates core helium fusion into carbon, which can last about 10^8 years (Schwarzschild and Härm 1962). Härm was born and educated in Estonia, but moved to the United States during World War II (1939–1945), spending his entire subsequent career at Princeton University.

When the core helium is exhausted, the carbon and oxygen core collapses, helium is burned in a surrounding shell, and the star briefly rises to giant status a second time, along the asymptotic branch of the H-R diagram. A short period of instability then begins, where the star can pulsate, but it is approaching its end.

An even briefer stellar life, with more violent winds, is expected for the more massive main-sequence stars that become supergiants. They help explain where most of the heavier elements originate.

10.5 Where did the Chemical Elements Come From?

10.5.1 Advanced Nuclear Burning Stages in Massive Supergiant Stars

A supergiant star has a much greater mass, interior compression, central temperature, and luminosity than a giant star, which is why we call them "super." Supergiants pass through the same early stages of stellar life as giants but at a faster rate. Unlike their counterparts of lesser mass, massive stars with a mass above 10 solar masses quickly consume the hydrogen in their cores and transform into colossal supergiants that burn the next available nuclear fuel: helium. A star with a mass of 25 times that of the Sun, for example, will complete hydrogen burning by the CNO cycle in about 7 million years and helium burning in a mere 660,000 years. In contrast, the Sun will take roughly 10 billion years to complete hydrogen burning by the proton–proton chain and another 100 million years to consume the helium within its core.

When the core helium is consumed, the evolutionary paths of high-mass and moderate-mass stars diverge. The cores of supergiant stars are so massive that they can contract and heat up enough to burn carbon, thereby stopping the core collapse and shining with renewed vigor. Carbon nuclei have an electrical charge equivalent to six protons; therefore, a formidable electrical repulsion separates them, and collisions at great speed are required for its penetration. This can happen when the temperature rises to about a billion K, or 10^9 K.

In contrast, the giant stars are not sufficiently massive to burn anything heavier than helium because they never get hot enough inside. In technical terms, degeneracy pressure halts the contraction of the inert, nonburning carbon core in a giant star before it can become hot enough for carbon fusion. These stars expel their outer layers into surrounding space and collapse inside to an Earth-sized white dwarf star.

A supergiant is so massive that it can enter progressively more advanced nuclear-burning stages (Salpeter 1957; Weaver et al. 1978; Heger et al. 2003). The core helium is converted into carbon, and some of the newly formed carbon nuclei can fuse with helium nuclei to make oxygen. When the helium is gone, the core contracts until it becomes hot enough to burn carbon into neon, which temporarily re-stabilizes the core. Each time the core depletes the elements that it is fusing, or "burning," it shrinks and heats up until it becomes hot enough for fusion reactions of the nuclear ash, continuing up the chain of successively heavier abundant elements. At the same time, nuclear fusion of the earlier fuel continues in overlapping shells at lower temperatures. Layer upon layer of nuclear burning shells are created deep down inside a supergiant.

The aging of a supergiant star accelerates rapidly, consuming its internal fuel sources at ever-increasing central temperatures and rates (Table 10.8). Due to the higher temperatures needed for these nuclear reactions to occur, they also proceed at a much more rapid rate than hydrogen burning, and the thermonuclear lifetime

Table 10.8 Nuclear fusion processes in a supergiant star of 25 solar masses[a]

Core fusion process	Central temperature (K)	Central density (kg m^{-3})	Duration (years)
Hydrogen burning (H → He)	3.7×10^7	3.8×10^3	7,300,000
Helium burning (He → C and O)	1.8×10^8	6.2×10^5	660,000
Carbon burning (C → Ne)	7.2×10^8	6.4×10^8	165
Neon burning (Ne → Mg and Si)	1.4×10^9	3.7×10^9	1.2
Oxygen burning (O → Si)	1.8×10^9	1.3×10^{10}	0.5
Silicon burning (Si → Fe)	3.4×10^9	1.1×10^{11}	0.004

[a] Adapted from Weaver et al. (1978)

of the supergiant stars therefore is much shorter than those of even the giant stars. This is one reason that so few supergiant stars are observed.

In the terminal stages of a supergiant's life, its inner core is converting silicon into iron, and onion-like overlying layers are burning lighter elements, such as magnesium, oxygen, neon, carbon, and helium. However, when the iron nuclei in the core of such a star are pushed together, no energy is released. The iron does not burn, regardless of how hot the star's core becomes. The nuclear fires are extinguished; and the star has reached the end of its life. It can never again shine by any slow nuclear-fusion process, and there is no energy left to support the core. The star has become bankrupt, having completely spent all of its internal resources. There is nothing left to do but collapse.

The inert core collapses in less than 1 s, bounces and then explodes as a supernova with the light of 1 billion Suns. The shattered star and all of the elements made inside it are then dispersed into surrounding space. This material provides the seeds for future planets and stars, which explains where most of the heavy elements that are now found in the Earth and the Sun came from.

10.5.2 Origin of the Material World

What accounts for the origin of the chemical elements that make up our everyday world? The English astronomer Arthur Stanley Eddington (1882–1944) was one of the first to propose that the light elements are compounded into more complex elements within stars (Eddington 1920). Then, within a decade, the great stellar abundance of hydrogen had been established, and Robert d'Escourt Atkinson (1898–1982) showed how element synthesis in stars, starting with the proton–proton reaction, might account for both stellar energy and the origin of the elements (Atkinson 1931). In his view, the observed relative abundance of the elements could be explained by the creation of less abundant, heavy nuclei from

more abundant, lighter nuclei, particularly hydrogen and helium. The formation of heavy nuclei from the nuclear reactions of lighter nuclei is termed *nucleosynthesis*.

Baron Carl Friedrich von Weizsäcker (1912–2007) proposed two mechanisms for nucleosynthesis: either within stars or in the primeval "fireball" explosion, out of which the expanding universe arose (Weizsäcker 1937, 1938). Today, we know that our material world indeed was synthesized in both places. Most of the heavy elements that now are found in the universe were created within former stars by various nuclear reactions at different times and under different physical conditions. However, all of the hydrogen and most of the helium now observed in the universe was synthesized in the big-bang "fireball" explosion. Alpher and Herman (1950), Trimble (1975), Penzias (1979), and Fowler (1984) have reviewed the early history of ideas concerning element formation in stars and the big bang. Our understanding of the details of these processes is intimately related to the observed element abundance in the Sun.

10.5.3 The Observed Abundance of the Elements

An important key to understanding how stars synthesize the elements is obtained from their relative abundances, initially studied by chemists rather than astronomers. The American chemist William D. Harkins (1873–1951), for example, found an important clue to the mystery of the origin of the elements when he noticed that elements of low atomic weight are more abundant than those of high atomic weight and that, on average, the elements with even atomic numbers are about 10 times more abundant than those with odd atomic numbers of about the same number. These features led Harkins to conjecture that the relative abundances of the elements depend on nuclear rather than chemical properties and that heavy elements must have been synthesized from lighter ones (Harkins 1917, 1931).

Decades later, astronomers showed that stars are the crucibles in which all but the lightest elements are formed by nuclear reactions that convert light elements into heavier ones, and that the systematic decline in the abundance of heavier elements can be attributed to the relative scarcity of stars that have evolved to the stage that creates them.

Two other American chemists, Hans E. Suess (1909–1993) and Harold Clayton Urey (1893–1981), provided a detailed discussion of the elemental and isotopic abundances of the Sun and similar stars, calling attention to the many fluctuations that appear in the general trend of an exponential decline of abundance with increasing atomic weight (Suess and Urey 1956). This discussion served as a major stimulus for modern ideas concerning stellar nucleosynthesis (Fig. 10.17).

Asplund et al. (2009) have reviewed observations of the chemical composition of the Sun; Wilson and Rood (1994) provided an early summary of abundances in the interstellar medium; and Savage and Sembach (1996) reviewed interstellar abundances from absorption-line observations with the *Hubble Space Telescope*.

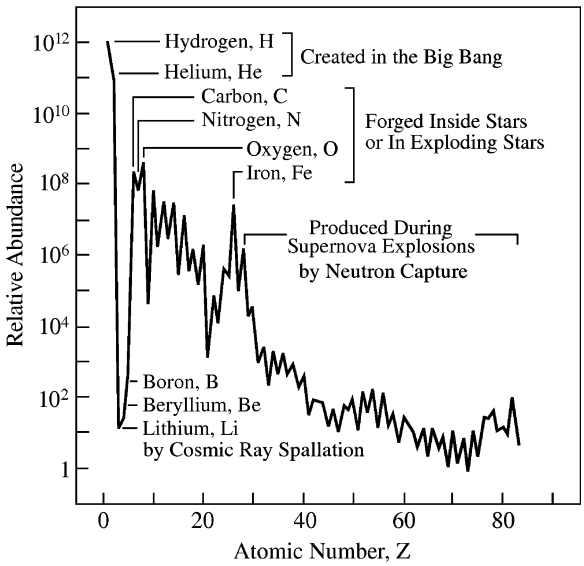


Fig. 10.17 Abundance and origin of the elements in the Sun The relative abundance of elements in the solar photosphere, plotted as a function of their atomic number Z, which is the number of protons in an atom's nucleus and roughly half the atomic weight. The abundance is plotted on a logarithmic scale and normalized to a value of 1 million million, or 1.0×10^{12}, for hydrogen. Hydrogen, the lightest and most abundant element in the Sun, was formed about 14 billion years ago in the immediate aftermath of the big bang that led to the expanding universe. Most of the helium now in the Sun was also created then. All of the elements heavier than helium were synthesized in the interiors of massive stars that were then wafted or blasted into interstellar space where the Sun subsequently originated. Carbon, nitrogen, oxygen, and iron were created over long time intervals during successive nuclear burning stages in former massive stars as well as during their explosive death. Elements heavier than iron were produced by neutron-capture reactions during the supernova explosions of stars that lived and died before the Sun was born. The light elements boron, beryllium and most lithium are believed to originate from heavier cosmic-ray particles that were stripped of some of their ingredients by collisions, in a process called *spallation*. The exponential decline of abundance with increasing atomic number and weight can be explained by the rarity of stars that evolved to later stages of life. (Data courtesy of Nicolas Grevesse.)

10.5.4 Synthesis of the Elements Inside Stars

The English astronomer Fred Hoyle (1915–2001) introduced the grand concept of nucleosynthesis in stars in the mid twentieth century (Hoyle 1946, 1954). He showed that both theoretical and experimental considerations were required, and placed the concept of stellar nuclear reactions within the framework of stellar structure and evolution, using the then-known nuclear data.

By 1957 the detailed abundance data of Suess and Urey had served as an inspiration for a comprehensive review of the major element-producing reactions in stars by Geoffrey R. Burbidge (1925–2010), E. Margaret Burbidge (1919–),

Fred Hoyle and William A. "Willy" Fowler (1911–1995). Their seminal publication, entitled "Synthesis of the Elements in Stars" was published in the *Reviews of Modern Physics* (Burbidge et al. 1957); it is known as the B^2FH paper, after the surnames of the four authors. It provided the fundamental framework on which most subsequent studies of stellar nucleosynthesis are based (Fig. 10.17).

As a star ages, it experiences successive core-contraction and stable-core burning stages, in which heavier elements are synthesized from lighter ones, and the nuclear ash of one stage becomes the fuel of the next one. These core nuclear-burning reactions proceed at a progressively hotter, denser, and faster pace, such as successive hydrogen, helium, carbon, neon, oxygen, and silicon burning. Elements with an even number of protons in their nuclei, such as carbon and oxygen, are more abundant because they are formed by nuclear reactions with helium, which contains two protons rather than one in its nucleus.

As realized by B^2FH, the temperature is not the same everywhere inside a star; therefore, its nuclear evolution is most advanced in the central regions and least or not at all advanced near its outer regions. Thus, the composition of an aging star is not uniform throughout. At the end of its life, a massive star would be layered with successively thinner shells of helium, carbon, oxygen, and silicon surrounding an inert, nonburning iron core that explodes as a supernova, casting out the heavy elements from the overlying layers of the star.

Although the general flow of the observed relative abundance of the more abundant elements can be explained by successive static, or nonexplosive, burning stages within stars, elements heavier than iron – as well as many of the detailed ups and downs of the abundance curve for lighter elements – were attributed to fast nuclear reactions during the explosive, supernova death of massive stars. The American hydrogen-bomb tests in the 1950s had indicated that heavy elements are created by rapid neutron bombardment during the explosions, and Hoyle and Fowler knew that similar processes occur in supernovae (Hoyle and Fowler 1960).

B^2FH used slow, s, neutron capture; rapid, r, neutron capture; and proton, p, capture to explain some of the details of the abundance curve. Some of the nuclear-burning reactions produced free neutrons, residing outside any atomic nucleus. These neutrons can slowly fuse with abundant heavy nuclei to produce the relatively rare ones, encountering no electrical obstacle because the neutron is electrically uncharged. Unhindered by electrical repulsion, the neutrons permit the extension of stellar nucleosynthesis from iron all the way to uranium. A free neutron also can decay into a proton, and the rapid capture of both neutrons and protons during supernova explosion helps to forge the heavier elements.

Busso et al. (1999) reviewed nucleosynthesis in asymptotic giant branch (AGB) stars. Arnett (1973, 1995) reviewed explosive nucleosynthesis in stars, whereas Meyer (1994) gave us a review of the r-, s- and p- processes in nucleosynthesis.

10.5.5 Big-Bang Nucleosynthesis

The synthesis of elements inside stars is an incomplete scenario for it does not explain the origin of any of the hydrogen and most of the helium in the observable universe. Moreover, because deuterium is destroyed rapidly inside stars, there also must be another explanation for its cosmic existence. As proposed by the eclectic George Gamow (1904–1968) and his colleagues, these elements must have been produced during the exceptionally hot and dense first moments of the big-bang explosion that gave rise to the expansion of the universe. In fact, it once was thought that all of the elements found today might have been created back then.

Working with his young colleague Ralph A. Alpher (1921–2007), Gamow proposed that all of the elements were produced in a chain of nuclear reactions during the earliest stages of the expanding universe. They supposed that the original substance of the material universe, the cosmic "ylem," consisted solely of neutrons at high temperature. Some of these neutrons decayed into protons, and successive captures of neutrons by protons led to the formation of the elements.

This novel idea was published in 1948, in a paper titled "The Origin of the Chemical Elements," with Hans A. Bethe (1906–2005) added as an author, even though he contributed nothing to the research or the writing. This was done to make a pun on the first letters of the Greek alphabet – alpha, beta, and gamma or α, β, and γ (Alpher et al. 1948).

Two years after the α–β–γ paper, the Japanese astrophysicist Chushiro Hayashi (1920–2010) showed that in the first moments of the expansion, the temperature was hot enough to create particles such as neutrinos and positrons, the anti-matter particles of electrons (Hayashi 1950). The mutual interaction of all of the subatomic particles present in the first moments of the big bang establishes the relative number of neutrons and protons, which in turn determine the amount of helium produced.

With this correction, modern computations by Robert V. Wagoner (1938–), David Schramm (1945–1997) and their colleagues conclusively demonstrated that all of the hydrogen and most of the helium found in the cosmos today were synthesized in the immediate aftermath of the big bang (Wagoner et al. 1967; Peebles et al. 1991).

These processes are known as *big-bang nucleosynthesis*, and they tell us how much matter is now in the universe in both visible and invisible forms. Boesgaard and Steiman (1985) have reviewed the theory and observations of big-bang nucleosynthesis.

So Gamow was partly right. The lightest and most abundant element, hydrogen – and therefore the majority of atoms that we see today – indeed were formed even before the stars existed, in the immediate aftermath of the big bang that produced the expanding universe. All of the hydrogen that is now found in stars, interstellar space, and the rest the universe was created then, about 14 billion years ago, and so was most of the helium, the second most abundant element.

Why weren't all of the heavier elements produced in the early stages of the big bang? By the time that the expanding universe became sufficiently cool to allow nucleosynthesis to occur, at a temperature of about 1 billion K, it had expanded into a greater volume and become less dense. The rate of helium burning by the triple alpha process is proportional to the square of the density at a given temperature, and at 1 billion degrees the density of the expanding universe is too low by many orders of magnitude for appreciable operation of the triple alpha process.

In simpler terms, the expanding universe was not and is not in equilibrium, so it rapidly cooled down and thinned out to low density, making the simultaneous collision of three helium nuclei – alpha particles – nearly impossible. The dense cores of giant stars, however, are in equilibrium for the long intervals of time needed to accrue noticeable amounts of carbon by the triple alpha process. Only the very light elements, such as deuterium and helium could be produced by two-body reactions during the big bang, whereas carbon had to be synthesized by three-body reactions in the interior of stars.

This completes our account of the origin of the elements inside stars and during the big bang, with one oversight: the under-abundant, light elements with atomic weights between hydrogen and helium. These light elements – boron, beryllium, and most of the lithium – probably were produced by *spallation reactions* in which energetic charged particles, known as cosmic rays, strip off the components of heavy nuclei to form light nuclei (Reeves et al. 1973; Reeves 1994). Wallerstein and Conti (1969) reviewed lithium and beryllium in stars. For references to research papers on the observations and explanations of the abundance of light elements see Lang (1999).

10.5.6 The First and Second Generation of Stars

Because big-bang nucleosynthesis produced no elements heavier than helium, the earliest stars had to be composed of the lightest abundant elements: hydrogen and helium. This first generation of stars is known as *Population I stars*. Then, as a result of ongoing stellar alchemy, the most massive first-generation stars forged heavier elements in their cores, scattering them into space by winds or explosions as they died. These heavy elements, known as *metals* to astronomers, then were recycled and incorporated into second generation stars. Because some of their material came from previous stars, these *Population II stars* are polluted somewhat by the heavier elements, the metals.

Baade (1952) proposed the existence of two stellar Populations, finding that pulsating variable stars of the two types have different period-luminosity relations. Ivezic et al. (2012) have provided a recent review of galactic stellar Populations. The first stars to be formed in the universe have been designated *Population III*.

Observations of stellar spectra confirm this scenario. Very old Population I stars, which formed when the universe was young, have less than 0.1 % of their mass in elements heavier than hydrogen or helium. We see these survivors of the

earliest times in the oldest globular star clusters. Astronomers have not yet found completely pure stars with absolutely no metals, but they are confident that the second-generation Population II stars contain a greater proportion of heavy elements, at 2–3 % of their mass.

Audouze and Tinsley (1976) provided an early discussion of the chemical evolution of galaxies; Rana (1991) subsequently reviewed the chemical evolution of our Galaxy; and McWilliam (1997) reviewed abundance ratios and galactic chemical evolution.

Beers and Christlieb (2005) reviewed the discovery and analysis of very metal-poor stars in the Galaxy, whereas Sandage (1986) reviewed the population concept and related topics such as the collapse of our Galaxy. Gratton et al. (2004) reviewed abundance variations within globular star clusters; Wheeler et al. (1989) have reviewed the abundance ratios as a function of metallicity.

10.5.7 Cosmic Implications of the Origin of the Elements

During the billions of years before the Sun was born, massive stars reworked the chemical elements, fusing lighter elements into heavier ones within their nuclear furnaces. Carbon, oxygen, nitrogen, silicon, iron, and most of the other heavy elements were created this way. The enriched stellar material then was cast out into interstellar space by the short-lived massive stars, gently blowing out in their stellar winds or explosively ejected within supernova remnants.

The Sun and its retinue of planets condensed from this material about 4.6 billion years ago. They are composed partly of heavy elements that were synthesized in stars that lived and died before the Sun and planets were born. The Earth and everything on it spawned from this recycled material.

Moreover, the first stars could not have had rocky planets like the Earth because there initially was nothing but hydrogen and helium. The only possible planets would have been icy balls of frozen gas. Without carbon, life as we know it could not evolve on these planets.

Perhaps the most fascinating aspect of stellar alchemy is its implications for life on the Earth. Most of the chemical elements in our bodies, from the calcium in our teeth to the iron that makes our blood red, were created billions of years ago in the hot interiors of long-vanished stars. Therefore, we are all made of "star stuff". If the universe were not very, very old, there would not have been enough time to forge the necessary elements of life in the ancient stars. The lightest element, hydrogen, needed for the water in our bodies, was synthesized when the observable universe was very young, in the first instants of the big bang. Therefore, we are the offspring of both the star and the big bang – true children of the cosmos.

Chapter 11
The Material Between the Stars

11.1 Gaseous Emission Nebulae

The space between the stars, which looks like an empty black void is filled with cold atoms of hydrogen. Stars form out of this supposed emptiness and eventually return to it. We see some of this interstellar material when the brightest stars illuminate nearby regions (Fig. 11.1). The energetic starlight heats, ionizes, and lights up the surrounding material. Such regions are known as *emission nebulae*. In the parlance of modern astronomers, they also are known as *H II regions*, pronounced "H two regions," because they contain ionized hydrogen, where H denotes hydrogen and II indicates that the atom is missing its one electron.

Bright, extended regions in the cosmos have been noticed ever since telescopes were first trained on the night sky. Since they looked like large, diffuse places with a definite size, unlike the unresolved stars, early astronomers called these places *nebulae*, the Latin word for "clouds." In the late eighteenth, century, the French astronomer Charles Messier (1730–1817) published a famous catalogue that listed the brightest nebulae as well as star clusters (Messier 1781); they are now designated by M followed by the number in his list. Some of these regions include the emission nebulae (Table 11.1).

In 1864 the English astronomer William Huggins (1824–1910) discovered emission lines in the visible-light spectra of these objects (Huggins 1864), which explains why the word emission is part of their name. Lines of ionized hydrogen, ionized nitrogen, and a then-unidentified substance named *nebulium* were found.

As it turned out, an emission nebula, or H II region, is nothing more than a great ball of ionized gas with an exceptionally hot, massive star at its center. The ultraviolet radiation of a luminous blue star, of spectral class O5, ionizes and heats the surrounding gas to temperatures of about 10,000 K (Eddington 1926a), and the hydrogen emission lines are emitted when the electrons freed by ionization recombine with the protons to make hydrogen atoms (Zanstra 1927, 1928).

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_11, © Springer-Verlag Berlin Heidelberg 2013

Example: Ionization wavelength for hydrogen and oxygen

The radiation photon energy $h\nu = hc/\lambda$ required for ionization of an atom to a given state of ionization is given by:

$$h\nu = \frac{hc}{\lambda} = E_i, \tag{11.1}$$

or

$$\lambda = \frac{hc}{E_i} = \frac{1.9864 \times 10^{-25}}{E_i} \text{ m}, \tag{11.2}$$

where ν is the radiation frequency, λ is the radiation wavelength, the Planck constant $h = 6.6261 \times 10^{-34}$ J s, the speed of light $c = 2.9979 \times 10^{8}$ m s^{-1}, and E_i is the ionization potential for the stage of ionization in the joule unit of energy. The ionization potential is often given in electron volts (eV), where $1 \text{ eV} = 1.6022 \times 10^{-19}$ J. The number of electrons freed from the atom is equal to the ion's identifying Roman numeral number minus one.

The H II ion is a hydrogen atom missing its one electron with $E_i = 13.5$ eV and O III is an oxygen atom missing two electrons with $E_i = 35.116$ eV. The radiation wavelength λ at which the photon energy is just equal to the ionization energy is given by $\lambda = (hc)/E_i = 1.9864 \times 10^{-25}/E_i = 91.8$ nm and 35.3 nm for ionized hydrogen and doubly ionized oxygen, respectively, where $1 \text{ nm} = 10^{-9}$ m. These wavelengths lie in the ultraviolet range, on the short-wavelength side of blue light, or roughly from 10 to 400 nm.

As shown by the American astronomer Ira S. Bowen (1898–1973), spectral lines of emission nebulae also arise from ionized atoms of oxygen and nitrogen undergoing "forbidden" transitions in the nebulae; such transitions are improbable in the laboratory situation. In a remarkable piece of detective work, Bowen was able to match previously unidentified nebular lines with [O II], [N II], [O III] and [N III], where O and N denote oxygen and nitrogen, respectively; II and III denote that the atom is missing one or two electrons, respectively; and the square brackets [] denote a forbidden line (Bowen 1927, 1928). Thus, in addition to hydrogen, the dominant spectral lines of emission nebulae are those of cosmically abundant oxygen and nitrogen (Table 11.2), as well as sulfur, S II.

As a luminous star's ultraviolet radiation moves out into surrounding space, it is absorbed and consumed by interstellar hydrogen atoms and other less abundant atoms (Fig. 11.2). Eventually, all of the available ultraviolet rays are used up in ionizing atoms close to the star. As a result, the rays cannot travel farther than the immediate vicinity of the star. This is one of the reasons why the night sky remains black outside the periphery of emission nebulae; the other reasons are the low temperature and density of the interstellar material.

Fig. 11.1 Rosette Nebula This large, circular emission nebula, or H II region, known as the Rosette Nebula, lies at a distance of about 1.6 kpc, or 5,200 light-years from the Earth and is about 40 pc or 130 light-years in diameter. Parts of this region include nebulae designated as NGC 2237, NGC 2238, and NGC 2239, as well as the open star cluster NGC 2244. The mass of the Rosette Nebula is estimated to be 10,000 solar masses, or $10^4\ M_\odot$. Hot O and B stars in the core of the Rosette Nebula exert pressure on the nearby interstellar material, triggering star formation, and heat the surrounding gas to a temperature of about 6 million K, causing it to emit x-rays observed from the *Chandra x-ray Observatory*. (Courtesy of KPNO/CTIA.)

Table 11.1 Bright named emission nebulae

	RA (2000)		December (2000)		θ^a	Distance[b](light-years)
Name	h	m	°	′	′×′	
Lagoon Nebula, M 8	18	03.6	−24	23	40 × 90	4,100
(Also known as NGC 6523, central region Hourglass Nebula and star cluster NGC 6530)						
Omega Nebula, M 17	18	20.4	−16	11	37 × 47	5,000
(Also known as NGC 6618, Horseshoe Nebula, and Swan Nebula)						
Orion Nebula, M 42	05	35.3	−05	24	60 × 65	1,344
(Also known as NGC 1976, northwestern part is M 43, NGC 1982)						
Rosette Nebula, NGC 2237	06	33.8	+05	00	78 × 78	5,200
(Nebulous region includes NGC 2238, NGC 2239, NGC 2246 and cluster NGC 2244)						
Trifid Nebula, M 20	18	02.4	−23	02	27 × 29	7,600
(Also known as NGC 6514)						

[a] Angular diameter θ in minutes of arc, designated ′

[b] Distance in light-years, where 1 light-year = 9.46×10^{15} m. For comparison, 1 pc = 1 parsec = 3.0857×10^{16} m

Table 11.2 Intense spectral lines of emission nebulae

Element	Wavelength (nm)
[O II]	372.62
[O II]	372.89
[O III]	436.32
Hβ	486.1332
[O III]	495.891
[O III]	500.684
[N II]	654.81
Hα	656.28
[N II]	658.36
[S II]	671.6440
[S II]	673.0816

A central blue star of spectral class O5 has a main-sequence lifetime of only about 3.7 million years, and these massive, relatively young stars are still embedded within the material from which they formed. Radiation pressure of the hot, luminous star will drive most of the surrounding gas away in a few million years. So an emission nebula, or H II region, is no older than its central star, which is enveloped by dense material that coalesced to form the central star.

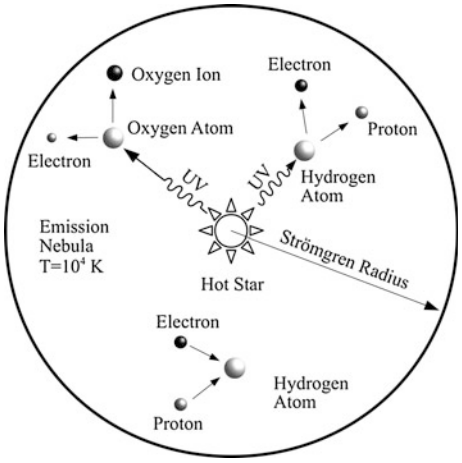


Fig. 11.2 Spheres of ionization Ultraviolet radiation, denoted by UV, from a hot star ionizes hydrogen and other atoms in its immediate vicinity, creating a nebulous region that radiates emission lines and contains abundant ionized hydrogen, denoted H II. They are known as emission nebulae or H II regions. In this figure, the size of the atoms, ions, electrons, and protons are exaggerated greatly. The ionization by the UV creates numerous free electrons and protons that are not attached to atoms. They subsequently recombine to make atoms in a process of continued ionization and recombination. The free electrons can emit two types of radiation, illustrated in Fig. 11.3. At large distances from the star, its ultraviolet rays are all absorbed and can travel no farther into surrounding space. This limits the radius of the emission nebula, or H II region, to the Strömgren radius at about 30 light-years. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

Table 11.3 Physical properties of emission nebulae (H II regions)

N = number density = N_e = electron density = 10^7–10^{10} m^{-3}
T_e = electron temperature = 10,000 K = 10^4 K
R_s = Strömgren radius = 0.65–326 light-years = 0.2–100 pc $\approx$ (0.6–300) $\times$ 10^{16} m
T_{eff} = effective temperature of ionizing star, spectral class O9–O5 = (3 − 5) $\times$ 10^4 K

Because hydrogen is by far the most abundant element in interstellar space, as it is in the stars, the emission nebulae contain large amounts of ionized hydrogen. They contain electrons that were freed from their atomic bonds during ionization. An electron number density, N_e, and electron temperature, T_e, characterize these free electrons. A large, extended emission nebula might have an electron density of $N_e = 10^7$ m^{-3}, while a compact one could have $N_e = 10^{10}$ m^{-3} or more (Table 11.3). These number densities can be significantly greater than the typical density of interstellar hydrogen atoms at about 10^6 m^{-3}. The electron temperature of an emission nebula is $T_e = 10^4$ K, the temperature required to ionize its contents and roughly equal to the disk temperature of its central star.

We can infer the density of an emission nebula, or H II region, once we understand how it shines. When the ultraviolet radiation of a hot star ionizes a nearby hydrogen atom, an electron breaks free of its former atomic bond with a proton. The electron moves away at high speed due to its relatively low mass compared to that of the proton, but when the free electron encounters another proton, the two particles attract one another by their opposite electrical charge, and one of two things can happen: The electron can continue moving into surrounding space, deflected from a straight path; or the electron can be captured, joining together with the proton to again form a hydrogen atom. The electron emits radiation in either situation (Fig. 11.3), and the intensity of the radiation increases with the number of free electrons participating in the process. When a free electron is captured by a proton to make a hydrogen atom, the electron cascades down through the ladder of possible orbital energies, emitting radiation in the process. This radiation is known as *recombination radiation*, because the two recombine to make an atom; it also is called *free-bound radiation* because the former free electron is bound once again.

We have encountered the allowed electron orbits when discussing the Bohr atom and the Balmer series of hydrogen lines (Sect. 6.2), and the recombination transitions are between those electron orbits, designated by integer $n = 1, 2, 3\ldots$ and $m = 1, 2, 3,\ldots$. The alpha transition, designated α, is for $m - n = 1$. At visible wavelengths, the most intense hydrogen emission line is the Balmer α transition at a red wavelength of 656.28 nm while the ultraviolet α transition is known as Lyman α at 121.567 nm.

Recombination lines are also detected from emission nebulae, or H II regions, at radio frequencies, for the cosmically abundant elements of hydrogen, helium

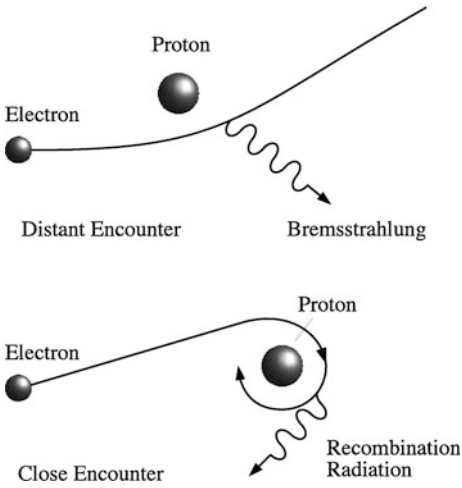


Fig. 11.3 Radiative interactions between electrons and protons When an electron moves rapidly and freely outside an atom, it inevitably passes near a proton in the ambient gas. There is an electrical attraction between the electron and proton because they have equal and opposite charge and this pulls the electron toward the proton. If the interaction is distant, it bends the electron's trajectory and alters its speed. The electron then emits electromagnetic radiation known as *bremsstrahlung* from the German for "braking radiation"; this also is called *free–free radiation* because the electron remains free and unattached to the proton. In a close encounter, the electron goes into orbit around the proton, forming a hydrogen atom and cascading down through allowed orbital energies. In this case, the electron emits *recombination radiation*, also known as *free-bound radiation*. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

and carbon (Fig. 11.4). Initial suggestions and pioneering observations of radio recombination lines are given in Lang (1999).

The frequency ν_{mn} of the $m - n$ transition when recombining to an ion is given by:

$$\nu_{mn} = \frac{1}{h}|E_m - E_n| = cR_i\left|\frac{1}{n^2} - \frac{1}{m^2}\right| \tag{11.3}$$

where $||$ denotes the absolute value, E_m designates the energy of the m^{th} electron orbit, and the wavelength of the $m - n$ transition is given by $\lambda_{mn} = c/\nu_{mn}$. Here the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, the Planck constant $h = 6.6261 \times 10^{-34}$ J s, and the Rydberg constant R_i for an ion is:

$$R_i = R_\infty\left(1 + \frac{m_e}{M_i}\right)^{-1} \tag{11.4}$$

where M_i is the mass of the ion and the Rydberg constant, R_∞, is:

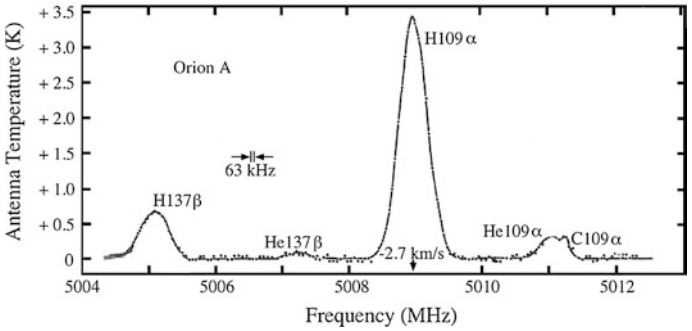


Fig. 11.4 Recombination lines Radio-frequency spectrum of the Orion Nebula, also designated as M 42, showing the hydrogen, H, and helium, He, recombination lines from the α ($m - n = 1$) and β ($m - n = 2$) transitions at high quantum numbers $n = 109$ and $n = 137$. The hydrogen lines are more intense than the helium or carbon lines because of the greater abundance of hydrogen. (Adapted from Lang 1979.)

$$R_\infty = \frac{2\pi^2 m_e e^4}{c\varepsilon_0^2 h^3} \approx 1.09737316 \times 10^7\ \mathrm{m}^{-1}, \tag{11.5}$$

for an electron of mass $m_e = 9.1094 \times 10^{-31}$ kg and elementary charge $e = 1.6022 \times 10^{-19}$ C, where $\varepsilon_0 = 8.8542 \times 10^{-12}$ F m^{-1}, the electric constant.

For hydrogen, the ion is a proton with a mass $M_i = m_p = 1{,}836.153\ m_e$, and the frequency of the $m - n$ transition is given by:

$$\nu_{mn} \approx 3.28805 \times 10^{15} \left(\frac{1}{n^2} - \frac{1}{m^2} \right) \mathrm{Hz} \tag{11.6}$$

Example: What is the recombination line frequency of the H 109α transition?

For the quantum number $n = 109$ and the α transition with $m - n = 1$, the frequency, ν, of the hydrogen recombination line, denoted H 109α, is:

$$\nu = 3.28805 \times 10^{15} \left[\frac{1}{(109)^2} - \frac{1}{(110)^2} \right] \approx 5.0089 \times 10^9\ \mathrm{Hz} \approx 5008.9\ \mathrm{MHz}. \tag{11.7}$$

This line has been detected from many emission nebulae, or H II regions, such as the Orion Nebula (Fig. 11.4).

When the electron continues on its way, it has changed speed during the encounter with the proton, and emitted electromagnetic radiation in the process. This radiation is known as *bremsstrahlung*, from the German word for "braking

radiation"; it also is called *free–free radiation* because the electron begins and remains free. It is emitted at all wavelengths, but at different intensities, and is therefore known as continuum radiation.

The free–free radiation, or thermal bremsstrahlung, from emission nebulae, or H II regions, is also detected at radio frequencies. The ionized gas within the H II region is in thermal equilibrium, characterized by a single temperature and a blackbody brightness distribution given by the Rayleigh-Jeans law at radio frequencies (Sect. 2.4). The ratio of the emission and absorption coefficients is equal to this brightness (Sect. 2.7). When the absorption coefficient is integrated over the line of sight through the H II region we obtain an optical depth, τ_v, given by:

$$\tau_v \approx 3.28 \times 10^{-7} T_e^{-1.35} \nu^{-2.1} EM \tag{11.8}$$

where the electron temperature, T_e, is in units of 10^4 K, the frequency, ν, is in units of GHz or 10^9 Hz, and EM denotes the emission measure in parsecs cm^{-6}. It is the integral of the square of the electron density, N_e, over the line of sight, with $EM \approx 2N_e^2 R_S$ for a region of radius R_S. The optical depth becomes unity at a frequency ν of about 10^9 Hz. For lower frequencies the optical depth is large, and the brightness depends only on the electron temperature. At higher frequencies, the brightness also depends on the emission measure, or on the square of the electron density and the size of the H II region.

The detailed calculations are complex and have a long history. The relevant formulae and pioneering references for the emission coefficients and absorption coefficients for recombination radiation and thermal bremsstrahlung are given in Lang (1999). In our introductory treatment, we only need to realize that the total emission depends on the square of the electron density and the size of the H II region, and this brings us to estimates for their size.

As shown by the Danish astronomer Bengt Strömgren (1908–1987), the central star creates a sphere of ionized gas that envelops it (Strömgren 1939). He found that although the interstellar hydrogen usually is electrically neutral, or un-ionized, the very hot O and B stars could generate enormous but sharply bounded spheres of ionization (Fig. 11.2). For some years, these regions were referred to as "Strömgren spheres," and their size still is designated as the Strömgren radius, which depends on the temperature of the star and the density of the surrounding material. The hotter and more luminous the exciting star, the larger the Strömgren radius becomes, and the denser the surrounding interstellar hydrogen gas, the smaller the Strömgren radius.

For typical interstellar hydrogen densities of 10 million–20 billion atoms per cubic meter, the Strömgren radius ranges from 30 to 300 light-years, or roughly 1–100 parsecs, for central stars with disk temperatures between 26,000 and 48,000 K.

The Strömgren radius, R_s, depends on the rate, S, of emission of the ionizing radiation from the exciting star and the atomic number density, N, of the gas surrounding it. Here S is the total number of ionizing photons that leave the star. An ionizing photon is one with a wavelength short enough to ionize hydrogen, or

at a wavelength less than 91.8 nm. Under steady state conditions, the total number of hydrogen recombinations per unit time inside the Strömgren sphere must equal the total number of photoionizations per unit time, which means that

$$S = \frac{4}{3}\pi R_S^3 \alpha(T) N^2, \tag{11.9}$$

or:

$$R_s = \left(\frac{3S}{4\pi\alpha(T)N_e^2}\right)^{1/3}, \tag{11.10}$$

where we assume the surrounding hydrogen is fully ionized within the Strömgren sphere, where $N = N_e$, the electron density, for a fully ionized gas, and the temperature-dependent recombination coefficient, $\alpha(T)$, of hydrogen to all its excited states except the ground state can be calculated from quantum mechanics and has the value of:

$$\alpha(T) \approx 2.6 \times 10^{-16}\, T_e^{-0.75}\ \mathrm{m^3\ s^{-1}}, \tag{11.11}$$

for $T = T_e$ the electron temperature. For a typical equilibrium temperature of $T_e = 10^4$ K of an emission nebula (Osterbrock 1965), we have $\alpha(T) = 2.6 \times 10^{-19}$ m^3 s^{-1}, and S $= 3 \times 10^{49}$ s^{-1} for an O5 V star with an effective temperature of 48,000 K.

The number density of the fully ionized gas surrounding the emission nebula is $N = N_e = N_p = 10^7$ m^{-3}–10^{10} m^{-3}, with a respective Strömgren radius $R_S = 21 - 0.2$ pc, where 1 pc = 1 parsec $= 3.085678 \times 10^{16}$ m. A lower density corresponds to a larger Strömgren radius, since the ionizing radiation can travel farther from the central star before becoming fully absorbed. The mass, M, of the emission nebula, or H II region, is given by:

$$M = \frac{4\pi}{3} m_P N_e R_S^3, \tag{11.12}$$

where the proton mass $m_P = 1.6726 \times 10^{-27}$ kg and the proton density is equal to the electron density. For a Strömgren radius of $R_s = 1$ pc and an electron density of $N_e = 10^9$ m^{-3}, the mass is $M \approx 2 \times 10^{32}$ kg $\approx 100\ M_\odot$, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg.

Example: Temperature, extent, and mass of an emission nebula or H II region

A star of spectral class O5 V has an effective $T_{eff} \approx 0.5 \times 10^5$ K for its visible disk, and it can ionize the surrounding hydrogen to a temperature of $T = T_e \approx 10^4$ K, where T_e denotes the electron temperature, emitting $S = 3 \times 10^{49}$ ionizing photons every second. If the number density of the surrounding gas $N = N_e = 10^7$ m^{-3}, where N_e denotes the electron density, the O5 V star would have the Strömgren radius of $R_S = 0.6 \times 10^{18}$ m

$= 21$ pc $= 68.7$ light years, where 1 pc $= 3.085 \times 10^{16}$ m $= 3.26$ light-years is the typical spacing between adjacent stars. For an electron density of $N_e = 10^{10}$ m^{-3}, we obtain a Strömgren radius of $R_S \approx 0.617 \times 10^{16}$ m ≈ 0.2 pc.

The mass, M, of the H II region is given by $M = 4\pi m_P N_e R_s^3/3$, where the proton mass $m_p = 1.6726 \times 10^{-27}$ kg, the proton density is equal to the electron density N_e, and the protons dominate the mass rather than lighter electrons. For $N_e = 10^7$ m^{-3} we obtain $M \approx 1.92 \times 10^{34}$ kg $\approx 10^4\, M_\odot$, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, and for $N_e = 10^{10}$ m^{-3} the mass of the H II regions is $M \approx 1.6 \times 10^{31}$ kg $\approx 8\, M_\odot$. The higher the electron density, the lower the total mass, which is in inverse proportion to that density.

11.2 Solid Dust Particles in Interstellar Space

Vast, dark regions provide contrast to the bright light of luminous stars (Fig. 11.5). The blackness initially suggested emptiness – literally, holes in the sky. The dark regions are instead filled with interstellar dust particles that absorb and scatter the light from stars that lie behind them.

Salpeter (1977) has reviewed the formation and destruction of dust grains. Savage and Mathis (1979) provided an early review of the properties of interstellar dust, and Mathis (1990) has discussed interstellar dust and extinction. Draine (2003) provided a review of interstellar dust grains; Mann (2010) reviewed interstellar dust in the solar system; and Stein and Soifer (1983) described dust in galaxies.

It was the American astronomer Edward E. Barnard (1857–1923) who contributed most to our early awareness of these dark and bright regions of the Milky Way. His systematic photographic survey spanned 30 years, first at the Lick Observatory on remote Mount Hamilton in California and then at the Yerkes Observatory near Chicago, culminating in two stunning catalogues of the regions, which he noncommittally called dark markings (Barnard 1919, 1927). Their nebulous form and shape suggested to Barnard that the dark places were not empty; in some cases, they even seemed to interact with the bright regions that enfolded them.

Whereas Barnard was interested chiefly in the peculiar shapes of the dark regions, the German astronomer Maximillian Wolf (1863–1932) was concerned with measuring their distances and absorbing powers. By counting stars in an obscured and an adjacent unobscured region, Wolf was able to demonstrate that the dark areas absorb the light of distant stars (Wolf 1923).

Because Wolf could not detect any substantial difference in the colors of stars that lie outside of and behind the dark nebulae, he concluded that the dark regions must be composed of solid dust particles; their scattering properties depend weakly

Fig. 11.5 Dark dust and bright gas Interstellar dust blocks the light of distant stars, while bright young stars illuminate nearby gas in this part of the North America Nebula, designated NGC 7,000. The infrared detectors aboard the *Spitzer Space Telescope* have penetrated the dark clouds pictured here in optically visible light, viewing young stars in many stages of formation, including gas and dust cocoons, disks, and jets. The North America Nebula is about 1,500 light-years from Earth and spans about 50 light-years. (Courtesy of Karl-Schwarzschild Observatorium, Tautenburg.)

on wavelength, unlike gas atoms that scatter light much more effectively at shorter wavelengths (Sect. 2.7).

The interstellar absorption of starlight was demonstrated by Robert J. Trumpler's (1886–1956) ten-year study of open star clusters at the Lick Observatory (Trumpler 1930). The Swiss-born astronomer used the colors of the brightest stars in each open cluster to infer their luminosity, or absolute magnitude, which he combined with their observed brightness, or apparent magnitude, to determine their distances. However, when these distance estimates were combined with the measured angular diameters of the open clusters to infer their size, he found that the linear diameters increase with distance – and the difference was not trivial. In whatever direction he looked, more remote clusters seemed to be about twice as large as the closer ones. Moreover, the effect was systematic. The farther away a cluster was, the larger it appeared to be.

Concluding that this pervasive, systematic change in physical size was impossible, Trumpler instead assumed that all open clusters actually have the same linear diameters or physical extent. This meant that the initial distance estimates were overestimated due to the absorption of starlight by an amount that increases with distance. The greater the distance, the more the absorption, making the remoter clusters look systematically fainter and even farther away than their actual distance.

The removal of short-wavelength, blue-colored light that occurs when starlight passes through dust is known as *reddening*, because it makes the observed star redder than it would be without the intervening dust (Schild 1977). When radiation passes through a longer path containing a greater density of interstellar dust, the amount of reddening increases, making more distant stars appear redder.

The unfolding story of interstellar dust took an unexpected turn when two American astronomers, William A. Hiltner (1914–1991) and John Scoville Hall (1908–1991), independently observed highly polarized light from reddened stars (Hiltner 1949; Hall 1949). The polarization was oriented along a common plane due to an alignment of the elongated dust particles by the interstellar magnetic field (Davis and Greenstein 1951) – (Fig. 11.6). A magnetic field strength of roughly 10^{-10} tesla is required to align the interstellar dust, and the Zeeman effect from such an interstellar magnetic field was eventually observed.

What is this interstellar dust? The particles must be smaller than 1/10,000th of a meter, or 10^{-4} m, across or they would completely block starlight and not scatter it; also they must be larger than gas molecules, the scattering of which depends strongly on wavelength (Oort and Van de Hulst 1946; Van de Hulst 1949). To weakly absorb and scatter starlight, interstellar dust particles must be comparable in size to the wavelength of visible light, or roughly 10^{-7} to 10^{-6} m across, which is about the same size as the particles of cigarette smoke or the dust motes seen when we look back toward a movie projector.

Cosmic dust absorbs and scatters starlight, which means that interstellar space is not transparent, and the combined effects of scattering and absorption of starlight by interstellar dust is called *extinction*. The observed intensity of radiation, I_λ, at wavelength, λ, is given by:

$$I_\lambda = I_{\lambda o} \exp(-\tau_\lambda) \tag{11.13}$$

where I_{λ_o} is the intensity that would be received at the Earth in the absence of interstellar extinction and τ_λ is the optical depth at the observed wavelength. The approximate interstellar extinction in the optical, or visible wavelength, region varies as λ^{-1} with an optical depth:

$$\tau_\lambda = \frac{C}{\lambda}, \tag{11.14}$$

where the constant C depends on the star (Whitford 1958; Savage and Mathis 1979).

The shape of the curve of starlight extinction as a function of wavelength provides information on the size and composition of the interstellar dust grains; when observations were extended to ultraviolet and infrared wavelengths, more mysteries arose. The amount of extinction increases in a "bump" centered at an ultraviolet wavelength of about 217.5 nm; it may be caused by graphite.

Ice, carbon, and silicates, or some combination of these, are most likely the principal ingredients of interstellar dust. Most of the solid silicate dust found in interstellar space probably came from the cool expanding outer envelopes of evolved giant stars that are rich in oxygen or carbon; the outer atmospheres of these stars exhibit the absorption bands of silicates.

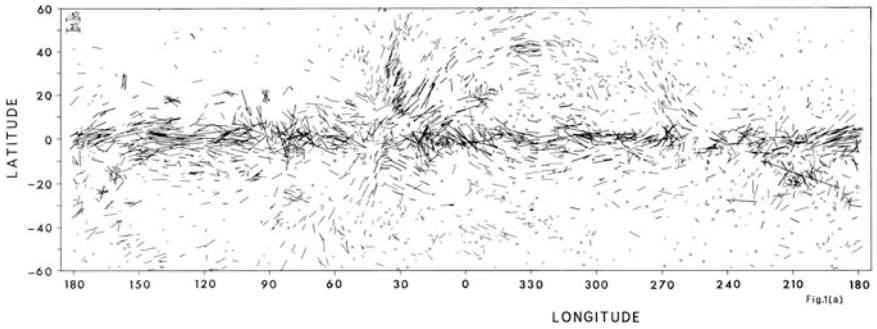


Fig. 11.6 Polarized starlight The light from nearly 7,000 stars is polarized, with the strength and direction of polarization designated by the short lines. The observations are plotted in galactic coordinates where the galactic plane, or the Milky Way, runs horizontally across the middle of the figure. The starlight polarization has been attributed to dust grains elongated along the interstellar magnetic field. (Courtesy of D. S. Mathewson.)

Dust produces strong infrared emission that is studied using instruments aboard the *Spitzer Space Telescope*. Interstellar dust is measured from the emission at long infrared wavelengths, while dust that is being formed in old, evolved stars is observed using medium-infrared wavelengths.

11.3 Radio Emission from the Milky Way

The American radio engineer Karl Jansky (1905–1950) inadvertently discovered radio noise of cosmic origin in the early 1930s, when radio waves were being used extensively for global communications. At that time, the Bell Telephone Laboratories assigned Jansky the task of tracking down and identifying natural sources of radio noise that were interfering with ship-to-shore radio communications at a wavelength of 14.6 m. He constructed a rotating antenna that pointed sideways at the horizon and enabled identification of the interference, including the radio static produced by lightning discharges from distant thunderstorms.

Fortunately, the antenna's wide field of view, of 25°, also pointed part way up into the sky; therefore it could detect a persistent extraterrestrial hiss of unknown origin that was comparable in intensity to terrestrial lightning. By observing the variation of its intensity as a function of direction and time of arrival, Jansky established that the radio source must lie outside the solar system (Jansky 1933a, b, 1935).

The astronomical community almost completely ignored Jansky's results, most likely because he did not publish them in an astronomical journal and his radio techniques were outside the conventional methods of traditional astronomy. It was not until a decade later that amateur astronomer and radio engineer Grote Reber

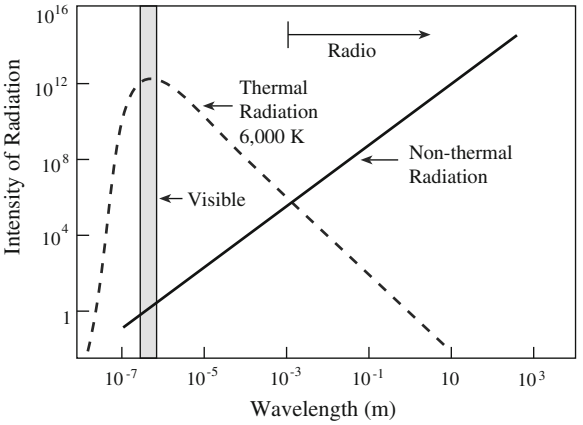


Fig. 11.7 Spectrum of thermal and nonthermal radiation When the intensity of thermal radiation is spread out as a function of wavelength, the resultant spectrum is most intense in a band of wavelengths that depends on the temperature. This peak occurs at visible wavelengths when the temperature is about 6,000 K. Such a hot gas emits radiation at longer wavelengths but at a lower intensity. In contrast, nonthermal radiation is more intense at longer radio wavelengths. High-speed electrons emit nonthermal synchrotron radiation in the presence of a magnetic field (see Fig. 11.8)

(1911–2002) confirmed and extended Jansky's investigations; he published these new results on the "cosmic static" in the *Astrophysical Journal* (Reber 1944).

Because the radio emission of even the nearby Sun could not be detected with Jansky's antenna, normal stars were ruled out as a possible source of the cosmic radio emission. Their thermal radio emission would be exceptionally faint, when compared to that at visible wavelengths, and further diluted by the greater distance. Some nonthermal source of radiation was required that emitted most intensely at the longer radio wavelengths (Fig. 11.7).

As it turned out, the extraterrestrial radio signals were not coming from stars but instead from interstellar space. The cosmic emission is attributed to energetic electrons traveling at a speed close to that of light, and spiraling about the interstellar magnetic field. The energy of the high-speed electrons is comparable to that of the cosmic-ray electrons entering the Earth's atmosphere (Sect. 7.5), with an energy $E \approx 10^9$ eV with 1 eV $= 1.6022 \times 10^{-19}$ J. The electrons are known as *relativistic electrons* because the equations of relativity apply at such high speeds. The speed of the electrons is so great that their kinetic energy of motion cannot be attributed to the thermal energy of a hot gas at any plausible temperature.

Example: Nonthermal velocities

The thermal velocity, $V_{thermal}$, of an electron of mass, m_e, at temperature, T_e, is obtained by equating its kinetic energy $m_e V^2/2$ to the thermal energy $3kT_e/2$, where the electron mass $m_e = 9.1094 \times 10^{-31}$ kg and the Boltzmann constant

$k = 1.38065 \times 10^{-23}$ J K^{-1}. If the electron is moving at the speed of light, with $V_{thermal} = V = c = 2.9979 \times 10^8$ m s^{-1}, then the suggested electron temperature would be $T_e = m_e c^2/(3k) \approx 2 \times 10^9$ K. And a proton moving at that speed would imply a proton temperature $T_p = m_p T_e/m_e \approx 3.6 \times 10^{12}$ K, where $m_p/m_e = 1836$ is the ratio of the proton mass to the electron mass. Cosmic objects in thermal equilibrium, even at the center of the Sun where the temperature is 1.5×10^7 K, cannot plausibly obtain such high temperatures. The radiation emitted by electrons moving at nearly the speed of light is therefore known as nonthermal radiation.

A moving charged particle cannot move straight across a magnetic field, but instead gyrates around it (Focus 11.1), so the high-speed electrons move away from, or around, the magnetism.

Focus 11.1 Charged particles gyrate around magnetic fields

If a charged particle approaches a magnetic field in the perpendicular direction, a magnetic force pulls it into a circular motion about the magnetic field line. Since the particle can move freely in the direction of the magnetic field, it spirals around it with a helical trajectory. The English scientist Oliver Heaviside (1850–1925) first derived the radius, R_g, of this circular gyration (Heaviside 1904). It is given by:

$$R_g = \left[\frac{m}{Ze}\right]\left[\frac{V_\perp}{B}\right], \tag{11.15}$$

where $V_\perp$, is the velocity of the particle in the perpendicular direction, B is the magnetic field strength, and m and Ze respectively denote the mass and charge of the particle. Thus, a stronger magnetic field tightens the gyration into smaller coils, and faster particles will gyrate in larger circles.

For an electron with mass $m_e = 9.1094 \times 10^{-31}$ kg and charge $e = 1.602 \times 10^{-19}$ C, with Z = 1.0, the corresponding gyration radius is:

$$R_g(\text{electron}) \approx 5.7 \times 10^{-12}[V_\perp/B] \text{ m}, \tag{11.16}$$

where $V_\perp$ is in m s^{-1} and B is in tesla, and for a proton it is:

$$R_g(\text{proton}) \approx 1.05 \times 10^{-8}[\text{V}_\perp/\text{B}] \text{ m}, \tag{11.17}$$

where the mass of the protons is $m_p = 1{,}836\, m_e$, and the charge of the proton is the same as that of the electron, just opposite in sign.

The angular gyrofrequency, denoted ω_g and also called the angular cyclotron frequency, is given by:

$$\omega_g = \frac{V_\perp}{2\pi R_g} = \frac{ZeB}{2\pi m}, \tag{11.18}$$

and the *gyrofrequency*, ν_g, also called the *cyclotron frequency*, is given by $\nu_g = \omega_g/(2\pi)$. For an electron the gyrofrequency is:

$$\nu_g = \frac{\omega_g}{2\pi} = \frac{eB}{2\pi m_e} \approx 2.8 \times 10^{10} B \text{ Hz}, \tag{11.19}$$

with a magnetic field strength B in tesla.

These expressions only apply if the velocity is not close to the velocity of light, c. At high particle velocities approaching that of light, the radius equation is multiplied by a Lorentz factor $\gamma = [1 - (V/c)^2]^{-1/2}$ and the gyrofrequency expression is divided by γ, owing to the increase in mass from a rest mass m to γm, which becomes unimportant at low velocities when $\gamma = 1$.

The frequency, ν_s, of synchrotron radiation from high-speed electrons is amplified by an additional factor of γ^3 so $\nu_s = \gamma^2\ \nu_g$ (see subsequent text).

High-speed electrons spiral about the interstellar magnetic field and emit nonthermal *synchrotron radiation* at radio wavelengths (Fig. 11.8). The name is derived from General Electric Company's synchrotron particle accelerator where the linearly polarized light was first seen (Elder et al. 1947). Theoretical studies by Schott (1912) indicated that a charged particle moving in the presence of a magnetic field would emit radiation, and the discovery of synchrotron radiation in a terrestrial particle accelerator stimulated further theoretical treatments of this type of non-thermal emission (Schwinger 1949).

Ginzburg (1956) and Ginzburg and Syrovatskii (1965) have demonstrated how synchrotron radiation can explain the radio emission of supernova remnants and galaxies. Nonthermal synchrotron radiation is more intense at longer wavelengths. More energetic electrons have a shorter lifetime, expending their energy by synchrotron radiation at a greater rate than less energetic electrons. High-energy electrons also emit synchrotron radiation at a higher frequency or shorter wavelength than low-energy electrons. This accounts for the nonthermal radiation spectrum in which the source of synchrotron radiation is most intense at longer rather than the shorter wavelengths, since the high-energy electrons, that emit short-wavelength radiation, will be the first to radiate their energy away.

Synchrotron radiation is the non-thermal emission of a high-speed, relativistic electron, of energy, E, spiraling about a magnetic field of strength, B. By relativistic, we mean moving near the speed of light, $c = 2.9979 \times 10^8$ m s^{-1}. Such nonthermal radiation is more intense at longer wavelengths, λ. When the electron moves at high velocities near the speed of light, the radiation of a slow-moving, non-relativistic electron is modified by the Lorentz factor, γ, given by:

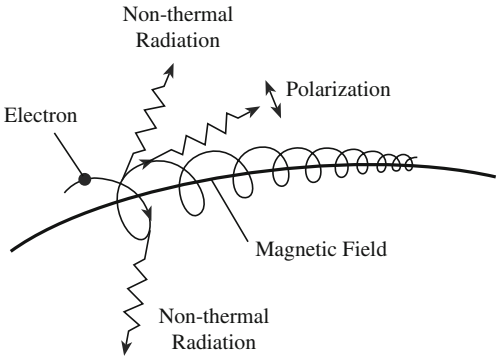


Fig. 11.8 Synchrotron radiation Electrons moving at velocities near that of light emit a narrow beam of *synchrotron radiation* as they spiral around a magnetic field. This emission sometimes is called *nonthermal radiation* because the electron speeds are much greater than those of thermal motion at any plausible temperature. The name *synchrotron* refers to the manmade ring-shaped synchrotron particle accelerator where this type of radiation was first observed; a synchronous mechanism keeps the particles in step with the acceleration as they circulate in the ring

$$\gamma = \left[1 - \left(\frac{V}{c}\right)^2\right]^{-1/2} = \frac{E}{m_e c^2} \tag{11.20}$$

for an electron moving at a velocity, V, with energy, E, and an electron mass $m = \gamma m_e$ with a rest-mass $m_e = 9.1094 \times 10^{-31}$ kg, or non-moving, energy $m_e c^2$ given by

$$m_e c^2 = 8.187 \times 10^{-14}\,\mathrm{J} = 5.11 \times 10^{-4}\,\mathrm{GeV} = 0.511\,\mathrm{MeV}, \tag{11.21}$$

where $1\ \mathrm{GeV} = 1.6022 \times 10^{-10}\ \mathrm{J} = 10^3$ MeV. High-speed electrons produce synchrotron radiation at a frequency v_s given by:

$$v_s = v_g \gamma^2 = \frac{V_\perp \gamma^2}{2\pi R_g} = \frac{eB\gamma^2}{2\pi m_e} = 2.8 \times 10^{10}\, B\, \gamma^2\,\mathrm{Hz}, \tag{11.22}$$

where v_g is the gyrofrequency, $V_\perp$ is the velocity of the electron in the perpendicular to the magnetic field direction, R_g is the gyroradius, the magnetic field strength, B, is in tesla, and the gyrofrequency, v_g, of a non-relativistic, slow-speed electron moving in a circle about a magnetic field of strength B is given by $v_g = eB/(2\pi m_e) = 2.8 \times 10^{10}\, B$ Hz with electron charge $e = 1.6022 \times 10^{-19}$ coulombs and electron mass $m_e = 9.109 \times 10^{-31}$ kg.

The average energy loss rate, or luminosity L_S, of a single electron by synchrotron radiation is given by:

$$L_S = \frac{4}{3}\left(\frac{e^4}{6\pi\varepsilon_0^2 c^4 m_e^2}\right)\left(\frac{V}{c}\right) c \frac{B^2}{2\mu_0}, \tag{11.23}$$

or equivalently

$$L_S = \frac{4}{3}\sigma_T c U_B \left(\frac{V}{c}\right)^2 \gamma^2, \tag{11.24}$$

where the electron velocity $V \approx$ c, the Thomson scattering cross section, σ_T, is given by (Sect. 2.7):

$$\sigma_T = \frac{8\pi}{3} r_e^2 = \frac{8\pi}{3}\left(\frac{e^2}{4\pi\varepsilon_0 m_e c^2}\right)^2 \approx 6.65246 \times 10^{-29}\ \mathrm{m}^2, \tag{11.25}$$

and the energy density of the magnetic field is:

$$U_B = \frac{B^2}{2\mu_0}, \tag{11.26}$$

where the magnetic constant $\mu_0 = 1.2566 \times 10^{-6}$ N A^{-1}.

Collecting terms, evaluating the constants, and noting V/c ≈ 1, the luminosity, L_s, of the synchrotron radiation of a single electron is given by:

$$L_s = \frac{4\sigma_T c \gamma^2 B^2}{6\mu_0} \approx 10^{-14} B^2 \gamma^2. \tag{11.27}$$

The lifetime, τ_s, for a relativistic electron against synchrotron radiation is given by:

$$\tau_s = \frac{E}{L_s} = \frac{\gamma m_e c^2}{L_s} = \frac{8.187}{B^2 \gamma}\ \mathrm{s}, \tag{11.28}$$

where the rest mass energy of the electron is $m_e c^2 = 8.187 \times 10^{-14}$ J. Electrons which have a higher Lorentz factor γ, have a shorter lifetime, and because such electrons radiate at higher frequencies and shorter wavelengths than those with lower γ, it means that there is less synchrotron emission at shorter wavelengths and more at longer wavelengths, which explains its nonthermal spectrum.

The total synchrotron luminosity, L_{tot}, for a source of radius, R, and relativistic (high-speed) electron density, N_e, is:

$$L_{tot} = \frac{4}{3}\pi R^3 N_e L_s. \tag{11.29}$$

Example: Synchrotron radiation of high-speed electrons in the Milky Way

Cosmic ray electrons that enter the Earth's atmosphere have an energy $E \approx 1$ GeV, where 1 GeV $= 1.6022 \times 10^{-10}$ J. The cosmic-ray electrons have a lower flux than cosmic ray protons, with comparable energies, but it is the electrons that give rise to synchrotron radiation, not the protons. The Lorentz factor for an electron with this energy is $\gamma = E/m_e c^2 = 1.96 \times 10^3$, where the rest mass energy of the electron is $m_e c^2 = 0.511$ MeV. If an

electron with an energy similar to those of cosmic rays encountered an interstellar magnetic field, of strength $B = 10^{-10}$ tesla, the frequency of the synchrotron radiation from the electron will be $\nu_s = 2.8 \times 10^{10}\ B\ \gamma^2 \approx 10^7$ Hz $\approx$ 10 MHz, and the wavelength of the radiation is $\lambda = c/\nu_s \approx 30$ m, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. So these electrons give rise to synchrotron radiation at radio wavelengths. For one electron, the synchrotron luminosity $= 10^{-14}\ B^2\gamma^2$ J s$^{-1} \approx 3.84 \times 10^{-28}$ J s^{-1}, and its lifetime for expending all its energy by synchrotron radiation is $\tau_s = E/L_s \approx 0.42 \times 10^{18}$ s $\approx$ 13 billion years, which is comparable to the age of the observable universe, where 1 year $= 3.1557 \times 10^7$ s. However, electrons that emit synchrotron radiation at higher frequencies and shorter wavelengths have a much shorter lifetime. To produce a synchrotron luminosity at radio wavelengths comparable to the solar luminosity in visible light $L_\odot = 3.828 \times 10^{26}$ J s^{-1} in a volume whose radius R is comparable to the distance between adjacent stars, or $R = 1$ pc $= 3.0857 \times 10^{16}$ m, the number density N_e of the electrons would be obtained from $L_\odot = 4\pi R^3 N_e L_s/3$, or $N_e = 8.15 \times 10^3$ m^{-3} and, for example, about 100 times less than the density of interstellar hydrogen atoms.

11.4 Interstellar Hydrogen Atoms

Only rare, hot, luminous stars can ionize nearby hydrogen, and we wonder what is in the vast dark places outside them, other than the dust already discussed. Stars still are forming out of something and because the stars are composed mostly of hydrogen, there should be hydrogen atoms within interstellar space. However, it is so cold out there that we cannot detect any atoms at visual wavelengths. None of the atoms are ionized and the electrons are all in their lowest energy ground state, unable to change their atomic orbits.

But the spin, or rotation, of the electron and proton in the hydrogen atom may be oriented in the same direction or opposite directions, or in parallel and anti-parallel configurations. In technical terms, the lowest orbital energy state of atomic hydrogen has a hyperfine splitting arising from the spins of the proton and electron changing from a parallel to anti-parallel configuration. This transition is highly forbidden with an extremely small probability, but when it occurs, radiation is released with a photon energy equal to the energy difference between these two levels, about 6×10^{-6} eV, which corresponds with photons with a frequency of $\nu = 1.420 \times 10^9$ Hz or wavelength $\lambda = 21.1$ cm.

Hendrik C. "Henk" Van de Hulst (1918–2000) predicted that a radio wavelength spectral line might be detected from interstellar regions of electrically neutral, or un-ionized, hydrogen atoms (Van de Hulst 1945). He realized that these regions, now designated *H I regions* and pronounced "H one regions", would be

cold and that most of the atoms would be in their lowest energy state. The electron of the hydrogen atom in this state has two possibilities in the direction of its spin, or rotation; a rare collision between two atoms could result in a change of the spin direction of an electron in one of the atoms. The atom then is in an unstable configuration, so its electron will soon return to the original spin direction. This releases a small amount of energy and produces radiation at a wavelength of 21 cm.

The prediction was published in an obscure Dutch journal, *Nederlands tijdschrift voor natuurkunde*, but the Soviet theorist Iosif S. Shklovskii (1916–1985) confirmed the prediction 4 years later, with greater detail, in the Russian language (Shklovskii 1949). At about this time, Harold I. "Doc" Ewen (1922), a graduate student at Harvard University, became interested in radio astronomy. His advisor, Edward M. Purcell (1912–1991), asked his wife, Beth, who was good with languages, to translate Shklovskii's paper about the 21 cm line. After reading it, Purcell encouraged Ewen to build a radio receiver to search for the transition. They also had a copy of Van de Hulst's original work, translated from the Dutch.

Ewen constructed a radiometer to detect the hypothetical radiation using electronic components scavenged from other places or bought with a $500 grant and $300 from Purcell. Much of the equipment was borrowed every Friday and returned each Monday from Harvard's Cyclotron Laboratory, using a wheelbarrow. Accurate measurements of the expected hydrogen-line wavelength, using terrestrial atomic hydrogen in a laboratory at Columbia University, enabled Ewen to tune his receiver to the precise wavelength of 21.106 cm. At Purcell's suggestion, the receiver was switched between the wavelength of the expected signal and an adjacent one, with a difference that might contain the expected 21 cm line. Such a wavelength-switched, or frequency-switched, receiver has since been widely adapted by radio astronomers to remove unwanted noise from an observed signal containing a spectral line.

The receiver was connected to a simple horn antenna constructed of plywood, lined with copper sheeting, and mounted on a ledge outside a window in the university's physics building. Shortly past midnight, when the Earth's rotation brought the plane of the Milky Way through the beam of the horn antenna, Ewen succeeded in detecting the 21 cm transition with the novel receiver,

As it turned out, Van de Hulst was in Cambridge, Massachusetts, as a visiting professor at Harvard College Observatory and the Australian radio astronomer Frank Kerr (1918–2000) was visiting Harvard on a Fulbright grant. So Ewen and Purcell invited them over to describe the discovery of the 21 cm line and urge them to have their people confirm the result.

At the meeting, the Harvard team learned for the first time that the Dutch group at Leiden had been actively trying to detect the radio transition for several years. A description of the wavelength-switched receiver was provided to Van de Hulst, leading to the conversion of the Dutch system and, in a gracious move Purcell insisted that publication of their discovery be delayed until the Dutch group confirmed it. Their discovery was published when it was confirmed by the Dutch and Australian radio astronomers. A coordinated report from all three centers then

was published in the journal *Nature* (Ewen and Purcell 1951, Müller and Oort 1951).

The detection of the 21 cm line revolutionized studies of the interstellar medium. Snow and McCall (2006) reviewed diffuse atomic and molecular clouds, whereas Kalberla and Kerp (2009) have reviewed the H I distribution of the Milky Way – Kerr (1969) previously reviewed the large-scale distribution of hydrogen in our Galaxy. Frisch et al. (2011) have reviewed the interstellar medium surrounding the Sun. Cox and Reynolds (1987) reviewed our knowledge of the local interstellar medium; McCray and Snow (1979) discussed the violent interstellar medium; and Spitzer (1990) reviewed theories of the hot interstellar gas.

The H I regions do not emit visible light; hence they are invisible at optical wavelengths, but they do emit radio waves that are 21 cm long, with typical parameters given in Table 11.4.

As Van de Hulst pointed out, these spin transitions occur rarely in the tenuous interstellar gas; however, an observer might well detect them when looking through the vast extent of interstellar space. The probability, A_{10}, of a single atom undergoing such a change in its electron spin is:

$$A_{10} = 2.8689 \times 10^{-15}\ \mathrm{s}^{-1}. \tag{11.30}$$

The radiative half-life of this hyperfine transition is $\tau_{1/2} = 1/A_{mn} \approx 11.1$ million years, where 1 year $= 3.1557 \times 10^7$ s. At an average interstellar number density of $N_H = 10^6\ \mathrm{m}^{-3}$, one 21 cm transition of a hydrogen atom would occur every second when observing an H I region with a radius of about 440 m, and when looking through thousands of parsecs, with 1 parsec $= 3.0856 \times 10^{16}$ m, one could detect about 10^{17} transitions every second.

An interesting aspect of the first observations of the 21 cm transition was that the detected line was seen in emission. Collisions between atoms determined the line excitation temperature, which is equal to the kinetic temperature of the interstellar gas. This can be inferred from the Doppler broadening, $\Delta\lambda$, at $\lambda = 21$ cm, with a velocity $V/c \approx \Delta\lambda/\lambda$, where the speed of light $c = 2.9979 \times 10^8\ \mathrm{m\ s}^{-1}$ and the kinetic temperature, T_K, of hydrogen atoms of mass, $m_H \approx 1.67 \times 10^{-27}$ kg moving at velocity, V, obtained from equating thermal energy to kinetic energy. It is given by:

$$\text{Kinetic temperature} = T_K = \frac{m_H V^2}{3k}, \tag{11.31}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}\ \mathrm{J\ K}^{-1}$. The excitation temperatures of the observed 21 cm lines confirmed theoretical expectations that the neutral hydrogen in interstellar space has a temperature of about 100 K (see Spitzer 1948; Spitzer and Savedoff 1950).

Table 11.4 Physical properties of atomic hydrogen (H I) regions

N_H = density of hydrogen atoms = 10^6–10^8 m^{-3}
T_k = kinetic temperature = 10–100 K
R = radius = 3–33 light-years = 1–10 pc ≈ (3–30) × 10^{16} m
ν_{10} = frequency of hyperfine spin transition = 1.420 405 751 768 × 10^9 Hz ≈ 1,420 MHz
λ_{10} = wavelength of hyperfine spin transition = 21.106 114 054 13 cm ≈ 21.1 cm
A_{10} = probability of hyperfine spin transition = 2.8689 × 10^{-15} s^{-1}

Example: Column density of interstellar hydrogen atoms

Radio astronomers use observations of the 21 cm line to infer the line-integrated column density N_{CH} of hydrogen atoms along the line of sight. For a spin excitation temperature, T_S, and optical depth τ, the column density inferred from a 21 cm line of width Δv_L is:

$$N_{CH} = 1.82 \times 10^{22}\ \tau T_S \Delta v_L\ \mathrm{m}^{-2}, \tag{11.32}$$

where the numerical coefficient gives the number of hydrogen atoms per square meter when the full-width to half maximum of the line is in units of km s^{-1}.

For the optically thin case, the measured brightness temperature, T_B, which is equal to τT_S, will be proportional to the column density per unit velocity. For a line width of about 1 km s^{-1} and a brightness temperature of $T_B \approx 10$ K, we would have $N_{CH} \approx 2 \times 10^{23}$ m^{-2}, and since the number density of interstellar hydrogen is $N_H \approx 10^6$ m^{-3}, you would have to look through a length $L = N_{CH}/N_H \approx 10^{17}$ m to detect such a line, which is just a few parsecs.

The kinetic temperature can be inferred from the Doppler broadening of the line, and used to infer the spin temperature. If the spin temperature is equal to the kinetic temperature of $T_K = 100$ K, then the thermal velocity is $V_{th} = (3kT_K/m_H)^{1/2} \approx 1.57 \times 10^3$ m s^{-1} = 1.57 km s^{-1}, where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and the mass of the hydrogen atom $m_H \approx 1.67 \times 10^{-27}$ kg.

11.5 Interstellar Molecules

Soon after the discovery of interstellar atomic hydrogen at radio wavelengths, astronomers began to speculate about the possibility of detecting molecules with radio waves, which are sensitive to the coldest clouds of interstellar matter where molecules might survive (Shklovskii 1953; Townes 1957). Observations of any molecule first required accurate measurements of the wavelength of its radio spectral features in the terrestrial laboratory. Furthermore, radio telescopes with surfaces accurate to a few centimeters or better had to be constructed for receiving

the short-wavelength emission of molecules, and new methods of spectral analysis needed to be developed.

Charles H. Townes (1915–) and his colleagues at Columbia University made the first precision laboratory measurements of the radio-frequency transitions of the hydroxyl (OH) molecule, composed of an atom of oxygen, O, and an atom of hydrogen, H. At an international symposium of radio astronomy in 1955, he presented laboratory measurements of the rotational transitions of other molecules that might be detected, including carbon monoxide and water (Townes 1957).

Alan Barrett (1927–1991) and his colleagues at the Massachusetts Institute of Technology built the sophisticated equipment needed to obtain the first observations of interstellar OH, at the wavelength of 18.005 cm specified by Townes (Weinreb et al. 1963). The discovery was followed closely by intensive searches for OH and eventually resulted in the realization that some sources are as small as stars and act like cosmic masers – *maser* is an acronym for "microwave amplification by stimulated emission of radiation." Reid and Morgan (1981) have reviewed cosmic masers; many pioneering research papers for radio detection of interstellar molecules are mentioned in Lang (1999).

Zuckerman and Palmer (1974) have given us a review of radio radiation from interstellar molecules. Ho and Townes (1983) have described interstellar ammonia; Combes (1991) has reviewed the distribution of CO in the Milky Way – also see Gordon and Burton (1976). Herbst and Van Dishoeck (2009) have reviewed complex organic interstellar molecules; Ehrenfreund and Chamley (2000) reviewed organic molecules in the interstellar medium, comets and meteorites. Snow and McCall (2006) have discussed both diffuse atomic and molecular clouds; Lada and Lada (2003) have reviewed embedded clusters in molecular clouds; Fukui and Kawamura (2010) have reviewed molecular clouds in nearby galaxies; and Solomon and Vanden Bout (2005) have discussed molecular gas at high redshift. The dense molecular clouds are the sites of star formation, which we next discuss.

Townes moved to the University of California at Berkeley and, within a year, he and his graduate students discovered ammonia and water in interstellar space (Cheung et al. 1968). This was followed soon by the detection of the embalming fluid formaldehyde (Snyder et al. 1969), as well as carbon monoxide (Wilson et al. 1970).

As might be suspected, the interstellar molecules are composed of the most abundant atoms in the universe, beginning with hydrogen, but also carbon (C), oxygen (O), and nitrogen (N). They combine to form molecules such as ammonia, NH_3; carbon monoxide, CO; hydrogen cyanide, HCN; water, H_2O; and formaldehyde, H_2CO (Table 11.5).

The early findings triggered an avalanche of molecular searches in which groups of young radio astronomers engaged in an exciting pursuit of previously unseen interstellar molecules. The net result was the discovery of hundreds of interstellar molecules, including complex organic molecules such as ethyl alcohol, or ethanol – the substance that gives beer, wine, and liquor their intoxicating power. Although interstellar matter is generally sparse and tenuous, the molecules

Table 11.5 Abundant interstellar molecules[a]

Chemical symbol	Name of molecule	Year of discovery	Frequency[a]	Wavelength[a] (cm)
OH	Hydroxyl	1963	1665.4 MHz	18.0054
CO	Carbon monoxide	1970	115.27 GHz	0.25911
H_2O	Water	1968	22.235 GHz	1.3483
HCN	Hydrogen cyanide	1970	88.632 GHz	0.33824
NH_3	Ammonia	1968	23.694 GHz	1.2653
H_2CO	Formaldehyde	1969	4829.7 MHz	6.2072

[a] The transition frequencies are given in MegaHertz, abbreviated MHz, or GigaHertz, abbreviated GHz. One MHz $= 10^6$ Hz or a million Hertz and 1 GHz $= 10^9$ Hz or a billion Hertz. The transition wavelengths are given in units of centimeters, cm, where 1 cm $= 0.01$ m $= 0.01$ m

are concentrated within dark dust clouds that can be 1 million times as dense as a typical region of interstellar hydrogen atoms.

The transition frequencies of the numerous interstellar molecules that have been detected are given online at Splatalogue.net. For example, the CO line has a frequency of $\nu = 115.271208$ GHz, where 1 GHz $= 10^9$ Hz. This transition has a spontaneous emission coefficient of $A_{10} = 7.202 \times 10^{-8}$ s$^{-1} \approx 2.3$ yr^{-1}. The minimum temperature required to excite this transition is given by equating the photon energy $h\nu$ to the thermal energy $3kT/2$, to obtain $T \approx 3.7$ K, using the Planck constant $h = 6.6261 \times 10^{-34}$ J s and the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}.

The first discoveries of interstellar molecules had not been anticipated even a decade earlier because astronomers had overlooked the importance of interstellar dust grains in shielding molecules from destructive ultraviolet starlight and acting as a catalyst in forming complex molecules; for reviews of pioneering work see Lang (1999). Nevertheless, it had long been known that molecules could be found in the cool outer atmospheres of some stars (Tsuji 1986).

The most abundant cosmic molecule is molecular hydrogen, denoted H_2, which consists of two atoms of hydrogen, designated H. This molecule cannot be observed directly at radio wavelengths; however, the second most abundant interstellar molecule, carbon monoxide, maps out its distribution. Field, Somerville and Dressler (1966) discussed hydrogen molecules in astronomy, and Shull and Beckwith (1982) reviewed interstellar molecular hydrogen.

Chapter 12
Formation of the Stars and Their Planets

12.1 How the Solar System Came into Being

12.1.1 The Nebular Hypothesis

Where did the Sun and its attendant planets come from? How and when did they form? The most likely explanation is provided by the *nebular hypothesis*, which states that the Sun and planets formed together, as a result of the gravitational collapse of an interstellar cloud of gas and dust also known as the *solar nebula* (Fig. 12.1). This theory accounts for the orderly, aligned motions of the major planets. They all move in a narrow band across the sky, implying that their orbits all lie in nearly the same plane, which nearly coincides with the Sun's equatorial plane. All of the planets move in the same direction within their Sun-centered orbits, and both the Sun and most of the major planets rotate in this direction – Venus and Uranus are the exceptions.

The orbits of most of the planetary moons, or natural satellites, imitate those of the planets in being confined to the planet's equatorial plane and revolving about the planet in the same direction that the planet rotates. It is exceedingly unlikely that the major planets and large moons became aligned by chance.

Although Newton's laws and Kepler's laws describe the present motions of the solar system, they cannot explain the remarkable arrangement of its planets and satellites. Additional constraints are required, which describe the situation before the planets were formed and set in motion. These initial conditions are provided by the *nebular hypothesis*.

The German philosopher Immanuel Kant (1724–1804) introduced the basic idea (Kant 1755). He pictured an early universe filled with thin gas that collected into dense, rotating gaseous clumps. One of these primordial concentrations was the spinning solar nebula. The Sun formed at the center of the solar nebula, and the planets formed from swirling condensations in a flattened disk revolving around the Sun.

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_12, © Springer-Verlag Berlin Heidelberg 2013

Fig. 12.1 Formation of the solar system An artist's impression of the nebular hypothesis, in which the Sun and planets were formed at the same time during the collapse of a rotating interstellar cloud of gas and dust that is called the *solar nebula*. The center collapsed to ignite nuclear reactions in the nascent Sun, and the surrounding material was whirled into a spinning disk where the planets coalesced. (Courtesy of Helmut K. Wimmer, Hayden Planetarium, American Museum of Natural History.)

According to another version of the nebular hypothesis, suggested by Pierre Simon Laplace (1749–1827), the shrinking Sun shed a succession of gaseous rings, and each ring condensed into a planet (Laplace 1796). Then, each planet, in turn, became a small rotating nebula in which its own family of rings and satellites was born.

There is now so much evidence for the nebular hypothesis that it has acquired the status of a theory, whose basic tenets are still valid. The spinning solar nebula, attracted by its own gravity, fell in on itself 4.6 billion years ago, becoming increasingly dense, until the middle became so packed, tight, and hot that the Sun began to shine. The planets formed at the same time, within a flattened rotating disk centered on the contracting proto-Sun.

This is the essence of the original nebular theory, which explains qualitatively the fact that the major planets and their large moons all revolve in the same direction within the plane that coincides with the equator of the rotating Sun. This regular, aligned pattern of motion is a natural consequence of the rotation and collapse of a solar nebula composed of gas and dust from which the Sun and planets were produced.

12.1.2 Composition of the Planets

If the nebular theory is correct, we might expect that all of the planets would have the same composition as the Sun because they all formed from the same interstellar nebula. After all, they should have the same ingredients as the material from which they formed – they do, but with a varying mix.

The abundance of elements in the giant planet Jupiter does indeed mimic that of the Sun, with a predominance of the lightest element, hydrogen. Unlike the Sun, the Earth is mainly composed of heavier elements, and this difference must be explained. It is related to the fact that there are two main types of major planets – the *terrestrial planets* and *giant planets* – that differ in size, composition, and distance from the Sun.

The four planets closest 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, planets that have relatively low mean mass densities. Unlike the inner terrestrial planets, rings and numerous satellites encircle the outer giant planets.

The radius of a planet is determined from its distance and angular size, and the mass of a planet can be inferred from the orbital motions of its large satellites or moons. The radius of Jupiter, for example, is 11.2 times the radius of the Earth and its mass is 318 times that of the Earth. If a planet has no moon, 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 with a planet with 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 1,330 kilograms per cubic meter, abbreviated 1,330 kg m^{-3}. This is comparable to the Sun's mean mass density of 1,409 kg m^{-3}, but it is almost four times lower than that of the Earth at 5,513 kg m^{-3}.

A clue to these differences comes from the locations of the two types of planets when they originated. The terrestrial planets formed in the warm regions of the flattened solar nebula, close to the bright, young Sun, whereas the giant planets formed farther from the Sun in the colder outer regions of the solar nebula (Lewis 1974, 2004).

Modern observations of the interstellar medium indicate that it consists of 71 % hydrogen, 27 % helium, and 2 % heavier elements, where the percentage numbers are by mass. In terms of the number of atoms, about 92 % are hydrogen atoms, nearly 8 % are helium atoms, and all the heavier elements make up less than 1 %, which is roughly consistent with the composition of the Sun.

When the temperature was low enough, some of the materials condensed out of the gas and dust of the solar nebula, but only a very modest fraction of the nebular material outside the Sun ever condensed into the planets. The substances that did contribute to the formation of planets can be divided into three categories, totaling up to no more than 2 % by mass of the original solar nebula. They are the metals (0.2 %), the rocks (0.4 %) and the ices (1.4 %), and they condensed at different temperatures and distances from the young Sun.

Metals, like iron, condensed from gaseous into solid form at the highest temperatures close to the Sun, rock condensed at moderate temperatures of about 1,000 K and water, methane, and ammonia condensed into ices at temperatures

below 150 K. The asteroid belt, located between Mars and Jupiter, is located at the distance from the Sun where it is cold enough for ices to condense, marking the transition between the warm inner regions where the terrestrial planets formed and the cold outer domain of the giant planets.

In the inner regions of the solar nebula, the higher temperatures vaporized icy material that could not condense, leaving only rocky substances of relatively high mass density to coalesce and merge to form the terrestrial planets. Also, the low total mass and high initial temperature of these planets, as well as their proximity to the Sun, did not allow them to capture and retain the abundant lighter gases – hydrogen and helium – directly from the solar nebula.

The rocky terrestrial planets were 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. In a process known as *differentiation*, the denser material (iron) sank toward the center, whereas the less dense rocks (the silicates) remained closer to the surface. The planets then cooled from the outside in as time elapsed, so we can now walk across the Earth's solid surface. Our planet still has a molten core due to heat generated by radioactive elements inside it.

At larger distances from the Sun, where the solar nebula was colder, icy substances condensed to form the cores of giant planets. These cores became sufficiently massive to gravitationally capture some of the surrounding hydrogen and helium, which was pulled into the giant planets. The low temperatures at remote distances from the Sun thus enabled the giant planets to retain the abundant light gases and grow even bigger, with large masses and low mass densities. A'Hearn (2011) reviewed comets as building blocks of the cores of giant planets.

Jupiter's low mass density, for example, indicates that it is composed largely of hydrogen and helium, just as the Sun is. Under the enormous pressures inside massive Jupiter, the abundant hydrogen is compressed into liquid molecular form, and its central temperature of about 17,000 K is leftover heat from its formation. Since Jupiter is not as massive as the Sun, nuclear fusion reactions cannot occur within the giant planet.

When the masses of the Sun and planets are determined, we find that the Sun does not only lie at the heart of our solar system; it also dominates it, which means that most of the nebular mass outside the Sun never became part of any planet. Some 99.866 % of all of the matter between the Sun and halfway to the nearest star is contained in the Sun. All of the objects that orbit the Sun – the planets and their satellites, the comets, and the asteroids – add up to only 0.134 % of the mass of our solar system. Relative to the Sun, the planets are insignificant specks, left over from its formation and held captive by its massive gravity.

Almost all of the hydrogen and helium gas that enveloped the newly formed Sun must have disappeared. The powerful winds of the young Sun apparently cleaned out the solar system, blasting away all of the remaining gases that had not condensed to make planets. Some of the leftover rocky material not found in the terrestrial planets is located in the asteroid belt, and some of the remaining ice is located in the distant precincts of the solar system where the comets reside. Most

of the hydrogen and helium that was not near the giant planets or in the Sun also was blown away in the formative stages of the solar system.

12.1.3 Mass and Angular Momentum in the Solar System

The nebular hypothesis must also be adjusted to explain the current distribution of angular momentum in the solar system. Most of it is concentrated in the orbital angular momentum of Jupiter, and the rotating Sun has less than one percent of the amount of angular momentum carried by this giant planet. In other words, a very small fraction of the mass of the solar system has significant angular momentum, while most of the mass has relatively little angular momentum.

According to the law of conservation of angular momentum, the rotation of a shrinking object will speed up as the radius decreases (Focus 12.1). The young Sun should have therefore been rotating very rapidly when it formed, with a rotation period of just a few hours or less, but the Sun rotates quite slowly today, with each full rotation taking 25.7 days at the solar equator. That corresponds to a rotation velocity of about 2,000 m s^{-1}. The spinning Sun must have slowed down as it aged.

Focus 12.1 How fast was the young Sun rotating?
According to the nebular hypothesis, the Sun and planets formed together as the result of the gravitational collapse of a rotating interstellar cloud, called the solar nebula. Before the formation of the solar system, the solar nebula might have been rotating at a leisurely rate, but collapse would inevitably increase its rotation speed. This is a consequence of the conservation of angular momentum during gravitational collapse. For a body of mass, M, rotation velocity, V, and radius, R, the angular momentum is $M \times V \times R$, and since $V = 2\pi R/P$ for a rotation period, P, the conservation law means that:

$$MVR = \frac{2\pi MR^2}{P} = \text{Constant.} \tag{12.1}$$

Since the Sun contains 99.87 % of the mass of the solar system, we can assume that the mass remains constant during the collapse of the solar nebula to form the Sun and planets.

The nearest star, other than the Sun, is Proxima Centauri, located at a distance of $D = 4.23$ light-years $= 1.30$ parsecs $= 268{,}000$ AU. We might assume the solar nebula initially extended to half this distance or roughly to a radius of 134,000 AU, where 1 AU $= 1.496 \times 10^{11}$ m.

In fact, some comets in our solar system have orbits at about a distance of $a = 100{,}000$ AU from the Sun. Kepler's third law gives the orbital period P of such a comet as:

$$P^2 = \frac{4\pi^2}{GM_\odot} a^3, \tag{12.2}$$

where $\pi \approx 3.14159$, the gravitational constant $G = 6.674 \times 10^{-11}\ \mathrm{m^3\ kg^{-1}\ s^{-2}}$ and the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. Solving for the period of the comet we obtain $P \approx 10^{15}$ s $\approx 3 \times 10^7$ years, and an orbital velocity of $V = 2\pi a/P \approx 10^2\ \mathrm{m\ s^{-1}}$.

If we assume that the solar nebula began at the distance and orbital period of a comet, with a radius $R_{SN} = 100{,}000$ AU and rotation period of $P_{SN} = 10^{15}$ s, then by the time the Sun collapsed to its present size of $R_\odot = 6.955 \times 10^8$ m, its rotation period, $P_\odot$, should have been:

$$P_\odot = P_{SN}\left(\frac{R_\odot}{R_{SN}}\right)^2, \tag{12.3}$$

which gives $P_\odot \approx 2.1$ s and a rotation velocity of $V_\odot = 2\pi R_\odot / P_\odot \approx 2.2 \times 10^9\ \mathrm{m\,s^{-1}}$ if the angular momentum is perfectly conserved during gravitational collapse. But this is an impossibly fast rotation, for the outward centrifugal force of rotation would stop the collapse long before this occurred. The stars with the fastest rotation, have rotation periods of about 10 h or 36,000 s, and rotation velocities of about 300 km $\mathrm{s^{-1}} = 3 \times 10^5\ \mathrm{m\ s^{-1}}$ (Sect. 4.3).

This suggests that a star is not formed by gravitational collapse alone; some other phenomenon must slow the spin, perhaps through mass loss from powerful winds or by magnetic fields that link the star to the surrounding material.

However, the Sun now has an equatorial rotation period of $P_\odot = 25.7$ days, or 2.22×10^6 s, much longer than expected, and a rotation velocity of $V_\odot \approx 1{,}971\ \mathrm{m\,s^{-1}}$. Even in this situation, we have to conclude that some process other than gravitational contraction must have slowed the Sun's spin after its birth.

A powerful solar wind and intense magnetic field during the Sun's youth may have conspired to produce the Sun's current slow rotation. A strong magnetic field, generated by the fast solar rotation in the Sun's early epochs, might have connected the Sun to the distant, slowly rotating material in the surrounding disk, acting as a magnetic brake to solar rotation (Mestel 1968). The rapidly rotating Sun swept the magnetic field by the slow-moving charged particles in the outer solar nebula, producing a drag that slowed down the Sun (Fig. 12.2).

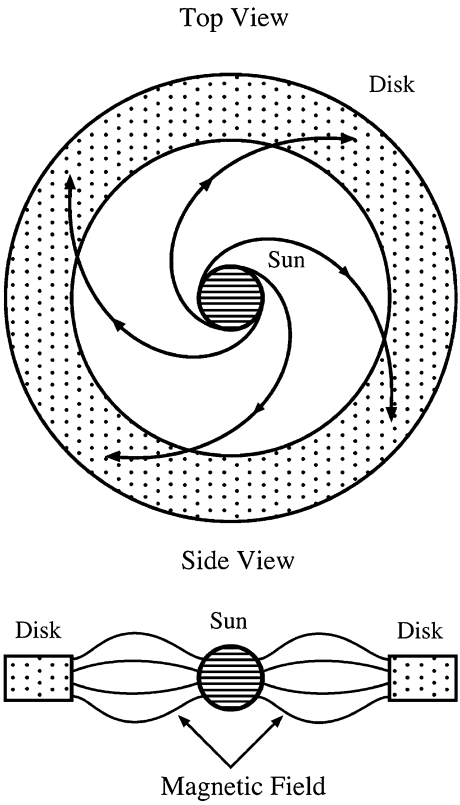


Fig. 12.2 Magnetic brakes The central, rapidly rotating Sun is connected to an ionized, slowly rotating disk by magnetic fields (*side view*). The magnetic field is twisted into a spiral shape (*top view*) and acts as a brake on the Sun's rotation, transferring angular momentum from the Sun to the proto-planetary disk

The solar wind may have been more intense during the Sun's youth, carrying greater amounts of mass away from the Sun and perhaps slowing its rotation. The young Sun might also have been more active than it is today, producing strong, explosive gusts in the solar wind and helping to diminish the Sun's angular momentum. Fiegelson and Montmerle (1999) have reviewed high-energy processes in young stellar objects.

Young stars rotate rapidly, generate powerful winds, and have strong magnetic fields. Older stars rotate slowly like the Sun, and exhibit calmer winds and reduced activity.

Whatever the exact mechanism of sweeping angular momentum away from the Sun, it therefore appears to apply to other stars.

If the nebular theory is correct, it should apply to the formation of other stars, not only the Sun. Massive molecular clouds, in fact, are even now in the process of creating young stars that are currently embedded in the interstellar gas and dust that spawned them. However, not every interstellar cloud is in the process of star formation. For the most part, the interstellar gas and dust is too hot and too tenuous to spontaneously collapse into stars, which is why there is still sufficient interstellar material around to create new stars after billions of years of continued star formation.

12.2 Star Formation

Woodward (1978), Shu et al. (1987), Evans (1999) and McKee and Ostricker (2007) have reviewed star formation. Bastian et al. (2010) discussed a universal initial mass function for stars. Zinnecker and Yorke (2007) reviewed massive star formation, whereas Churchwell (2002) has discussed ultra-compact H II regions and massive star formation. Luhman (2012) has reviewed the formation and early evolution of low mass stars and brown dwarfs; and Zuckerman and Song (2004) summarized our knowledge of young stars near the Sun.

12.2.1 Giant Molecular Clouds

Interstellar clouds of gas and dust provide the raw material for new stars, but they do not form everywhere within interstellar space. Stars form within particularly cold, dense, and massive regions known as *giant molecular clouds*. These contemporary incubators of newborn stars have temperatures as low as 10 K, span tens of light-years, and each one has a mass of up to 1 million solar masses, mainly in the form of hydrogen molecules. These giant molecular clouds are now the dominant star-forming component of the interstellar medium.

As many as 1 million million, or 10^{12}, hydrogen molecules can be packed into every cubic meter of such a giant molecular cloud (Table 12.1). In contrast, there are no more than 100 million, or 10^8, hydrogen atoms in 1 m^3 of the interstellar material outside the giant molecular clouds. These atoms are about 10 times hotter, at about 100 K, than molecular clouds. The dust in the dark molecular clouds blocks the harsh ultraviolet radiation in space and enables chemical reactions to form complex, delicate molecules from the atomic constituents of the interstellar gas.

If a giant molecular cloud becomes sufficiently massive and dense, the mutual gravitation of its parts will overcome the outward gas pressure from inside, and the cloud starts falling in on itself. Once this gravitational collapse is underway, the giant cloud fragments into smaller components; and the pieces collapse until their cores become hot enough to ignite nuclear fusion, burning hydrogen to become stars like the Sun. Some interstellar clouds are even now in the process of creating stars (Fig. 12.3). Thus, stars are continually reformed, as new stars arise in the dark spaces between the old ones.

Table 12.1 Physical properties of giant molecular clouds

N_{H2}	Density of hydrogen molecules = 10^{10}–10^{12} m^{-3}
T_k	Kinetic temperature = 10–30 K
R	Radius = 0.6–32 light-years = 0.2–10 pc $\approx$ (0.6–30) $\times$ 10^{16} m
M	Mass = 10^4–10^6 $M_\odot$ $\approx$ 2 $\times$ 10^{34} kg − 2 $\times$ 10^{36} kg

Fig. 12.3 Mountains of creation The infrared heat radiation of hundreds of embryonic stars (*white*/*yellow*) and windblown, star-forming clouds (*red*), detected from the *Spitzer Space Telescope*. The intense radiation and winds of a nearby massive star, located just above the image frame, probably triggered the star formation and sculpted the cool gas and dust into towering pillars. (Courtesy of NASA/JPL-Caltech/Harvard-Smithsonian CfA/ESA/STScI.)

Bergin and Tafalla (2007) have reviewed cold dark clouds as the initial conditions of star formation. Lada and Lada (2003) have reviewed embedded clusters in molecular clouds; Fukui and Kawamura (2010) have reviewed molecular clouds in nearby galaxies; and Solomon and Vanden Bout (2005) have discussed molecular gas at high redshift.

12.2.2 Gravitational Collapse

Near the beginning of the 20th century, the English physicist and mathematician James Jeans (1877–1946) considered the stability conditions of a gas subject to perturbations in mass density, showing that a fluctuation greater than a critical size – now called the *Jeans length* – or a mass greater than a critical mass – known as the *Jeans mass* – will become unstable to gravitational collapse (Jeans 1902). This collapse occurs when there is insufficient gas pressure to support a large, massive interstellar cloud against the combined gravitational attraction of its component parts.

A spherical gas cloud of a given radius, R, and temperature, T, will undergo gravitational collapse if the mass is greater than the Jeans mass, M_J, given by:

$$M_J = \frac{3kT}{Gm}R, \tag{12.4}$$

where m is the particle mass, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, and the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. The lower the temperature and size, the greater is the likelihood of gravitational instability. A cloud is stable for masses less than the Jeans mass, and only becomes unstable for a mass greater than this critical value.

Example: **Derivation of the Jeans mass**
Suppose the radius, R, of a mass, M, contracted by an amount ΔR. This would compress the gas, increasing the internal thermal energy, $E_{thermal}$, by an amount $\Delta E_{thermal}$ given by:

$$\Delta E_{thermal} = P\left(4\pi R^2 \Delta R\right) = NkT\left(4\pi R^2 \Delta R\right) = \frac{3MkT\Delta R}{mR}, \tag{12.5}$$

where the gas pressure, $P = NkT$, the ideal gas law, N is the number density, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the gas temperature is T, the gas particle mass is m, and the decrease in volume is $4\pi R^2\ \Delta R$.

The gravitational potential energy, E_{grav}, of the cloud (Sect. 3.2) will increase by an amount ΔE_{grav} given by:

$$\Delta E_{grav} = \frac{GM^2\Delta R}{R^2}. \tag{12.6}$$

Collapse will ensue if the increase in gravitational potential energy exceeds the pressure increase that opposes the collapse, or for $\Delta E_{grav} \geq \Delta E_{thermal}$, where the $\geq$ symbol denotes greater than or equal. This means that when the mass M exceeds the Jeans mass M_J, or for

$$M \geq M_J = \frac{3kT}{Gm}R, \tag{12.7}$$

there isn't any equilibrium between the gravitational and thermal forces. The situation is unstable and collapse ensues.

By way of comparison, the mass, M, of a cloud of particles of mass, m, number density, N, and mass density, $\rho = Nm$, is:

$$M = \frac{4}{3}\pi R^3 Nm = \frac{4}{3}\pi R^3 \rho, \tag{12.8}$$

where $\pi = 3.14159$. The cloud's mass density is given by:

$$\rho = \frac{3M}{4\pi R^3} = Nm. \tag{12.9}$$

An equivalent criterion for cloud collapse is that the cloud radius, R, is less than the Jeans radius, R_J, given by:

$$R_J = \frac{Gm}{3kT} M\,. \tag{12.10}$$

Collapse will also occur if the cloud mass density, ρ, is greater than the Jeans density, ρ_J, given by:

$$\rho_J = \frac{M}{\frac{4}{3}\pi R_J^3} = \frac{3}{4\pi M^2}\left(\frac{3kT}{Gm}\right)^3. \tag{12.11}$$

Giant molecular clouds can have a mass greater than the Jeans mass, a mass density greater than the Jeans density, and a radius less than the Jeans radius.

Example: **Gravitational collapse of a giant molecular cloud**
A giant molecular cloud with a number density of hydrogen molecules of $N_{H2} = 10^{12}\ \mathrm{m}^{-3}$, a radius of $R = 1$ parsec $= 3.0856 \times 10^{16}$ m, and a temperature of $T = 10$ K, has a Jeans mass, M_J, of:

$$M_J = \frac{3kT}{Gm} R, \tag{12.12}$$

where $m = 2m_H = 2 \times 1.67 \times 10^{-27}$ kg is the mass of the hydrogen molecule, the Boltzmann constant $k = 1.38065 \times 10^{-23}\ \mathrm{J\ K^{-1}}$, and the gravitational constant $G = 6.674 \times 10^{-11}\ \mathrm{m^3\ kg^{-1}\ s^{-2}}$. Substituting these numerical values into the equation we obtain $M_J = 5.73 \times 10^{31}$ kg for the giant molecular cloud. This is $28.8\,M_\odot$, where the solar mass $M_\odot = 1.989 \times 10^{30}$ kg.

The mass, M_C, for this cloud is:

$$M_C = \frac{4\pi R^3 \rho}{3} = \frac{8\pi m_H N_{H2} R^3}{3}. \tag{12.13}$$

where the mass density $\rho = N_{H2} \times 2m_H$. For the giant molecular cloud under consideration, $M_C = 4.110 \times 10^{35}$ kg $= 2.07 \times 10^5\ M_\odot$. So this giant molecular cloud contains a mass equivalent to 200,000 stars like the Sun, and exceeds the Jeans mass by a factor of about 7,000, definitely fulfilling the condition for gravitational collapse.

It is more difficult for clouds with a smaller mass than the Jeans mass to form stars. They require external compression to begin the collapse.

12.2.3 Triggering Gravitational Collapse

Why haven't all the interstellar gas and dust drawn together to make stars? In most regions of interstellar space, the temperatures are high enough and the mass densities low enough for a long-lived stable equilibrium between outward gas pressure and inward gravitational pull. Take a typical cloud of interstellar hydrogen, an H I region, for example. At a temperature of about 100 K, the hydrogen atoms have a typical kinetic energy that is greater than the gravitational potential energy of the region, and the kinetic energy of H II regions, with a temperature of 10,000 K, is even greater.

Example: **The equilibrium of H I and H II regions**
The number density of un-ionized, or electrically neutral, hydrogen atoms in an interstellar H I region is $N_H = 10^7\ \mathrm{m}^{-3}$, the temperature is $T = 100$ K, and the radius is $R = 1\ \mathrm{pc} = 3.0857 \times 10^{16}$ m, the typical spacing between adjacent stars. The mass density, ρ, of the region is $\rho = N_H m_H = 1.6739 \times 10^{-20}\ \mathrm{kg\ m}^{-3}$, where the mass of the hydrogen atom is $m_H = 1.6739 \times 10^{-27}$ kg. The mass of the region is $M = 4\pi R^3 \rho/3 \approx 2.1 \times 10^{30}\,\mathrm{kg} \approx M_\odot$, the mass of the Sun. This shows that the mass of neutral, unionized hydrogen in the space between the stars is about equal to the mass of hydrogen in stars. The Jeans mass, M_J, of the region is:

$$M_J = \frac{3kT}{Gm}R, \tag{12.14}$$

where $m = m_H$ is the mass of the hydrogen atom, the Boltzmann constant $k = 1.38065 \times 10^{-23}\ \mathrm{J\ K}^{-1}$, and the gravitational constant $G = 6.674 \times 10^{-11}\ \mathrm{m}^3\ \mathrm{kg}^{-1}\ \mathrm{s}^{-2}$. Substituting numerical values we obtain $M_J = 1.14 \times 10^{33}\,\mathrm{kg} \approx 575\,M_\odot$. This is nearly 600 times greater than the mass of the region, which means that an H I region is stable against gravitational collapse.

A typical emission nebulae, a region of ionized hydrogen, might have a proton density $N_P = 10^9\ \mathrm{m}^{-3}$, a temperature of $T = 10^4$ K, and radius of $R = 30\ \mathrm{pc} = 9.257 \times 10^{17}$ m. The mass density of the H II region is $\rho = N_P m_P = 1.6726 \times 10^{-18}\ \mathrm{kg\ m}^{-3}$, for a proton mass $m_P = 1.6726 \times 10^{-27}$ kg, and the mass is $M = 4\pi R^3 \rho/3 \approx 5.56 \times 10^{36}\,\mathrm{kg} = 2.8 \times 10^6\,M_\odot$. This indicates that this H II region contains about as much mass as a giant molecular cloud. Moreover, the Jeans mass for the H II region is $M_J = 3.43 \times 10^{36}\,\mathrm{kg} \approx 1.7 \times 10^6\,M_\odot$, which is so close to the H II region's mass that it seems on the verge of collapse. Many of the central stars of H II regions, or emission nebulae, must still be immersed in the material from which they formed.

The mixture of gas and dust also is stirred into motion here and there by gravitational tugs, the radiation pressure of hot stars, or waves expanding from stellar explosions. These movements also oppose gravitational collapse.

Magnetic fields do not like being pushed together any more than hot particles do, so magnetic pressure can also help an interstellar cloud resist gravity. The interstellar magnetic field can generate a magnetic pressure that is comparable to the interstellar gas pressure in its vicinity and magnetic energy can support a gas cloud against gravity (Focus 12.2),

Focus 12.2 Magnetic energy

A magnetic field of strength B produces a magnetic pressure, P_B, transverse to the direction of the magnetic field. This pressure is given by:

$$P_B = \frac{B^2}{2\mu_0}, \tag{12.15}$$

where the magnetic constant $\mu_0 = 4\pi \times 10^{-7} = 1.2566 \times 10^{-6}$ N A^{-2}. The interstellar magnetic field has a strength of $B \approx 10^{-10}$ tesla, to give $P_B \approx 0.4 \times 10^{-14}$ Pa. This is the pressure carried by the magnetic field (Focus 9.1, Sect. 9.1).

The magnetic pressure is a magnetic energy density, so the magnetic energy of a spherical volume of radius R will by $4\pi R^3 P_B/3$. If a gas cloud of mass, M, and radius, R, is in equilibrium with the magnetic pressure alone, the magnetic energy is equal to the gravitational potential energy GM^2/R, and the magnetic field strength, B, is given by:

$$B = \left(\frac{3\mu_0 GM^2}{2\pi R^4}\right)^{1/2} = \left(\frac{3\mu_0 G}{2\pi}\right)^{1/2} \frac{M}{R^2} \approx 6.33 \times 10^{-9} \frac{M}{R^2} \text{ tesla}, \tag{12.16}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. If the gas pressure, P_g, supports the cloud alone, then multiplying the gas pressure (Sect. 5.4) by the cloud volume and setting it equal to the gravitational potential energy gives:

$$P_g\left(\frac{4\pi R^3}{3}\right) = \frac{MkT}{m_H} = \frac{GM^2}{R}, \tag{12.17}$$

or the temperature required for the support is:

$$T = \frac{Gm_H M}{kR}, \tag{12.18}$$

where the mass of the hydrogen atom $m_H = 1.67 \times 10^{-27}$ kg, and the Boltzmann constant k $= 1.38065 \times 10^{-23}$ J K^{-1}. This is, within a factor of $2/3$, the expression used to calculate the temperature at the center of the Sun.

Example: **Supporting interstellar clouds by gas pressure and magnetic pressure**

The ratio of magnetic pressure, P_B, to gas pressure, $P_g = NkT$, is $P_B/P_g = B^2/(2\mu_0 NkT) = 2.88 \times 10^{28}\, B^2/(NT)$ for a magnetic field strength B, a gas number density N, and a gas temperature T, where the permeability of free space $\mu_0 = 4\pi \times 10^{-7} = 1.2566 \times 10^{-6}$ N A^{-2} and the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. For interstellar space $B = 10^{-10}$ tesla, the density of hydrogen atoms is about $N = 10^7$ m^{-3} and the temperature of interstellar hydrogen atoms is about $T = 100$ K. With these numbers $P_B/P_g \approx 0.3$, or the interstellar magnetic pressure is roughly comparable to the interstellar gas pressure.

For the most part, interstellar clouds merely swirl through space – too hot, agitated, and magnetic to collapse into stars. Compression by an external agent nevertheless can force an isolated cloud into gravitational collapse. Occasionally, gas clouds collide with one another, generating shock waves that can compress the colliding clouds, initiating their gravitational collapse.

A spectacular type of external compression is provided by a nearby exploding star, or *supernova*. When a massive star exhausts its thermonuclear fuel, it can explode and eject a spherical shock wave that expands at a speed of 10,000 km s^{-1}. The wave produced by the detonation of a nuclear bomb is analogous to the shock wave of a supernova.

As proposed by the Estonian astronomer Ernst Öpik (1893–1985), who spent the second half of his career in Ireland, the shocks and expanding remnants of the explosion can trigger the collapse of a normally stable interstellar cloud (Öpik 1953). The shock wave pushes nearby interstellar gas and dust together, compressing clumps of matter to sufficiently high density for gravitational collapse to ensue.

The solar nebula once may have been so spread out that its weak gravity was not sufficient for it to collapse to form the Sun and the planets. Instead, the explosion of a nearby star may have triggered the collapse. Some elements found in meteorites recovered on the Earth are apparently the decay products of radioactive elements that must have been produced in such a stellar explosion no more than a few tens of millions of years before our solar system formed. Because of the rate of radioactive decay, if these elements were created before that time, they would not be around now. Adams (2010) has provided a review of the birth environment of the solar system.

Emission from hot, massive, young stars also can compress nearby gas and dust into gravitational collapse and the formation of new stars. Associations of bright O and B stars that were formed about 1 million years ago, for example, now are expanding and dispersing into space (Ambartsumian 1949, 1955). The intense radiation and powerful winds associated with a previous and nearby generation of O and B stars could have triggered the collapse of neighboring material, giving rise

to the expanding stellar associations. These newer, younger stars may trigger the formation of other stars in the future, in an ongoing process of sequential star formation (Elmegreen and Lada 1977).

Thus, stars do not form just anywhere. They are born either in cold, dense molecular clouds or in proximity to exceptionally massive and short-lived stars by various mechanisms, including the pressure of stellar radiation, winds, and explosions.

12.2.4 Protostars

A star in the process of formation is commonly called a *protostar*. Such an embryonic star shines by the release of gravitational energy during the collapse of interstellar material, but it has not yet begun to shine by nuclear fusion in its core. Protostars are exceptionally bright at infrared wavelengths, which can be used to detect them within dark clouds.

Once an interstellar cloud becomes sufficiently dense, by either external compression or within a giant molecular cloud, the mutual gravitational attraction of its parts will overcome the gas pressure and cause this cloud to start collapsing. As this protostar falls inward, it gains gravitational potential energy, much in the way a waterfall gains energy when its water moves toward the ground. Some of the energy of the protostar is converted into heat as the gas particles fall inward and collide with one another (Fig. 12.4).

Before it becomes a Sun-like star, the entire collapsing protostar is a great, churning caldron of heated gas, with energy transported and released by convection. Eventually, the convection zone retreats toward the outer parts of the protostar, and radiation plays a role in transporting and carrying off energy.

As the gas particles fall inward, their gravitational potential energy is converted into the kinetic energy of motion, and particle collisions transform the kinetic energy into thermal energy, which heats the gas. As long as the thermal energy and heat can be radiated away, the protostar stays cool and the pressure remains too weak to slow the gravitational collapse. However, as the shrinking cloud becomes smaller and denser, some of the thermal energy cannot escape and is trapped inside the collapsing cloud. The internal temperature and pressure will then rise and slow the pace of contraction.

Eventually, the core temperature becomes high enough to ignite thermonuclear reactions, which heat the surrounding protostar material and completely halt the collapse. The protostar then arrives on the main sequence, and a star is born.

More massive protostars possess more gravitational potential energy and can thus collapse faster. They also dissipate the gravitational energy at a relatively faster rate during contraction because of their larger luminosity and greater radiation.

Studies of the youngest star clusters show that their hottest, most massive O and B stars have arrived on the main sequence of the Hertzsprung-Russell diagram,

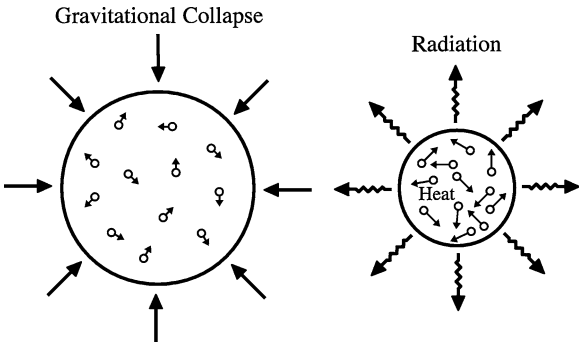


Fig. 12.4 Gravitational collapse produces heat and radiation The collapse of an interstellar cloud of gas and dust (*left*) compresses the cloud and heats it (*right*). When the cloud shrinks, gravitational potential energy is converted into heat as the gas particles fall inward and collide with one another. This also produces radiation that can carry off some of the energy. The velocities of the gas atoms are denoted by arrows that point in the direction of atomic motion and have lengths that increase with the speed of motion. Higher speeds occur in the compressed cloud, where the gas atoms move faster and in all directions

while their cooler, less massive stars have not reached it. The latter stars, of late spectral type, must still be undergoing gravitational contraction from the surrounding pre-stellar medium and have not yet had sufficient time to become hot enough to ignite thermonuclear reactions.

When the observations of young stellar clusters are combined with the theoretical studies of the Japanese astrophysicist Chushiro Hayashi (1920–2010), the pre-main sequence evolution of protostars of different masses can be deciphered (Hayashi 1961, 1966). As illustrated in Fig. 12.5, a protostar's track in the Hertzsprung-Russell (H-R) diagram initially moves straight down and subsequently turns to the left and continue that way until the protostar arrives on the main sequence. These paths in the H-R diagram have been successfully compared with observations of young star clusters (Walker 1956).

Upon arriving at the main sequence, the outward pressure of star's hot gas, which is now heated by nuclear fusion reactions, prevents the star from collapsing further. It has settled down for a long rather uneventful life as a main sequence star, the longest stop in its life history.

How long does it take for a collapsing protostar to become a star? One estimate for the time-scale on which clouds collapse is the free fall time. This is the time it would take a cloud to undergo gravitational collapse from its original shape to a single point, neglecting gas pressure that counteracts this force. The free fall time, τ_{ff}, for unopposed gravitational collapse is given by:

$$\tau_{ff} = \left(\frac{3\pi}{32G\rho}\right)^{1/2} \approx 66{,}430\,\frac{1}{\sqrt{\rho}}\ \text{s}, \tag{12.19}$$

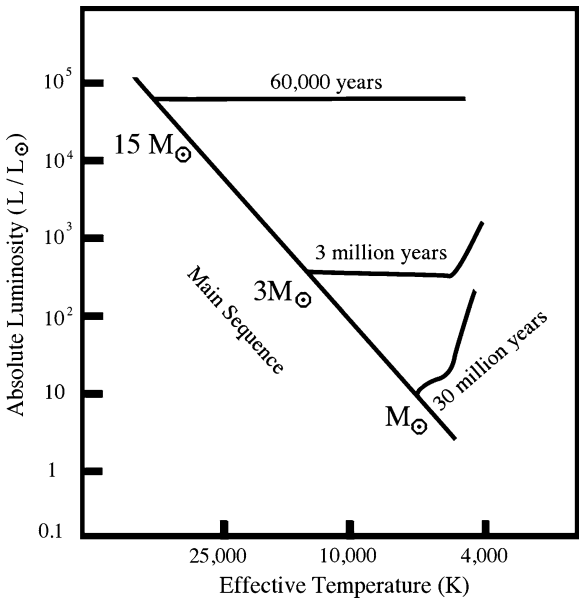


Fig. 12.5 Protostars on the Hertzsprung-Russell diagram Evolutionary tracks of protostars of various masses in the H-R diagram, ending with their arrival on the main sequence when stars have begun burning hydrogen in their cores. The absolute luminosity, L, is given in units of the Sun's absolute luminosity, denoted $L_\odot$. The star mass is given in units of the Sun's mass, designated $M_\odot$. The mass values are specified along the main sequence, from *upper left* to *lower right*. High mass stars, which have greater luminosity than low mass stars, are found at higher points on the main sequence and take a shorter time to arrive there. The protostar lifetimes are given above the relevant track. During star formation, transport of a protostar's internal energy is dominated by either radiation (*horizontal lines*) or by convection (*vertical lines*). Stars with lower mass ultimately have larger interior convective zones. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

where $\pi = 3.14159$, the gravitational constant $G = 6.674 \times 10^{-11}\ \mathrm{m^3\ kg^{-1}\ s^{-2}}$, and ρ is the mass density within the cloud at the time collapse began.

Example: **Free fall time for a giant molecular cloud and for the Sun**
A giant molecular cloud can have a hydrogen molecule number density of $N = 10^{12}\ \mathrm{m^{-3}}$, and the mass, m, of each hydrogen molecule is twice the mass of a hydrogen atom, with $m = 2m_H \approx 3.348 \times 10^{-27}$ kg. The mass density of the giant molecular cloud is $\rho = Nm \approx 3.348 \times 10^{-15}\ \mathrm{kg\ m^{-3}}$ and the free fall time is

$$\tau_{ff} = \left(\frac{3\pi}{32G\rho}\right)^{1/2} \approx 1.15 \times 10^{12}\ \mathrm{s} \approx 36{,}000\ \mathrm{years}, \tag{12.20}$$

where the Newtonian constant of gravitation $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and 1 year $= 3.15576 \times 10^7$ s.

A single cloud does not collapse to a point, with a radius of zero. As the cloud collapses, the density will increase and the Jeans mass of each piece of the cloud will decrease. These pieces will start to collapse on their own. So the cloud fragments into small, dense parts, which collapse faster than the overall cloud.

If an external force triggered the collapse of an interstellar cloud of temperature $T = 100$ K, to form the Sun of mass $M_\odot = 1.989 \times 10^{30}$ kg, collapse would start when the density exceeded the Jeans mass density.

$$\rho_J = \frac{3}{4\pi M_\odot^2}\left(\frac{3kT}{Gm_H}\right)^3 \approx 3.1 \times 10^{-12}\,\text{kg m}^{-3}, \tag{12.21}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. This mass density can be used to obtain a free fall time $\tau_{ff} \approx 3.8 \times 10^{10}$ s $\approx$ 1,200 years.

For the Sun, the unopposed free fall collapse time may have been about 1,200 years, where 1 year $= 3.157 \times 10^7$ s, but there are forces that opposed the collapse as it took place, making it take about 1 million years for the Sun to undergo gravitational collapse to the main sequence. This is much shorter than the thermonuclear lifetime of about 10 billion years once the Sun began to shine by hydrogen burning in its core. When compared to the Sun, more massive stars have shorter free fall times to the main sequence and shorter thermonuclear life times on the main sequence as well.

12.2.5 Losing Mass and Spin

The early stages of star life can be the most active. For example, young stars can have strong stellar winds that drive away protostellar material that may still envelop them. Even now, billions of years after its birth, the Sun generates a solar wind that removes about 10^{-14} solar masses every year; we suspect that its winds were more violent in its youth, clearing out the solar nebula when the planets were formed.

Hot, exceptionally luminous stars can generate immensely powerful winds with greater mass loss than the current solar wind. Mass loss rates of up to 1/10,000th or 10^{-4}, solar masses per year have been observed; at this rate, a star of say 10 times the mass of the Sun would blow itself away in just 100,000 years.

Stellar winds are not confined to youth alone, for stars of advancing age also can produce strong winds. When a star of relatively low mass evolves into a red

giant, the gravitational attraction at its inflated outer layers becomes much smaller than that during the star's former life on the main sequence. This reduces the star's ability to hold onto its outer atmosphere and increases the likelihood that some of it will escape. As a result, red giant stars also have strong winds, sometimes losing a significant fraction of their mass during this stage of stellar evolution.

Younger stars also rotate faster than older stars (Sect. 4.3). The rotation speed can be measured from the Doppler broadening of a star's spectral lines. Rapidly rotating stars have broad lines; slowly rotating stars have sharp and narrow lines. Observations of this line-broadening indicate that the rotation speed of main-sequence stars decreases from left to right on the H-R diagram, from luminous, young stars to fainter, older ones. Bright stars of spectral class O and B rotate at speeds of more than 100 km s^{-1} (Shajn and Struve 1929). The B3 type main sequence star Acherner rotates so rapidly, at more than 225 km s^{-1}, that its equatorial diameter is 56 % greater than its polar diameter. The rapid spin of the A7 type star Altair, at about 240 km s^{-1}, has similarly produced an oblate, non-spherical stellar shape.

Example: **How fast do stars spin?**
The Sun has an equatorial rotation period of $P_{\odot} = 25.67$ days $= 2.218 \times 10^6$ s, where 1 day = 86,400 s. At the Sun's equator the rotation velocity is $V_{\odot} = 2\pi R_{\odot}/P_{\odot} \approx 1.971 \times 10^3\,\mathrm{m\,s^{-1}} = 1.97\,\mathrm{km\,s^{-1}}$, where the Sun's radius $R_{\odot} = 6.955 \times 10^8$ m. This is hundreds of times slower than the fastest rotating main-sequence stars and the expected initial rotation period of the newly formed Sun. The Sun will eventually use up its nuclear fuel and collapse to a white dwarf star with a radius comparable to that of the Earth with $R_E = 6.378 \times 10^6$ m. Since angular momentum is conserved in gravitational collapse, the white dwarf rotation period will be $P_{WD} = P_{\odot}(R_E/R_{\odot})^2 \approx 186$ s. Its rotation velocity will be $V_{WD} = 2\pi R_E/P_{WD} \approx 2.15 \times 10^5$ m $s^{-1} \approx 215$ km s^{-1}, which is comparable to the rotation velocity the Sun might have had when it initially formed about 4.6 billion years ago.

Main-sequence stars of later spectral class rotate at significantly slower speeds. Stars of spectral class F5 have slowed to rotation speeds of about 30 km s^{-1}, while the Sun, at spectral class G2, spins at a leisurely speed of about 2 km s^{-1}. The older, late-type stars may rotate more slowly than early type stars of relatively young age because of the magnetic braking, which astronomers believe accounts for the Sun's unexpectedly slow rotation. The magnetic field that is embedded in a newly formed star and the surrounding nebula will act like a network of elastic cords that tie the star and distant regions together. The inner regions will be moving faster than the outer ones, and the magnetic field will transport some of the rotation from the star outward. Strong stellar winds of the young stars may also play a role in removing angular momentum and spin from the stars.

12.3 Planet-Forming Disks and Planets Around Nearby Stars

12.3.1 The Plurality of Worlds

Just as the ancient Greeks imagined that all matter consists of *atoms*, so they also believed that there were many planets like ours in the universe, created by the coalescence of atoms. In the second century BC, the Greek philosopher Epicurus of Samos (276–194 BC) proposed that the chance conglomerations of innumerable atoms, in an infinite universe, should result in the formation of a multitude of unseen Earth-like worlds.

Then, the Roman poet Lucretius (99–55 BC) wrote about the plurality of worlds, declaring that innumerable particle seeds are rushing on countless courses through an unfathomable universe, making it highly unlikely that our Earth is the only planet to have been created and that all of those other particles are not accomplishing anything (Lucretius 55 BC).

The belief in unseen worlds – some possibly inhabited – that are in orbit around stars other than the Sun, dates at least as far back as the late sixteenth century, to the Italian philosopher and priest Giordano Bruno (1548–1600). He reasoned that other planets would remain invisible to us because they are small and dim and would be hidden in the glare of their host star (Bruno 1584).

During the nineteenth and twentieth centuries astronomers used telescopes to explore the distant reaches of the Milky Way, showing that it contains about 100 billion stars. More recently, hundreds of planetary worlds, which were once only imagined, have been observed orbiting nearby stars.

12.3.2 Proto-Planetary Disks

Planetary systems probably formed around many stars as a result of the gravitational collapse of an interstellar cloud of gas and dust that created the stars, all in accordance with the nebular hypothesis of the origin of the solar system. The collapsing cloud would rotate faster and faster, giving spin to the material that then flattened into a planet-forming disk with a star at the center. Because rotation imparts motion to the colliding material in the direction of spin, the random gas motions of the original cloud are changed into a rotating disk. The centrifugal force of the rotation prevents gas and dust from raining directly onto the central star, instead making it settle into a rotating disk from which planets can form. The direction in which the disk is spinning coincides with the direction of the new star's rotation as well as the direction of the orbits of any planets that may be formed in the disk.

Astronomers have discovered flattened, rotating disks of gas and dust around nearby stars. The first evidence for these planet-forming disks was obtained in the

Table 12.2 Stars with an excess of infrared radiation detected from the *IRAS* satellite[a]

Star	Luminosity ($L_\odot$)	Spectral type (V)	Mass ($M_\odot$)	Distance (light-years)
Vega	37	A0	2.1	25.3
Fomalhaut	18	A3	2.1	25.13
Epsilon Eridani	0.34	K2	0.82	10.49
Beta Pictoris	8.7	A6	1.75	63.4

[a] The luminosity is in solar units of $L_\odot = 3.828 \times 10^{26}\,\mathrm{J\,s^{-1}}$ and the mass is in units of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg

early 1980s with instruments aboard the *InfraRed Astronomical Satellite* (*IRAS*), using technology pioneered by the military to detect the infrared heat of the enemy.

The *IRAS* instruments detected excess infrared radiation from four nearby stars, beyond what would be expected from the star alone (Aumann 1985, Table 12.2). This implied the presence of a circumstellar disk of cool dust in orbit around the star, which would radiate at infrared wavelengths and produce the excess. The hotter stars would shine brightly in optically visible light and emit relatively little infrared.

The *Spitzer Space Telescope* recently used its powerful infrared vision to detect hundreds of stars with excess infrared radiation, suggesting that they harbor planet-forming disks. In fact, the youngest nearby stars usually are found embedded in the dense clouds of the interstellar gas and dust that spawned them. Zuckerman (2001) has reviewed the available knowledge of dusty circumstellar disks.

The closest disk system to our own, surrounding the star Epsilon Eridani, contains two infrared-emitting belts: the first, at approximately the same position as the asteroid belt in our solar system; and the second, denser belt between the first one and a more remote ring similar to our own Kuiper belt.

Instruments aboard the *Hubble Space Telescope* (*HST*) discovered flattened disks of dust swirling around at least half of the young stars in the Orion nebula, that are shining in reflected visible light. The high-resolution and sensitivity of the *HST* also have been used to obtain detailed images of dusty, planet-forming disks surrounding Sun-like stars, providing insights to the beginnings of our solar system (Fig. 12.6). The flattened, rotating disks suggest that the nebular hypothesis applies to them, and material in the disks is expected to coalesce into full-blown planets if it has not done so already.

In the meantime, the circumstellar disk around one of the *IRAS* stars, Beta Pictoris, became the first to be imaged at visible wavelengths by using an occulting disk to block the star's bright light (Smith and Terrile 1984). Detailed observations of the disk were obtained more than a decade later with the *HST* and the Keck telescope (Golimowski et al. 2006). Eventually, the ground-based Very Large Telescope (VLT), located in Chile, was used with adaptive optics to show that a Jupiter-sized world is moving around the star (Fig. 12.7, Bonnefoy et al. 2011). This giant planet, which has been called Beta Pictoris b, is located from its host

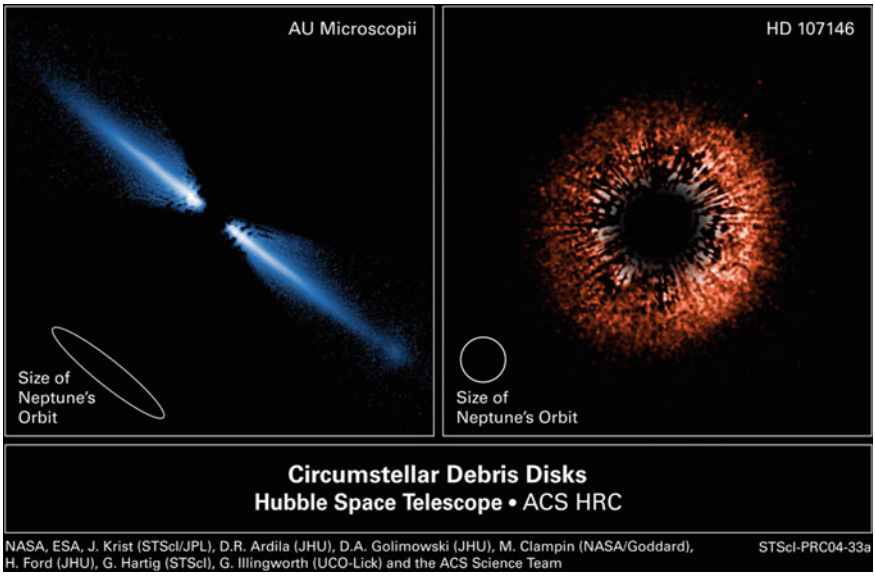


Fig. 12.6 Dusty disks around Sun-like stars Instruments aboard the *Hubble Space Telescope* have obtained these images of the visible starlight reflected from thick disks of dust around two young stars that still may be in the process of forming planets. Viewed nearly face-on, the debris disk surrounding the Sun-like star known as HD 107146 (*right*) has an empty center large enough to contain the orbits of the planets in our solar system. Seen edge-on, the dust disk around the reddish dwarf star known as AU Microscopii (*left*) has a similar cleared-out space in the middle. HD 107146 is 88 light-years away and is thought to be between 50 and 250 million years old, whereas AU Microscopii is located 32 light-years away and is estimated to be just 12 million years old. [Courtesy of NASA/ESA/STScI/JPL/David Ardila – JHU (*right*), and John Krist – STScI/JPL (*left*).]

star Beta Pictoris at a distance between 9 and 15 times the Earth-Sun distance of 1 AU or at about the same distance as Saturn from the Sun at 9.539 AU.

Circumstellar dust around another *IRAS* star, Fomalhaut, also has been imaged with the *HST*. The sharp inner edge of the dust ring suggests that a planet was clearing out the material inside the ring. The *HST* detected the light of a Jupiter-size world orbiting Fomalhaut in the expected place, which is an enormous 115 AU from the star (Kalas et al. 2005, 2008). The fantastic images of Beta Pictoris and Fomalhaut confirmed that infrared-emitting circumstellar disks are indeed signposts of planet formation, but they were obtained more than a decade after the even more astounding detection of the first planets orbiting a Sun-like star. These were also Jupiter-sized worlds but they were orbiting unexpectedly close to their host star.

Williams and Cieza (2011) discussed protoplanetary disks and their evolution; Kley and Nelson (2012) reviewed planet-disk interactions and orbital migration; and Armitage (2011) described the dynamics of protoplanetary disks. Dullemand and Monnier (2010) reviewed the inner regions of protoplanetary disks, and Blum

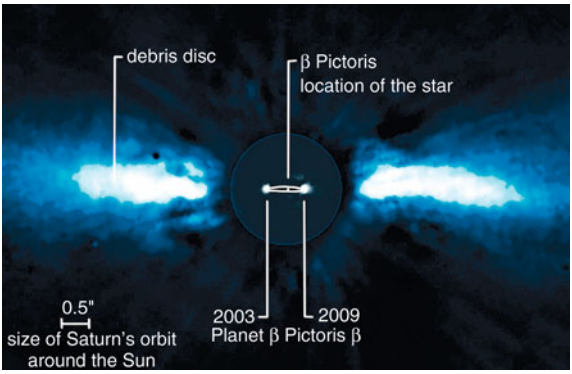


Fig. 12.7 Exoplanet on the move An exoplanet's orbital motion, denoted by the central white elliptical line, was imaged from an adaptive optics instrument attached to the Very Large Telescope (*VLT*) in Chile. The small white spot at the center shows the location of the host star, Beta Pictoris. Observations in 2003 are located at the *left side* of the planet's orbital ellipse and those in 2009 are on the *right side*. The larger dust disk surrounding the host star also is shown by the large flattened *blue* image at the *left* and the *right*. (Courtesy of ESO/A.M. Lagrange.)

and Wurm (2008) have discussed the growth mechanisms of macroscopic bodies in protoplanetary disks.

12.3.3 The First Discoveries of Exoplanets

Individual planets shine by reflecting light that is much fainter than the light of the star that illuminates them. The visible light reflected by Jupiter, for example, is about 1 billion or 10^9 times dimmer than the light emitted by the Sun, and that which is reflected by the Earth is 10 billion times fainter. As a result, planets are almost always too small and too faint to be seen directly in the luminous radiation of their nearby star. Their presence only recently has been inferred from their miniscule gravitational effects on the motions of the star around which they revolve or when they chance to pass in front of a star, momentarily blocking the star's light when viewed from the Earth. Such extrasolar planets that orbit around stars other than the Sun are called *exoplanets*.

The presence of an unseen planet orbiting a normal star like the Sun was first deduced by recording the way its gravity pulls at the star it orbits. The planet and star orbit a common center of mass where their gravitational forces are equal. This fulcrum is closest to the massive star in the stellar case. So the star moves in a much smaller circle, a miniature version of the planet's larger path. The more massive the planet and the closer it is to the star, the stronger the planet's gravitational pull on the star and the more the planet perturbs it.

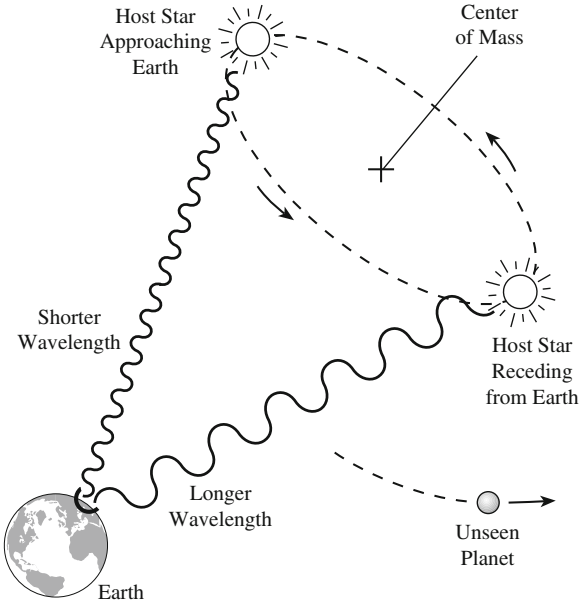


Fig. 12.8 Starlight shift reveals invisible planet An unseen planet exerts a gravitational force on its visible host star. This force tugs the star in a circular or oval path, which mirrors in miniature the planet's orbit. As the star moves along this path, it approaches and recedes from Earth, changing the wavelength of the starlight seen from the Earth through the Doppler effect. When the planet pulls the star toward us, its light waves pile up slightly in front of it, shortening or "blueshifting" the wavelength that we detect. When the planet pulls the star away from us, we detect light waves that are stretched, or "redshifted." During successive planet orbits, the star's *spectral lines* are periodically shortened and lengthened, revealing the presence of the planet orbiting the star, even though we cannot see the planet directly

To detect this tumbling motion, astronomers had to look for the subtle compressing and stretching of starlight as an unseen planet tugged on a star, pulling it first toward and then away from the Earth, causing a periodic shift of the stellar radiation to shorter and then longer wavelengths (Fig. 12.8). To measure the effect, astronomers must observe the wavelength of a well-known spectral feature, called a *line*, and measure the Doppler shift of its wavelength.

However, an orbiting planet produces an exceedingly small variation in the wavelength of spectral lines emitted from its star. Massive Jupiter, for example, makes the Sun wobble at a speed of only about 12 m s^{-1}. To detect the Doppler effect of a star moving periodically with this speed, astronomers would have to measure the wavelengths with an unheard accuracy of at least 1 part in 30 million and use a computer to search for a periodic back-and-forth wavelength change.

Therefore, the effect could not be detected until sensitive spectrographs were constructed to precisely spread out the light rays. The enhanced light-collecting powers of electronic CCDs were then used to record the dispersed starlight. Because no single line shift is significant enough to be seen, computer software

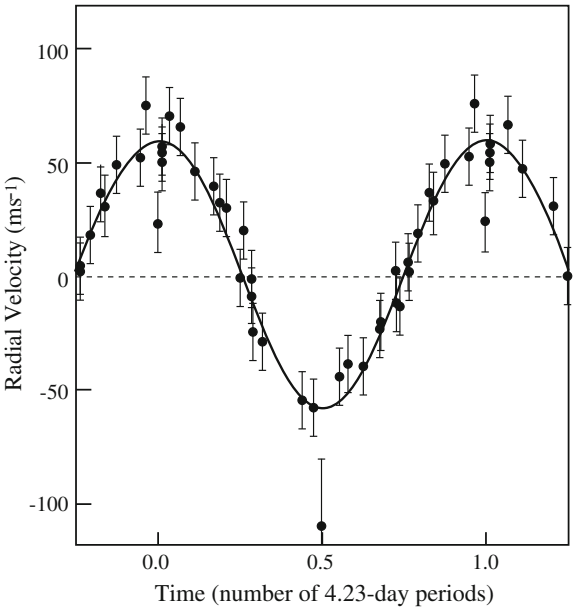


Fig. 12.9 Unseen planet orbits the star 51 Pegasi Discovery data for the first planet found orbiting a normal star other than the Sun. The giant, unseen planet is revolving around the solar-type star 51 Pegasi, located 50 light-years away from the Earth. The radial velocity of the star, in units of meters per second, designated m s^{-1}, was measured from the Doppler shift of the star's spectral lines. The velocity exhibits a sinusoidal variation with a 4.23 day period, caused by the invisible planetary companion that orbits 51 Pegasi in this period. The observational data (*solid dots*) are fit with the solid line, whose amplitude implies that the mass of the companion is roughly 0.46 times the mass of Jupiter. The 4.23 day period indicates that the unseen planet is orbiting 51 Pegasi at a distance of 0.05 AU, where 1 AU is the mean distance between the Earth and the Sun. [Adapted from Mayor and Queloz A Jupiter-mass companion to a solar-type star. *Nature* **378**:355–359 1995).]

had to be written to add up all of the star's spectral lines, which shift together, and to combine them repeatedly at all possible regularities, or orbital periods, and with continued comparison to nonmoving laboratory spectral lines.

It took decades for astronomers to develop these complex and precise instruments. Then, in the 1990s, two Swiss astronomers from the Geneva Observatory in Switzerland, Michel Mayor (1947–) and Didier Queloz (1966–), discovered the first planet that orbits an ordinary star: the faintly visible, Sun-like star 51 Pegasi, only 48 light-years away from the Earth (Mayor and Queloz 1995).

They had detected the back-and-forth Doppler shift of the star's light with a regular 4.23 day period, measured by a periodic change of the star's radial velocity of up to 50 m s^{-1} (Fig. 12.9). To produce such a quick and relatively pronounced wobble, the newfound planet must be large, with a mass comparable to that of Jupiter – which is 318 times heftier than the Earth – and it was moving in a tight close orbit around 51 Pegasi, at a distance of only 0.05 AU (Focus 12.3).

Focus 12.3 Determining the mass and orbital distance of an exoplanet
Planet hunters record the spectral lines of a nearby star, and look for periodic variations in the line-of-sight velocities, V_{obs}, detected from the measured Doppler shifts of the lines. Because the orbital plane is normally inclined to the line of sight, the true orbital velocity, V, is related to the observed velocity V_{obs} by:

$$V_{obs} = V \sin i, \tag{12.22}$$

where i is the inclination angle between the perpendicular to the orbital plane and the line of sight.

The period, P, of the velocity variations is given by Kepler's third law:

$$P^2 = \frac{4\pi^2 a^3}{G(M_1 + M_2)}, \tag{12.23}$$

where M_1 and M_2 respectively denote the mass of the star and its planet, their separation is a, and the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. If r_1 and r_2 denote their respective distances from a common center of mass, and we assume circular orbits, then $r_1 M_1 = r_2 M_2$ with $a = r_1 + r_2 = r_1 \ (M_1 + M_2)/M_2$. Since the orbital velocity $V = 2\pi r_1/P = V_{obs}/\sin i$, we obtain:

$$a = \frac{P V_{obs}}{2\pi \sin i}\left[\frac{M_1 + M_2}{M_2}\right]. \tag{12.24}$$

Substituting this expression into Kepler's third law gives:

$$M_2^3 \sin^3 i = \frac{P V_{obs}^3}{2\pi G}(M_1 + M_2)^2, \tag{12.25}$$

and since the mass of the star will greatly exceed the mass of the planet, or $M_1 \gg M_2$,

$$M_2 \sin i \approx \left(\frac{P}{2\pi G}\right)^{1/3} V_{obs} M_1^{2/3}. \tag{12.26}$$

For the first exoplanet to be discovered, we have $P = 4.23$ days $= 3.655 \times 10^5$ s, for 1 day = 86,400 s, and $V_{obs} = 50$ m s^{-1}. Under the assumption that $\sin i = 1$ and the star's mass is comparable to the Sun, with $M_1 \approx M_\odot = 1.989 \times 10^{30}$ kg, we obtain a planet mass of $M_2 = 7.55 \times 10^{26}$ kg, which is comparable to the mass of Jupiter $M_J = 1.90 \times 10^{27}$ kg. But the exoplanet is nowhere near as far away from

its star as Jupiter is from the Sun. The separation, a, of the newfound exoplanet from its star is given by:

$$a = r_1 + r_2 = r_1\left(1 + \frac{M_1}{M_2}\right) \approx \frac{r_1 M_1}{M_2} = \frac{M_1}{M_2}\frac{PV_{\text{obs}}}{2\pi \sin i} \approx 7.66 \times 10^9 \text{ m} \approx 0.0512 \text{ AU}, \tag{12.27}$$

for 1 AU $= 1.496 \times 10^{11}$ m, the mean distance from the Earth and the Sun. So the new-planet is about one hundredth of Jupiter's distance from the Sun, which is 7.78×10^{11} m $= 5.2$ AU. The exoplanet is even closer to the star than Mercury is from the Sun, at 5.79×10^{10} m or 0.387 AU. Jupiter orbits the Sun once every 11.86 years, Mercury has an orbital period of about 88 Earth days; the exoplanet orbits its star once every 4.23 Earth days.

Planets that are closer to a star move around it with greater speed and take less time to complete an orbit, in accordance with Kepler's third law. Thus, the Earth takes a year, or 365 days, to travel once around the Sun at a mean distance of 1 AU, whereas Mercury, the closest planet to the Sun, orbits our star in a period of 88 days at 0.387 AU. A short orbital period of only 4.23 days meant that the newfound planet was located at a distance of only 0.05 AU from its parent star, or about one-eighth the distance between Mercury and the Sun. Thus, a completely unanticipated planet had been found, rivaling Jupiter in size and revolving around 51 Pegasi in an orbit smaller than Mercury's.

No one anticipated that a giant planet would orbit so close to its star. The intense radiation and powerful winds of the newly formed star were expected to keep any hydrogen from gathering together into a planet, explaining why Jupiter and the other giant planets were formed far from the Sun in the cold, outer precincts of our solar system. However, this was good for planet hunters, for the large mass of a giant world would produce a more pronounced velocity change than the smaller mass of an Earth-sized world, and the close orbit meant a short orbital period that might be detected in weeks instead of years.

Fewer than 2 weeks after the announcement of a giant planet circling 51 Pegasi, two American astronomers, Geoffrey W. Marcy (1955–) and R. Paul Butler (1962–), used their own past observations to confirm the result. Once they knew that giant planets could revolve unexpectedly near a star, with short orbital periods, they used powerful computers to reexamine their observations of other nearby stars accumulated during previous years. They subsequently announced the discovery of two more Jupiter-sized companions of Sun-like stars (Marcy and Butler 1996).

12.3.4 Hundreds of New Worlds Circling Nearby Stars

After scientists realized that a large planet could be so near to its star, they knew where and how to look. By monitoring thousands of nearby Sun-like stars for years, American and European teams found hundreds of planets revolving about other nearby stars, most of them massive Jupiter-sized planets. The accelerating pace of discovery is documented at the extrasolar planets encyclopedia at http://exoplanet.eu/ and at http://planetquest.jpl.nasa.gov/.

Udry and Santos (2007) reviewed the statistical properties of exoplanets; Seager and Deming (2010) have reviewed exoplanet atmospheres, and Marcy and Butler (1998) have reviewed the detection of extrasolar giant planets.

Some of the newfound worlds travel in nearly circular orbits, like those in the solar system, but they are much closer to their host star than Mercury is to the Sun. Dubbed "hot Jupiters" because of their size and proximity to the intense stellar heat, they are much too hot for human life to survive or water to exist. Their temperatures can soar to more than 1,000 K, far hotter than the surface of any planet in our solar system. Other newfound planets follow eccentric, oval-shaped orbits that deviate from a circular path, so they venture both near and far from their star. Many flat multi-planet systems also have been found as a result of longer and improved observations, as expected from the nebular hypothesis.

Most of these worlds were discovered by the wobble they create in the motion of their host star, but some of them were discovered when they passed in front of the star, causing it to dim, or blink. If a planet happens to have a near edge-on orbit, as seen from the Earth, it periodically will cross directly in front of, or *transit*, its host star. Such a transit can be seen only if the orbit of the distant planet crosses the line of sight from the Earth, blocking a tiny fraction of the star's observed light and causing it to periodically dim, repeatedly during the planet's endless journey around its star.

The size of a planet can be derived from the size of the dip. The fractional change in brightness, or *transit depth*, is equal to the ratio of the area of the planet to the area of the star. For the Earth and the Sun, as an example, the transit depth is 0.000084. The planet's temperature can be estimated from the characteristics of the star that it orbits and the planet's orbital period.

These have all been indirect detections of exoplanets. As previously mentioned, the important direct confirmation of a planet circling another star was obtained from the ground-based Very Large Telescope in Chile, obtaining images of a Jupiter-sized planet moving around the star Beta Pictoris. Astronomers have also used the Keck I telescope in Hawaii to directly image the orbital motion of three planets around the star HR 8799, using adaptive optics at infrared wavelengths. The host star is roughly 1.5 times as massive as the Sun, about 5 times as luminous, and located 129 light-years away from the Earth. The planets, designated HR 8799 b, c and d, orbit inside a massive dusty disk at distances of roughly twice those of Neptune, Uranus and Saturn from the Sun. Their masses lie between 8 and 10 times the mass of Jupiter.

12.3.5 Searching for Habitable Planets

From a human perspective, the most interesting planets will be those as small as the Earth, in circular orbits at the precise distance from the heat of a Sun-like star to provide a haven for life. Scientists call this location a *habitable zone*, meaning that it could be inhabited – but not necessarily that it is. Such a planet might be detected by the transit method.

The orbital size can be calculated from the period of the repeated transit and the mass of the star. From the orbital size and the luminosity of the star, the planet's temperature can be calculated. This information would indicate whether the planet resides within the warm habitable zone – that is, the range of distances from a star where liquid water can exist on the planet's surface and life might exist. At closer distances, the water would boil away, and at more remote distances it would freeze solid.

The *Kepler* mission is specifically designed to detect hundreds of planets comparable in size to the Earth or smaller and located at or near the habitable zone. By measuring the brightness of 100,000 stars, it detects the periodic dimming of starlight produced when the planets pass in front of the stars. A transit by an Earth-sized planet produces a small change in the star's brightness of about 1/10,000, lasting for 2–16 h.

The *Kepler* mission discovered several hundred new-planet candidates orbiting nearby stars. A few of the potential planets are nearly Earth-sized and orbit in the habitable zone of smaller, cooler stars than our Sun. Because these stars are less luminous than the Sun, the habitable zone is closer and planets within it have orbital periods that are shorter than our year, so they can be recognized in an observation time of a few years. The *Kepler* planet candidates require follow-up observations with the world's best ground-based telescopes to verify that they are actual planets.

In the meantime, the world's best telescopes are being employed to find new exoplanets using the velocity method. The European Southern Observatory's 3.6 m telescope in La Silla, Chile, has discovered many new ones, including several super-Earths, and the 10 m Keck I telescope atop Mauna Kea in Hawaii has been used to discover many more, including a super-Earth with about four times the mass of the Earth.

Example: **An exoplanet in the habitable zone**
The nearby star Gliese 581, the 581th star in the nearby star catalogue of Wilhelm Gliese (1915–1963), has at least six planets. The star is located at a distance of 6.2 pc or 20.3 light-years from the Earth. It has an apparent visual magnitude of $m_V = 10.55$, an absolute visual magnitude of $M_V = 11.56$ and an absolute luminosity of $L = 0.013\ L_\odot$; from the mass-luminosity relation for such stars a mass of $M = 0.31\ M_\odot$ is determined, and from the mass-radius relation a radius of $R = 0.29\ R_\odot$ is found.

The subscript $\odot$ for the luminosity, mass and radius denotes the Sun's value. An effective temperature, $T_{eff} = 3{,}480$ K is inferred from the star's luminosity and radius. The sixth exoplanet detected for this star, named GJ 581 g, has an orbital period of 36.6 days, an estimated distance from its star of 0.146 AU, and a minimum estimated mass of $M_P = 3.1\ M_E$, where M_E, denotes the mass of the Earth (Vogt et al. 2010). [The designation GJ comes from the nearby catalogue by Gliese and Jahreiss (1979)]. For an albedo of $A = 0.3$ and the star's luminosity, the estimated temperature of the exoplanet at its orbital distance is $T_P = 228$ K (Sect. 2.5). That is below the freezing temperature of water, at 273 K, but still thought to be in the habitable zone of the star; an atmospheric greenhouse effect might raise the temperature above freezing as it does on the Earth.

The atmospheres of transiting exoplanets also are being investigated using the *Hubble Space Telescope*, the *Spitzer Space Telescope*, and ground-based infrared telescopes. As a planet passes in front and behind its star, astronomers can subtract the light of the star alone – when the planet is blocked – from the light of the star and planet together prior to eclipse. This isolates the emission of the planet and enables the detection of the infrared spectral signatures of gases in the planet's atmosphere. Water vapor and methane, for example, have been found in the atmosphere of at least one exoplanet, HD 189733 b (Tinetti et al. 2007; Swain et al. 2008); it is a hot Jupiter-size planet that orbits its star in just 2.2 days and is nearly 63 light-years away from the Earth.

Chapter 13
Stellar End States

13.1 A Range of Destinies

No material object can exist forever, and stars are no exception. Although their lives may be measured in millions or billions of years, stars do stop shining when all of the available sources of subatomic energy have been exhausted. Their central thermonuclear reactions, which keep the star hot inside, are then turned off. There is no heat and pressure being generated inside such a star, so the internal support has been removed and it begins its ultimate contraction. The demise of such a star results in the simultaneous creation of a new star from its collapsing core, with a final resting state that depends on the star's mass.

As in the beginning of their lives, the central regions of all dying stars are subject to the unsupported, inward pull of gravity from all sides, and the entire stellar mass is compressed into an increasingly smaller radius. It only stops when some outward pressure grows sufficiently large to halt the contraction, which means that there is an enormous range in stellar size and mass density (Table 13.1). The mean mass density of the Sun, for example, is comparable to that of water, whereas the density of a neutron star is similar to that of the nucleus of an atom.

In its earliest stages, a protostar's gravitational contraction is stopped when the star begins to fuse hydrogen nuclei into helium nuclei, generating the internal heat and pressure that halts the formative collapse. A long time later, when the hydrogen runs out, gravity takes over again and compresses the core, heating it up until helium can be consumed in synthesizing carbon. Enough heat is then generated to balance the relentless force of gravity, and the star's outer atmosphere expands to giant or supergiant size.

What happens next depends on the mass of the star. Stars that have a mass comparable to the Sun's mass begin their ultimate collapse when the helium is gone, ending up as burned-out, Earth-sized white dwarf stars. The ultimate destinies of the rare, more massive, and luminous supergiants are explosive. They can leave a city-sized neutron star behind or be crushed into a stellar black hole. Thus,

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_13,

Table 13.1 Representative mass, radius, and mean mass density of the stars[a]

Star	Mass ($M_\odot$)	Radius ($R_\odot$)	Mean mass density[b] (kg m^{-3})
Red giant star	1.2	100	0.0014
Sun	1.0	1.0	1,400
White dwarf star	0.6	0.01	0.84×10^9
Neutron star	1.5	0.00001	2.1×10^{18}
Black hole[c]	10.0	0.000004	0.2×10^{18}

[a] The mass is in units of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, and the radius is in units of the Sun's radius $R_\odot = 6.955 \times 10^8$ m

[b] The mean mass density $= 3\,M/(4\pi R^3)$ for a star of mass M and radius R

[c] A representative radius for the black hole formed by the collapse of a star of mass M is taken as the Schwarzschild, or gravitational, radius $R_g = 2GM/c^2$ for gravitational constant $G = 6.673 \times 10^{-11}$ N m^2 kg^{-2} and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}

there is a range of destinies: Giants turn into tiny white dwarfs; supergiants turn into even smaller neutron stars; and the bigger, heavier supergiants turn into black holes – all compressed into their end states by the never-ending force of gravity. The more massive the star, the smaller it eventually becomes.

13.2 Planetary Nebulae

After discovering Uranus, in 1781, the English astronomer William Herschel (1738–1822) discovered a small glowing object that he designated a *planetary nebula* because of its round shape, which resembled the disks of planets as seen through a small telescope (Herschel 1786). However, planetary nebulae are not made of planets, and planets are not visible in them. The designation *nebula* is from the Latin word for "cloud", and it was used to distinguish the diffuse planetary nebulae, which have resolved disks, from unresolved, point-like stars.

Herschel and other astronomers soon discovered more of these objects, and one of them, named the Cat's Eye Nebula, had a "condensation" in its center, which turned out to be a star (Herschel 1786). It was eventually realized that every planetary nebula has a star at its center, the exposed core of a dying red giant that illuminates the nebula (Fig. 13.1).

Any star with a moderate mass, comparable to that of the Sun, eventually balloons into a red giant star. As the core nuclear reactions cease, the giant sheds its outer layers, which are blown away. All that remains of that part of the star is gas and dust, the planetary nebula. The central regions of a red giant star will collapse into a smaller white dwarf star.

Another English astronomer, William Huggins (1824–1910), used his spectroscope to find a trio of emission lines in the Cat's Eye Nebula (Huggins 1864, 1868). When heated, a low-density gas radiates these emission lines; therefore, their presence indicated that the planetary nebulae contain hot, rarefied gas.

Fig. 13.1 Planetary nebula When a Sun-like star uses up its nuclear fuel, the star's center collapses into an Earth-sized white dwarf star and its outer gas layers are ejected into space. Such a planetary nebula is named after its round shape, which resembles a planet as seen visually in small telescopes, and is not related to planets. The shells of gas in the planetary nebula NGC 6751, shown here, were ejected several thousand years ago. The hot stellar core, exposed by the expulsion of the material surrounding it, has a disk temperature of about 140,000 K. Its intense ultraviolet radiation causes the ejected gas to fluoresce as a planetary nebula. (A *Hubble Space Telescope* image courtesy of the NASA/STScI/AURA/Hubble Heritage Team.)

However, at that time, no one knew how hot the gas was, or much about its elemental constitution.

The wavelength of one of the emission lines detected by Huggins coincided with hydrogen—the Balmer emission line at 486.1 nm, but the chemical identification of the other two emission lines remained a mystery for more than half a century. Although these green nebular lines, at wavelengths of 495.9 and 500.7 nm, were even stronger than hydrogen, they defeated attempts to identify them with elements known on the Earth. These spectral features initially were attributed to a previously unknown element called "nebulium," but the emission lines eventually were shown to be due to known elements that had become ionized by the ultraviolet light of the bright central stars.

When it was realized that the central stars of planetary nebulae are very hot, the English astronomer Arthur Stanley Eddington (1882–1944) proposed that the ultraviolet starlight ionizes surrounding material, and that the electrons liberated by this photoionization will heat the gas to about 10,000 K (Eddington 1926). The American astronomer Donald Menzel (1901–1976) and the Dutch astronomer Herman Zanstra (1894–1972) then showed that the hydrogen emission lines of planetary nebulae are produced when the free electrons liberated by the ultraviolet

Table 13.2 Physical properties of planetary nebulae

N = number density = N_e = electron density = (0.5–20) × 10^9 m^{-3}
T_e = electron temperature ≈ (0.6–1.8) × 10^4 K
R = radius = 0.2–0.8 light-years = 0.07–0.25 pc ≈ (2–7) × 10^{15} m
M = mass = $4\pi R^3 N m_P/3 = 10^{21}$ kg = 5 × 10^{-10} $M_\odot$
V_{exp} = expansion velocity = (1–9) × 10^4 m s^{-1}
τ_{exp} = expansion age ≈ 16,000 years

photoionization recombine with protons to make hydrogen atoms, cascading through the atoms' various allowed electron orbits or energy levels and radiating the Balmer emission line (Menzel 1926; Zanstra 1927, 1928).

In a brilliant piece of detective work, the American astronomer Ira S. Bowen (1898–1973) interpreted the two strong green emission lines as forbidden transitions of doubly ionized oxygen (Bowen 1928). His solution depended on the rarity of atomic collisions in the extremely tenuous planetary nebulae, which allows the occurrence of "forbidden" transitions. They are not actually forbidden but rather so improbable that they seldom take place in a higher-density laboratory situation, where an atom almost always is jostled by collisions into a different state before the forbidden radiation can be emitted.

The observed emission lines indicate gas temperatures of about 10,000 K and electron or ion number densities of about 10 billion per cubic meter, denoted as 10^{10} m^{-3} (Table 13.2). Although this is a big number, such densities are lower than the best vacuum used in a terrestrial laboratory. Kaler (1985) has provided a review of planetary nebulae and their central stars.

The mass density and temperature of planetary nebulae resemble those of the emission nebulae (Sect. 11.1); however, the planetary nebulae are about 10 times smaller and, unlike the emission nebulae, they also are expanding. Both types of nebulae are illuminated by a bright central star, and they both emit similar spectral lines: those of ionized hydrogen and the forbidden emission lines of oxygen and nitrogen ions, designated [O III], [O II], and [N II] (see previous Table 11.2).

As Zanstra realized, the intensity of the hydrogen emission line can be related to the temperature of the exciting star through the theory of the hydrogen atom and the Planck spectrum of thermal radiation. He found that these stars are enormously hot, and modern investigations show that they are the hottest stars known. The luminous central star radiates thousands of times more energy than the Sun and has a temperature of 100,000 K and even over 200,000 K, much higher than any main-sequence star. This places the central star of a planetary nebula right off the scales of the Hertzsprung – Russell diagram, on the far left side (Fig. 13.2). Powerful winds have removed the star's relatively cool, outer layers to reveal its hot interior.

Most of the radiation of such a hot star is at ultraviolet wavelengths, which brighten the surrounding nebula, but the star is relatively dim at the longer visible wavelengths and may even become invisible. However, as a young planetary nebula is blown outward by powerful winds, it slowly grows in size, thins out, and becomes transparent, revealing its source – the exposed core of a dying red giant.

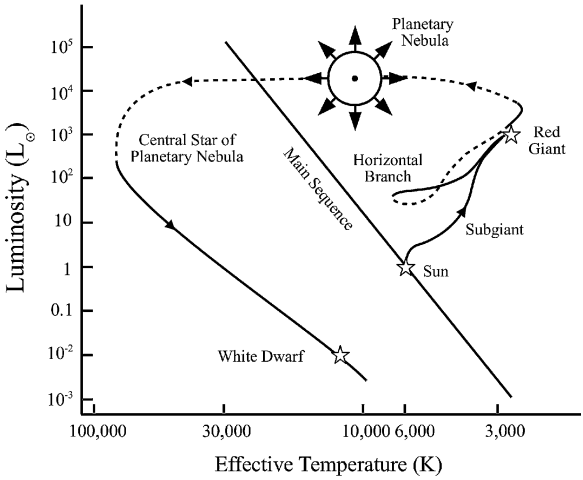


Fig. 13.2 Formation of a planetary nebula and white dwarf star The evolutionary track of a dying Sun-like star in the Hertzsprung-Russell diagram. When the star has exhausted its nuclear hydrogen fuel, which makes the star shine, it expands into a red giant star; after a relatively short time, the giant star ejects its outer layers to form a planetary nebula. The ejected gas exposes a hot stellar core, which collapses to form an Earth-sized white dwarf star that gradually cools into dark invisibility. The luminosity is in units of the Sun's luminosity, denoted $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and the effective temperature of the stellar disk is in units of degrees kelvin, denoted K. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

When modern telescopes are used to zoom in and resolve the expanding gas and dust, they show that it has not been expelled in a single puff of stellar wind, but instead in multiple gusts that can slam into each other (Fig. 13.3). Fast and slow winds may also play a role in producing the various shapes and forms of planetary nebulae. Balick and Frank (2002) have reviewed the shapes and shaping of planetary nebulae.

The observed expansion speeds of about 10 km s^{-1} and nebular dimensions of about a light-year across indicate the expanding shells of gas were ejected about 16 thousand years before the expansion and size were measured. Their luminescent gas will expand and disperse into interstellar space, cooling into invisibility and becoming indistinguishable from their surroundings in about 20,000 years. This is a relatively brief existence, only about 1 millionth of the stellar lifetime of many billions of years. As a result, planetary nebulae are much less numerous than the stars.

Example: Expansion age of a planetary nebula
The named planetary nebulae have expansion velocities of at least $V_{exp} = 10$ km s^{-1}, and a radius, R, of about 5×10^{15} m or half a light-year. The expansion time $\tau_{exp} = R/V_{exp} \approx 5 \times 10^{11}$ s $\approx$ 16,000 years, where 1 year $= 3.1557 \times 10^7$ s.

Fig. 13.3 The Eskimo Nebula About 10,000 years ago, a dying Sun-like star began flinging material into nearby space, producing this planetary nebula that is formally designated as NGC 2392. When first observed more than two centuries ago, it was dubbed the "Eskimo" Nebula because it resembled a face surrounded by a fur parka like those worn by Eskimos. It is located about 5,000 light-years from the Earth. This detailed image, obtained by instruments aboard the *Hubble Space Telescope*, reveals several episodes of ejection from the central star, including an outer ring of objects that are shaped like teardrops pointing outward and elongated, filamentary bubbles, each about 1 light-year in diameter. Dense material enveloping the star's equator has blocked ejected material, and intense winds moving at about 420 km s^{-1} have swept material above and below the equatorial regions. The bright central region contains another wind-blown bubble. (Courtesy of NASA/Andrew Frucher/ERO Team, Slyvia Baggett/STScI/Richard Hook, ST-ECF, and Zolan Levay/STScI.)

Despite their infrequent appearance on cosmic time-scales, thousands of planetary nebulae are known. Some of them are listed in Table 13.3 with the names associated with them, and their number in the New General Catalogue (NGC). Table 13.3 also provides the celestial position, distance, radius, expansion

Table 13.3 Bright named planetary nebulae[a]

Catalog designation	Popular name[c] (Nebula)	RA (2000)		Dec. (2000)		D[b] (ly)	R (ly)	V_{exp} (km s^{-1})	m_V[d]	T[d] (K)
		h	m	°	′					
NGC 650-1	Little Dumbbell	01	42.4	+51	34.5	2,400	0.85	39	17.5	175,400
NGC 2392	Eskimo	07	29.2	+20	54.7	3,000	0.18	53	10.5	65,000
NGC 3242	Eye of Jupiter	10	24.8	−18	38.5	1,600	0.09	30	12.1	90,000
NGC 3587	Owl	11	14.8	+55	01.0	2,000	0.98	30	16.0	112,000
NGC 3918	Blue Planetary	11	50.3	−57	10.9	3,260	0.16	25	13.2	–
IC 3568	Lemon Slice	12	33.0	+82	34.0	4,500	0.4	–	–	–
MyCn18	Hourglass	13	39.6	−67	22.9	10,100	0.33	10	14.4[e]	–
Menzel 3	Ant	16	17.2	−51	59.2	4,140	0.26	–	17.6[e]	–
M 2-9	Butterfly	17	05.6	+10	08.6	5,542	0.20	31	15.7	–
Hen 3-1357	Stingray	17	16.4	−59	29.6	18,000	0.08	–	15.0	–
NGC 6369	Little Ghost	17	29.3	−23	45.6	2,000	0.15	41	15.9	58,000
NGC 6543	Cat's Eye	17	58.6	+66	38.0	3,000	0.20	20	11.0	50,000
NGC 6720	Ring	18	53.6	+33	01.8	2,300	0.47	30	15.7	150,000
NGC 6751	Dandelion	19	05.9	−05	59.6	≤8,000	≤0.43	40	13.9	76,000
NGC 6826	Blinking Eye	19	44.8	+50	31.5	3,600	0.25	11	10.7	47,000
NGC 6853	Dumbbell	19	59.5	+22	43.3	1,200	1.0	28	14.0	160,000
NGC 7009	Saturn	21	04.2	−11	21.8	2,000	0.38	20	13.0	90,000
NGC 7027		21	07.0	+42	14.2	2,900	0.13	18	16.3	185,000
NGC 7293	Helix	22	29.6	−20	50.2	715	2.0	25	13.4[e]	110,000
NGC 7662	Blue Snowball	23	25.9	+42	32.1	2,500	0.20	30	13.2	110,000

[a] The distance, D, and radius, R, are in units of light-years, abbreviated ly. For conversion use 1 kiloparsec = 1 kpc = 3,260 light-years and 1 light year = 9.46×10^{15} m. The nebula's angular diameter $= 2R/D$ radians, where 1 radian = 2.06265×10^5 s of arc. Data courtesy of James B. Kaler, University of Illinois

[b] The distances of some planetary nebulae are not well known

[c] Other names: Little Dumbell Nebula = M 76, Barbell Nebula or Cork Nebula; Eskimo Nebula = Clownface Nebula; Eye of Jupiter Nebula = Ghost of Jupiter Nebula or Eye Nebula; Owl Nebula = M97; Blue Planetary Nebula = The Southerner Nebula; M 2-9 = Minkowski 2-9 or Butterfly Nebula or Twin Jet Nebula; Ring Nebula = M 57; Dandelion Nebula = Dandelion Puff Ball Nebula; Dumbbell Nebula = M 27; Blue Snowball Nebula = Snowball Nebula

[d] The estimated apparent visual magnitude, m_V, and temperature, T, of the central star

[e] Apparent blue magnitudes

velocity, and apparent visual magnitude and temperature of the central star for the planetary nebulae. The distances are often uncertain, and the radii are between 0.1 and 3 light-years, but it is the popular names that are so fascinating. They describe the resemblance of planetary nebulae to everyday objects, such as the twin weights of a dumbbell, a fur parka enveloping an Eskimo's face, the head of an owl, a cat's eye or simply a ring.

Planetary nebulae are produced by the winds of dying stars. Cassinelli (1979) has provided a review of stellar winds; Kudritzki and Puls (2000) have reviewed winds from hot stars; Willson (2000) has discussed mass loss from cool stars with their impact on the evolution of stars and stellar populations; and Dupree (1986) has discussed mass loss from cool stars.

Stars that produce the planetary nebulae end up as small, dense white dwarf stars. In the very distant future, our Sun will become one, as will the majority of other stars. Their discovery was entirely unexpected.

13.3 Stars the Size of the Earth

13.3.1 The Discovery of White Dwarf Stars

The first white dwarf to be known is a companion of a much brighter star, 40 Eridani, also known as Omicron Eridani from its Greek letter designation. The fainter star is designated 40 Eridani B to distinguish it from the brighter member, A, of the pair. The American astronomer Walter S. Adams (1876–1956) first drew attention to the A0 spectral type of 40 Eridani B, which suggested a disk temperature of about 10,000 K, and noticed that it was surprising that such a hot star should exhibit such a very low luminosity (Adams 1914).

Unfortunately, 40 Eridani B is so far away from its bright companion, with an orbital period of at least 7,300 years, that its mass could not be inferred from its orbital motion. However, this was not the case for Sirius B, the second white dwarf to be discovered. It is also a member of a binary star system, with a luminous companion designated Sirius A, the brightest star in the night sky.

The irregular motion of Sirius A first suggested the presence of its dim companion. As discovered by Friedrich Wilhelm Bessel (1784–1846), the bright star swerves from side to side of a straight-line trajectory, and this swerving motion was attributed to the gravitational attraction of a nearby, unseen companion (Bessel 1844). The American astronomer and telescope maker Alvan Clark (1804–1887) first detected the diminutive star in 1862.

The masses of Sirius A and B have been estimated from the orbital motion of the pair, weighing in at roughly the mass of the Sun; because the companion was about twice as far as Sirius from their common center of mass, the companion had to have about half the mass of the bright star. Modern determinations indicate that the mass of Sirius A is 2.02 solar masses, and that the mass of Sirius B is 0.978 times that of the Sun.

Adams showed that this low-luminosity star has a spectral class A0, like 40 Eridani B (Adams 1915). What Adams did not point out explicitly was that the high surface temperature in combination with the low total luminosity meant that both 40 Eridani B and Sirius B must be very small – only about the size of the Earth. Furthermore, the rather ordinary mass of Sirius B meant that the average mass density of the star must be enormous – about 10^9 kg m^{-3} and about 1 million times the average mass density of the Sun. Later on, it was found that the white dwarf 40 Eridani B has it own companion, designated 40 Eridani C, whose orbital motion allowed a mass determination of about half a solar mass for this white dwarf star, with a mass density comparable to that of Sirius B.

These stars now are known to be the inner, collapsed leftovers of dying red giant stars, exposed by the planetary nebulae that carried off the outer stellar atmospheres.

13.3.2 Unveiling White Dwarf Stars

Having depleted the hydrogen in their cores, the central regions of solar-mass stars contract to become hot enough to fuse helium into carbon and oxygen; however, there is not enough mass to generate a temperature hot enough to fuse carbon into neon, at about 1 billion K, and an inert carbon–oxygen core is surrounded by an inner helium-burning shell and an outer hydrogen-burning shell.

After shedding most of its outer material to form a planetary nebula, these giant stars leave behind a hot core of carbon and oxygen. Because it cannot generate additional heat by nuclear reactions, the core collapses to form a white dwarf with about the Sun's mass compressed to 1/100th of its former size and about the same radius as the Earth. Such a collapsed star is called a *white dwarf star* because initially it is white in color and it is relatively small for a star.

The concept of a white dwarf star as the exposed carbon–oxygen core of a former giant star initially was difficult to reconcile with the fact that the emitted light of the first white dwarfs contained strong spectral lines of hydrogen and it did not include the oxygen lines found in the surrounding planetary nebula. The French astronomer Evry Schatzman (1920−2010) explained this paradox when he noted that the hydrogen resides in a thin outer atmosphere and that the heavy elements such as carbon and oxygen had to sink out of sight into the dense stellar interior (Schatzman 1945). These elements are drawn by gravity into the unseen interior of the white dwarf and remain hidden by a hydrogen-rich layer that is only about 1/10,000th of the star's mass.

An inert, nonburning star, with no nuclear reactions in its core, which is composed mainly of carbon and oxygen nuclei, represents the final destiny, the end state, of main-sequence stars with a mass of between 0.5 and 8 solar masses, and it accounts for most of the observed white dwarf stars. Main-sequence stars of lower mass, below 0.5 solar masses, bypass the giant stage – never becoming hot enough inside to fuse helium – and they collapse directly into white dwarfs composed of

helium. However, these stars have such a low luminosity and temperature that they burn hydrogen slowly and have not yet exhausted their internal supply, even during the 14-billion-year age of the observable universe.

When first exposed, the white dwarf can be very hot because it was previously in the hot interior of a giant star. It may have an initial temperature of 200,000 K due to its hot origin and the collapse that created it, but because there is no thermonuclear fuel, there is nothing left to heat a white dwarf star.

A white dwarf's temperature will drop from an initial high of up to 200,000 K to an observed low that is colder than the Sun, steadily becoming fainter and dimmer. Liebert (1980) has reviewed white dwarf stars, Hansen and Liebert (2003) have provided a review of cool white dwarf stars, and D'Antona and Mazzitelli (1990) have reviewed the cooling of white dwarfs. This cooling and dimming is so slow that the oldest white dwarf has not yet chilled to the point of invisibility. That is, observations indicate that the oldest white dwarf stars, which are remnants of the earliest stars, have not yet had time to cool to the lowest possible luminosity that might be observed. The amount of time that the oldest white dwarf has cooled is estimated to be about 9 billion years, which when combined with its former lifetime as a main−sequence and giant star gives a rough estimate to the age of the observable universe of about 14 billion years (Winget et al. 1987; Wood 1992).

13.3.3 The High Mass Density of White Dwarf Stars

The rather ordinary stellar mass of a white dwarf has been compressed within a star that is comparable to the size of the Earth, which means that it has an enormous mean mass density up to 1 million times that of the Sun or about a billion, or 10^9, kg m^{-3} (Focus 13.1).

Focus 13.1 Radius and mass density of a white dwarf star
The effective disk temperature of Sirius B is $T_{eff} = 25{,}200$ K, while its absolute luminosity, L, is $L = 0.026\ L_{\odot}$, where the Sun's absolute luminosity is $L_{\odot} = 3.828 \times 10^{26}$ J s^{-1}. From the Stefan−Boltzmann law,

$$L = 4\pi\sigma R^2 T_{eff}^4, \tag{13.1}$$

where $\pi \approx 3.1416$ and the Stefan−Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}, we obtain an expression for the radius, R:

$$R = \left[\frac{L}{4\pi\sigma T_{eff}^4}\right]^{1/2}, \tag{13.2}$$

which gives a radius $R \approx 5.9 \times 10^6$ m $\approx 0.01\ R_\odot$, for the radius of Sirius B, where the solar radius $R_\odot = 6.955 \times 10^8$ m. By way of comparison, the radius of the Earth is $R_E = 6.378 \times 10^6$ m.

The mass, M, of Sirius B is $M = 0.98\ M_\odot$, where the Sun mass $M_\odot = 1.989 \times 10^{30}$ kg. The mass density, ρ, of Sirius B is therefore $\rho = 3\ M/(4\pi R^3) \approx 2.26 \times 10^9$ kg m^{-3}.

Arthur Stanley Eddington (1882–1944) pointed out that there is nothing inherently absurd about the high mass densities of white dwarf stars (Eddington 1924). Because all of the electrons are stripped away from their atomic nuclei in the hot stellar interiors, the free electrons can be packed closely with the bare nuclei, within the former space of the empty atoms.

Eddington also predicted that the gravitational redshift of Sirius B might be observed once allowance was made for the orbital motion of the double-star system (Eddington 1924). The gravitational redshift is the Doppler shift of a spectral line caused by the loss of energy in overcoming the gravity of the emitting object (Sect. 6.5). Adams apparently confirmed the effect, providing an independent confirmation of the small size of white dwarf stars that had been inferred from their high temperature and low luminosity (Adams 1915). This result was substantiated with greater clarity for Sirius B, 40 Eridani B, and other white dwarf stars using ground-based telescopes (Greenstein et al. 1971; Greenstein and Trimble 1972; Shipman 1972; Wegner 1980; Shipman et al. 1997; Huber et al. 1998), as well as the *Hubble Space Telescope* (Barstow et al. 2005).

The expression for the change in wavelength, $\Delta\lambda$, for a line emitted at wavelength λ_e can be derived from Newton's theory of gravity by assuming that the radiation photons have an energy $h\nu$ at frequency ν, and that this energy can be expressed in terms of an imaginary radiation mass, m, times the square of the speed of light, c, or that $h\nu = mc^2$, where the Planck constant $h = 6.626 \times 10^{-34}$ J s and $c = 2.9979 \times 10^8$ m s^{-1}. When the photon leaves the surface of a mass, M, with radius, R, the photon loses the energy, ΔE, given by:

$$\Delta E = h\Delta\nu = \frac{GMm}{R} = \frac{GMh\nu}{Rc^2}, \tag{13.3}$$

where, $\Delta\nu$ is the change in frequency ν. The gravitational redshift, z_g, caused by this loss of photon energy is given by

$$\begin{aligned} z_g &= \frac{V_r}{c} = \frac{\Delta\nu}{\nu} = \frac{\nu_L - \nu_{observed}}{\nu_{observed}} = \frac{\Delta\lambda}{\lambda_L} = \frac{\lambda_L - \lambda_{observed}}{\lambda_L} = \frac{GM}{Rc^2} \\ &= 2.12 \times 10^{-6}\left(\frac{M}{M_\odot}\right)\left(\frac{R_\odot}{R}\right), \end{aligned} \tag{13.4}$$

where V_r is the radial velocity corresponding to the gravitational redshift, ν_L and $\nu_{observed}$ respectively denote the emitted line frequency and observed line

frequency, λ_L and $\lambda_{observed}$ respectively denote the emitted line wavelength and observed line wavelength, $\Delta\lambda$ is the change in wavelength caused by overcoming the gravity, and the gravitational constant $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}.

Example: Gravitational redshift and thermal velocity of Sirius B
Precise measurements indicate that the radius of Sirius B is $R = 0.0084$ $R_\odot = 5.84 \times 10^6$ m, for a solar radius $R_\odot = 6.955 \times 10^8$ m, and that its mass is $M = 0.978$ $M_\odot$ (Barstow et al. 2005). This indicates a gravitational redshift of $z_g = GM/(Rc^2) \approx 2.5 \times 10^{-4}$, where the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2} and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. This redshift corresponds to a radial velocity $V_r = z_g c \approx 7.5 \times 10^4$ m s^{-1} $\approx$ 75 km s^{-1}, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The measured gravitational redshift of Sirius B is 80.42 ± 4.83 km s^{-1}, which is consistent with the estimate given the uncertainty of the measurement.

The bolometric luminosity of Sirius B is $L = 0.0026$ $L_\odot = 9.95 \times 10^{23}$ J s^{-1}, where the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. Measurements indicate an effective temperature of $T_{eff} = 25{,}200$ K for Sirius B. Using the star's luminosity and radius with the Stefan–Boltzmann law $L = 4\pi\sigma R^2 T_{eff}^4$, where $\pi = 3.1416$ and the Stefan–Boltzmann constant $\sigma = 5.670 \times 10^{-8}$ J m^{-2} K^{-4} s^{-1}, we obtain $T_{eff} \approx 14{,}000$ K. Using $T_{eff} = 25{,}200$ K in the expression for the thermal velocity $V_{thermal} = (3kT/m)^{1/2}$, where the Boltzmann constant $k = 1.381 \times 10^{-23}$ J K^{-1} and a hydrogen atom of mass $m = 1.67 \times 10^{-27}$ kg, we have $V_{thermal} = 2.5 \times 10^4$ m s^{-1} = 25 km s^{-1}, about three times smaller than the gravitational redshift.

For a white dwarf star, R $\approx$ 0.01 R$_\odot$ and $z_g = V_r/c \approx 0.02$ with $V_r \approx$60 km s^{-1}. The observed gravitational redshifts for Sirius B and 40 Eridani B, in radial velocity units, are 80.42 ± 4.83 km s^{-1} and 23.9 ± 1.3 km s^{-1}, respectively. Observations of these gravitational redshifts confirm the small size of white dwarf stars that had been inferred from their high temperature and low luminosity.

The small size and high mass density of the white dwarf stars have also been substantiated by measurements of their magnetic-field strength. During gravitational collapse, magnetic flux is conserved, and the surface magnetic-field strength increases as the surface area decreases. For a sphere of surface magnetic field strength, B, and radius R, the product BR^2 is a constant, and in solar units:

$$B = B_\odot \left(\frac{R_\odot}{R}\right)^2. \tag{13.5}$$

Table 13.4 Physical properties of white dwarf stars

M_{WD} = mass of white dwarf star $\approx$ 0.6 $M_\odot$ $\approx$ 1.2 $\times$ 10^{30} kg
M_{CWD} = critical upper mass limit for white dwarf star = 1.4 $M_\odot$ $\approx$ 2.3 $\times$ 10^{30} kg
R_{WD} = mean radius of white dwarf stars = 0.01 $R_\odot$ $\approx$ 6 $\times$ 10^6 m $\approx$ R_E = radius of Earth
ρ_{WD} = mass density of white dwarf star $\approx$ 10^9 kg m^{-3}
V_{esc} = escape velocity of white dwarf star = $(2GM_{WD}/R_{WD})^{1/2}$ = 9,000 km s^{-1} = 0.03 c.
L_{WD} = absolute luminosity of white dwarf star $\approx$ 10^{-3} $L_\odot$ $\approx$ 3.8 $\times$ 10^{23} J s^{-1}
T_{WD} = effective temperature of white dwarf star's visible disk = 4 $\times$ 10^3 K to 7 $\times$ 10^4 K
B_{WD} = surface magnetic field strength of white dwarf star = 10^2 tesla to 10^4 tesla
z_g = gravitational redshift of white dwarf star = $GM_{WD}/(R_{WD}c^2)$ $\approx$ 0.02 = V_r/c (or radial velocity $V_r = z_g c \approx$ 60 km s^{-1}, where c is the speed of light)

The dipolar magnetic field of the solar disk is $B_\odot \approx$ 0.01 tesla. For a white dwarf of radius R of 1/100th the radius of the Sun, or $R = R_{wd} \approx 0.01\ R_\odot$, the surface magnetic field strength will be amplified to $B_{wd} \approx$ 100 tesla. Surface magnetic field strengths of 100 to 10,000 tesla were inferred from the circularly polarized light of white dwarf stars, and roughly 10 % of them have magnetic fields in excess of 100 tesla (Kemp et al. 1970). Angel (1978) has reviewed magnetic white dwarfs. Individual papers on magnetism in white dwarfs are referenced in Lang (1999).

The physical properties of white dwarf stars are given in Table 13.4.

13.4 The Degenerate Electron Gas

13.4.1 Nuclei Pull a White Dwarf Together as Electrons Support It

The matter deep inside a white dwarf star is completely ionized and composed of equal numbers of atomic nuclei and electrons. Because most white dwarfs are the crushed remnants of red giant stars, which previously fused helium into carbon, their collapsed cores consist mainly of carbon nuclei and electrons. Stars that are somewhat more massive than the Sun leave behind white dwarfs containing oxygen nuclei. Therefore, white dwarf stars contain various amounts of carbon nuclei or oxygen nuclei, depending on the star, and it is these nuclei that supply the mass and gravity of a white dwarf star.

The nuclei supply the mass and gravity of a white dwarf star, with a mass density, ρ, given by:

$$\rho = \left(\frac{A}{Z}\right) m_p N_e, \tag{13.6}$$

for a white dwarf composed of elements of atomic number Z, the number of protons, and atomic mass number A. The mass number of helium is $A = 4$, that of carbon is $A = 12$ and the mass number of oxygen is $A = 16$, the number of protons and neutrons. Equivalently, $A/Z = N_e/N_p = \mu_e$, the ratio of electron number to proton number, since the number of electrons in an atom is equal to the number of nucleons or protons plus neutrons. For helium, carbon and oxygen $A/Z = 2$ or $Z/A = 0.5$. We also have $A/Z = \rho/(m_p N_e)$. In this expression, m_p is the proton mass, or $m_p = 1.67262 \times 10^{-27}$ kg, and N_e is the electron number density, which is related to the mass density by:

$$N_e = ZN_i = \frac{Z\rho}{Am_P} \approx \frac{0.5\rho}{m_p}, \tag{13.7}$$

where N_i is the ion number density.

Why does the core collapse stop at the white dwarf stage? In other words, what is holding up the star? There are no internal nuclear reactions to provide energy, generate heat, and create pressure to oppose gravity. As the white dwarf radiates away the heat left over from its former life in a red giant star, it eventually might cool down to a temperature of absolute zero, and there would be no motion or thermal energy left to support the white dwarf star.

When the material cooled enough, it could be expected that the electrons would return to their former orbits around the nuclei, making larger atoms, which would force the nuclei apart and make the white dwarf expand in size. However, the stars have no internal energy to push against gravitation and accomplish this feat; therefore, it seemed that a white dwarf star could not become that cold.

This paradox was not resolved until the development of quantum mechanics and the realization that it is the electronic properties of the crushed matter that hold up a white dwarf. The high-speed motions of the densely packed electrons, rather than the nuclei, produce an outward pressure that holds the gravitational forces at bay. These motions are not due to the star's internal temperature or heat; in fact, the internal pressure of a white dwarf star is unaffected by temperature.

The quantum–mechanical description of a very dense, crushed state of matter is statistical, and it is related to the quantum numbers that specify the state of subatomic particles. That is, the properties of the particles, such as energy or location, can take on only specific quantized values and no others. This situation is related to two principles that govern the quantum state of the very small: (1) The *uncertainty principle* that states that at any given time we cannot know exactly both where a particle is and where it is going; and (2) the *exclusion principle* that forbids the existence of two or more particles in exactly the same quantum state.

The German physicist Werner Heisenberg (1901–1976) first stated the uncertainty principle (Heisenberg 1927). It states that the more we know about the location of a subatomic particle, the less we know about its momentum and velocity, and vice versa. In an alternative interpretation, the more we know about the energy of a subatomic particle, the less we know about when it had that energy,

and vice versa. Mathematically, the product of the uncertainties, in location and momentum or energy and time, equals the Planck constant h.

The Austrian physicist Wolfgang Pauli (1900–1958) proposed the exclusion principle (Pauli 1927), which states that two identical subatomic particles cannot occupy the same quantum state at the same time. It applies to electrons, protons, and neutrons, and it dictates how electrons behave in an atom, occupying their various energy levels. Each electron in an atom has its own space, which prevents the electrons from either joining together in the same location or falling into their atomic nucleus. The exclusion principle ensures the very existence of atoms.

It also means that the free electrons in a white dwarf star, which are not attached to atoms, cannot be in precisely the same place at the same time. They instead resist being squeezed into one another's territory, darting away at high-speeds just to keep their own space. This provides the pressure of the crushed state of matter, which is caused by the electrons' resistance to crowding.

The Italian physicist Enrico Fermi (1901–1954) first worked out the statistical description of a large number of identical subatomic particles based on the exclusion principle (Fermi 1926, 1928). He specified the conditions in which all of the particles have the least possible energy without violating Pauli's exclusion principle, which specifies that these particles cannot all occupy the very lowest energy, called the *ground state*, at the same time. In Fermi's solution, all of the states up to a certain limiting energy can be occupied, and all of those above that energy are not occupied.

Under such conditions the collection of subatomic particles, or gas, is said to be *degenerate*. In mathematics, a degenerate case is a limiting one in which a class of objects changes its nature so as to belong to another, usually simpler, class; for example, a point is a degenerate circle.

As shown by Ralph H. Fowler (1889–1944), who was Eddington's colleague at Cambridge University, it is the degenerate pressure of the electron gas that supports a white dwarf star. Fowler applied Fermi's statistical description to such a dense star, showing that its electrons are completely degenerate. They produce an outward push, known as *degeneracy pressure*, to keep their own space and support the star (Fowler 1926). This kind of pressure is proportional to the $5/3$ power of mass density, and accordingly increases rapidly with it.

Not only is the electron degeneracy pressure strong enough to withstand the crushing gravity of a white dwarf star; it is also independent of the temperature of the electrons and involves no nuclear reactions. Because the pressure does not depend on temperature, it will persist even if the star cools to absolute zero, without the electrons ever rejoining the nuclei. As Fowler showed, the individual electrons in the crushed matter, even at a temperature of absolute zero still would have a kinetic energy comparable to the thermal energy of particles in an expanded gas with a temperature as large as 10 million K.

Because the equation of state of the degenerate electron gas is unaffected by temperature, any heating by hypothetical nuclear reactions will increase the temperature and rates of those reactions. The temperature would continue increasing until the star exploded, so we conclude that white dwarf stars do not

shine by nuclear reactions. Their light must come from the slow leakage of the heat contained in the nondegenerate nuclei. Eventually, the white dwarf star will fade into a gigantic black molecule, a frozen star.

When the electrons in a white dwarf star are moving at non-relativistic speeds, considerably less than the speed of light, the degenerate electron pressure, P_e, is given by:

$$P_e = \left(\frac{3}{8\pi}\right)^{2/3} \frac{h^2}{5m_e} N_e^{5/3}. \tag{13.8}$$

where the Planck constant $h = 6.626 \times 10^{-34}$ J s, the electron mass $m_e = 9.109 \times 10^{-31}$ kg, and N_e is the electron density. Notice that the degenerate electron pressure does not depend on the temperature. An equivalent expression for the equation of state of non-relativistic degenerate electron pressure is (Fowler 1926).

$$P_e \approx \left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{20 m_e m_p^{5/3}} \left(\frac{Z}{A}\right)^{5/3} \rho^{5/3}, \tag{13.9}$$

for a fully ionized gas composed of elements of atomic number Z and atomic mass number A. The proton mass $m_P = 1.67262 \times 10^{-27}$ kg. For a white dwarf star, which is composed of helium, carbon and oxygen with $Z/A = 0.5$, and:

$$P_e \approx 3.074 \times 10^6 \rho^{5/3} \text{ Pa for } \rho \ll 10^{10} \text{ kg m}^{-3} \tag{13.10}$$

Example: Gas pressure, degenerate electron pressure, and magnetic pressure in a white dwarf

A white dwarf star has a mass density of $\rho = 10^9$ kg m^{-3} and an initial temperature of $T = 10^7$ K. The gas pressure, P_G, is given by:

$$P_G = N_i kT = \frac{\rho}{A m_p} kT, \tag{13.11}$$

where N_i is the ion number density, the mass number $A = 4$ for helium, $A = 12$ for carbon and $A = 16$ for oxygen, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, and the proton mass is $m_p = 1.67262 \times 10^{-27}$ kg. Using these numbers for A = 4, the gas pressure is $P_G \approx 2 \times 10^{19}$ Pa. The non-degenerate electron gas pressure is $P_e = 3.074 \times 10^6 \ \rho^{5/3} \approx 3 \times 10^{21}$ Pa. So even on formation, before a white dwarf star has cooled, the degenerate electron pressure is about one hundred times greater than the gas pressure.

Since the white dwarf has no nuclear fusion reactions to supply heat, it gradually cools. The effective temperature of Sirius B, for example, is now 2.5×10^4 K, so its gas pressure is about 5×10^{16} Pa.

Since magnetic flux is conserved in gravitational collapse, the magnetic field strength B is amplified by the inverse square of the radius, to as much as $B \approx 10^4$ tesla for a white dwarf star which is 100 times smaller than the Sun. The magnetic pressure, $P_B = B^2/(2\mu_0) \approx 4 \times 10^{13}$ Pa, is much smaller than the degenerate electron gas pressure, where the permeability of free space is $\mu_0 = 1.2566 \times 10^{-6}$ N A^{-2}.

By 1929, the Estonian astrophysicist Wilhelm Anderson (1880–1940) demonstrated that the electrons in highly compressed, degenerate matter begin to attain velocities on the order of the speed of light, and that in this case the variation of the electron mass with velocity must be taken into account by using the equations of *Special Relativity*. For a relativistic degenerate electron gas the equation of state is (Anderson 1929; Stoner 1930):

$$P_e = \left(\frac{3}{8\pi}\right)^{1/3} \frac{hc}{4m_p^{4/3}} \left(\frac{Z}{A}\right)^{4/3} \rho^{4/3}, \tag{13.12}$$

which differs from the non-relativistic case in that the electron mass does not appear and the pressure varies as the $4/3$ power of the mass density, ρ, rather than the $5/3$ power for a non-relativistic degenerate electron gas. Using $Z/A = 0.5$ and evaluating $h = 6.626 \times 10^{-34}$ J s^{-1}, $c = 2.9979 \times 10^8$ m s^{-1}, and $m_p = 1.67262 \times 10^{-27}$ kg, we obtain

$$P_e \approx 0.49 \times 10^{10} \rho^{4/3} \text{ Pa} \quad \text{for } \rho \gg 10^{10} \text{ kg m}^{-3}, \tag{13.13}$$

for the relativistic, degenerate electron gas. The non-relativistic and relativistic degenerate electron pressures become equal at a mass density of $\rho \approx 4 \times 10^9$ kg m^{-3}.

13.4.2 Radius and Mass of a White Dwarf

A higher-mass white dwarf will be squeezed into a smaller space by its gravity, so the star's radius decreases with increasing mass. For a large enough mass, we might imagine that the star's radius would become very small, perhaps even shrinking to almost zero; however, this is preposterous and there must be a limit to the mass.

An approximate expression for the radius, R_{WD}, of a non-relativistic white dwarf star of mass, M, is:

$$R_{WD} \approx \frac{h^2}{20 m_e m_p^{5/3} G} \left(\frac{Z}{A}\right)^{5/3} M^{-1/3} \approx 4.82 \times 10^{16} M^{-1/3} \text{ m} \tag{13.14}$$

where h is the Planck constant, m_e is the electron mass, m_p is the proton mass, $Z/A = 0.5$ and the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. This means that

$$R_{WD} \approx 3.8 \times 10^6 (M_\odot / M)^{1/3} \text{ m}, \tag{13.15}$$

where the solar mass $M_\odot = 1.989 \times 10^{30}$ kg; see Provencal (1998) for observational tests of the mass-radius relation of white dwarf stars.

As the white dwarf shrinks in size, its mass and mass density become higher as does the degeneracy pressure of its electron gas. Under extreme compression, however, the average speed of the electrons increases and eventually approaches the speed of light. This means that there is an upper limit to the mass that can be supported by their pressure. This makes common sense, since the electrons will move at greater speeds with increasing stellar mass and density, although they cannot move faster than the speed of light.

The limiting mass for a white dwarf star is determined under high-speed, relativistic conditions, when the electrons approach the speed of light. As both the German-Estonian astrophysicist Wilhelm Anderson (1880–1940) and the English physicist Edmund C. Stoner (1899–1968) demonstrated, a star can contract only until the gravitational potential energy becomes insufficient to balance the increase in the kinetic energy of the electrons, which occurs for stellar masses of about 1 solar mass (Anderson 1929; Stoner 1929, 1930).

The two forms of energy are roughly equal at:

$$P_e \left(\frac{4\pi R_{WD}^3}{3} \right) \approx \frac{G M_c^2}{R_{WD}}, \tag{13.16}$$

or at a critical mass M_c given by

$$M_c \approx \left[\frac{4\pi P_e R_{WD}^4}{3G} \right]^{1/2}, \tag{13.17}$$

where P_e is the relativistic electron pressure, R_{WD} is the radius of the white dwarf star, and the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. Assuming a mass density of $\rho = 10^9$ kg and $R_{WD} = 6 \times 10^6$ m, we obtain $M_c \approx 10^{30}$ kg $\approx M_\odot = 1.989 \times 10^{30}$ kg.

Thus, for stellar masses larger than about 1 solar mass, there can be no equilibrium white dwarf configurations. More massive stars collapse under their own weight to form a neutron star or a black hole at the endpoints of stellar evolution.

During his voyage from India to Cambridge University for his graduate studies, the Indian astrophysicist Subrahmanyan Chandrasekhar (1910–1995) derived the detailed equation of state of a degenerate electron gas in the extreme relativistic limit (Chandrasekhar 1931). The exact solution for the critical mass is given by

$$M_c = 0.21\left(\frac{Z}{A}\right)^2\left(\frac{hc}{Gm_p^2}\right)^{3/2} m_p\,. \tag{13.18}$$

or for a white dwarf star with $Z/A = 0.5$,

$$M_C \approx 1.46\, M_\odot, \tag{13.19}$$

where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg.

Although both Anderson and Stoner previously called attention to the existence of this upper mass limit, it is known now as the *Chandrasekhar limit*, because he was the first to derive the detailed equilibrium conditions in which degenerate electron gases support a dense star's gravity.

In 1983, the Nobel Prize in Physics was awarded equally to Chandrasekhar, for his theoretical studies of the physical processes of importance to the structure and evolution of the stars, and to William A. "Willy" Fowler (1911–1995) for his theoretical and experimental studies of the nuclear reactions of importance in the formation of the chemical elements in the universe.

13.5 Exploding Stars

13.5.1 Guest Stars, the Novae

For at least 2,000 years, astronomers, hunters, mariners, and others familiar with the brightest stars must have been amazed by a *nova*, or "new star", that would appear suddenly at a place in the sky where no star previously had been seen. For a few days, the nova might be among the brightest stars in the dark night sky. But then the star would begin to fade away, and in about a month it would disappear back into invisibility, without a trace. The Chinese called them "guest stars" or "visiting stars" because they were not permanent members of the celestial sphere, instead appearing suddenly and then departing abruptly, like uninvited guests.

Every 20 years or so, a nova is luminous enough and close enough to be conspicuous without the aid of a telescope, attracting the attention of both astronomers and the superstitious. Like good wine, they are specified by the year of their occurrence, and their location is specified by the constellation in which they appear. Nova Aquilae 1918 was the brightest of the twentieth century, at apparent visual magnitude -1.1; Nova Herculis 1934 has historical importance; and Nova Cygni 1992 was the brightest nova in recent history.

Something of substance had to be at a nova's location before it appeared, to supply the energy of its outburst. By the mid-twentieth century, it was realized that this "mysterious something" was an inconspicuous star that previously had been recorded during systematic surveys of the dark night sky, using large telescopes and photographic exposures to record the dim light of faint stars.

Table 13.5 Physical properties of some novae[a]

Star Name	Year	m_{max}	L_{max} ($L_\odot$)	D (ly)	P_{orb} (hours)	M_1 ($M_\odot$)	M_2 ($M_\odot$)	V_{exp} (km s^{-1})
Classical novae								
GK Persei	1901	+0.2	$10^{5.3}$	1,500	47.92	0.9	0.25	1,200
V603 Aquilae	1918	−1.4	$10^{5.6}$	800	3.31	0.66	0.2	265
DQ Herculis	1934	+1.4	$10^{3.9}$	316	4.65	0.62	0.44	315
V1974 Cygni	1992	+4.4	$10^{4.9}$	10,430	19.53	0.83	–	–
Dwarf nova								
SS Cygni	b	8.3	≈10	≈541	6.603	0.60	0.40	–
Recurrent nova								
RS Ophiuchi	c	4.5	0.1	≥2,000	455.7 d	1.4	–	–

[a] Maximum visual magnitude, m_{max}, and maximum luminosity L_{max} in units of the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, distance in light-years, abbreviated ly, orbital period, P_{orb}, white dwarf mass, M_1, and companion mass, M_2 in units of the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, and expansion velocity V_{exp}

[b] The dwarf nova SS Cygni undergoes frequent and regular outbursts every 7–8 weeks, with an apparent visual magnitude of $m_{min} = 12.2$ at minimum and $m_{max} = 8.3$ at maximum. More than 800 outbursts have been observed since its discovery in 1896

[c] The recurrent nova RS Ophiuchi erupted in 1898, 1907, 1933, 1945, 1958, 1967, 1985 and 2006. It is a binary system with a red giant star in a 455.7-day orbit around a white dwarf star of mass near the Chandrasekhar limit

So, the bright, short-lived novae are neither new nor temporary but instead existing stars that suddenly increase in brightness by as much as 100,000 times, returning to their original states after several months or a few years.

Telescopic observations also have enabled the detection of faint dwarf novae. They brighten repeatedly, on a time-scale from days to decades, although by a smaller amount and with lower luminosity than the classical novae whose explosive outbursts are visible without the aid of a telescope. More than 900 outbursts of the dwarf nova SS Cygni, for example, have been observed since its discovery in 1896. It varies from apparent visual magnitude of 12.2 at minimum to 8.3 at maximum, every 7–8 weeks.

The properties of some of the classical novae, visible by the unaided eye, are listed in Table 13.5 with the dwarf nova SS Cygni.

13.5.2 What Makes a Nova Happen?

A major new understanding of novae occurred in the 1950s and 1960s when a few American astronomers began to examine the total light and spectra of ex-novae, long after the intense light of the nova outburst had faded to a relatively weak level. It then was discovered that a nova is not one star but rather two stars very close together. Twenty years after the 1934 eruption of Nova Herculis, for example, Merle F. Walker (1926–) found that this nova is an eclipsing binary

system with a remarkably brief orbital period of only 4.6 h (Walker 1954). The shortness of the period indicated that the two stars are very close together, practically touching one another. Nearly a decade later, Walker was able to show that Nova T Aurigae 1891 also is an eclipsing binary system with a short period, of 4.8 h (Walker 1963).

Alfred H. Joy (1882–1973) had examined the absorption and emission lines of the dwarf nova SS Cygni, identifying it as a binary-star system with a short orbital period of 6.6 h (Joy 1956). The emission lines originated in a blue-white dwarf star, whereas the absorption lines came from a red main-sequence star, the size of which was estimated to be roughly half the distance between the two stars. They were so close to one another that mass could spill from the red star into the blue star, and the nova process might be related to this mass flow.

In the meantime, Robert P. Kraft (1927–) demonstrated that membership in a short-period binary system is a necessary condition for a star to become a nova of either the classical or dwarf type (Kraft 1964). One of the stellar pair was usually a blue-white dwarf star; the other red component was usually a cool main-sequence star of spectral type G, K, or M or one of the same spectral type that is aging and expanding into a red giant. The short orbital period indicated that the two stars are so close that hydrogen flows from the red companion onto the white dwarf, reviving the "dead" star and giving it a brief new life in a cataclysmic nuclear explosion. The term *cataclysmic variable star* is used now to designate such close binary star systems, in which one of the components – conventionally called the *primary star* – is a white dwarf that accretes matter from its secondary companion. The category includes classical, dwarf, and recurrent novae.

Example: Dwarf nova SS Cygni
SS Cygni is a double star system with an orbital period of $P = 0.275$ days $= 6.603$ h $= 23{,}770$ s. It consists of a white dwarf of mass $M_1 = 0.60\ M_\odot$ and a main sequence star of mass $M_2 = 0.40\ M_\odot$, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. The linear separation, a, of the two stars can be inferred from Kepler's third law:

$$M_1 + M_2 = \frac{4\pi^2 a^3}{GP^2} \tag{13.20}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}. Using $M_1 + M_2 = M_\odot$ and solving for the linear separation we obtain $a \approx 1.24 \times 10^9$ m $\approx 1.78\ R_\odot$, where the Sun's radius $R_\odot = 6.955 \times 10^8$ m. The white dwarf star has a radius of about $0.01\ R_\odot$ and a low-mass main-sequence star of spectral class K5 might have a radius of about $0.8\ R_\odot$, so the two stars are practically touching each other. It therefore is not surprising that mass flows from the main-sequence star onto the white dwarf star, resulting in outbursts from SS Cygni every 7–8 weeks.

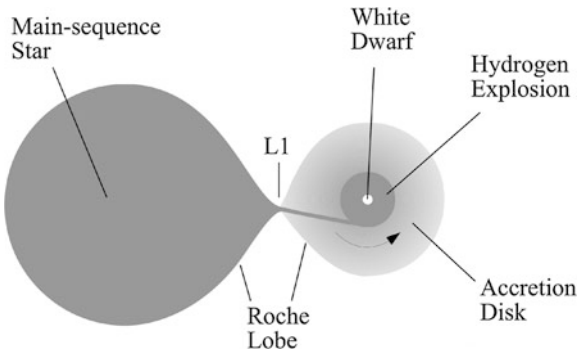


Fig. 13.4 Nova A classical nova is a thermonuclear explosion that occurs on the surface of a white dwarf star that is in a close orbit with a main-sequence star. The strong gravitational attraction of the white dwarf pulls its nearby companion into an elongated shape, the outer edge of which is designated the *Roche lobe*. Some of the hydrogen in the outer atmosphere of the main-sequence star spills over at the inner Lagrangian point, denoted L1, where the gravitational pull of the two stars is equal. This hydrogen spirals into a rotating accretion disk and down to the white dwarf, igniting an explosion, like a colossal hydrogen bomb. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

The immense gravity of the white dwarf star distorts the shape of its nearby companion, stretching it into an elongated configuration (Fig. 13.4). This is a tidal effect in which the side of the companion star that is nearest the white dwarf is pulled toward it, and the companion's center is pulled away from the side that is farthest from the white dwarf. Material that is pulled close enough to the white dwarf can pass outside the gravitational control of the red companion and go into orbit around the white dwarf, eventually spiraling into it.

The region in which the companion star retains gravitational control of its substance is known as its *Roche lobe*, named after the French astronomer Édouard A. Roche (1820–1883) who described it more than a century ago (Roche 1849/1850/1851). When the outer hydrogen atmosphere of the close companion star overflows its Roche lobe, it does not fall directly into the white dwarf star. In the absence of an intense magnetic field, the gas is accreted into an orbiting disk that resembles the proto-planetary disks that circle protostars (Sect. 12.3); the gas streams onto the magnetic poles when there is a strong magnetic field directing the flow.

In either case, the hydrogen is pulled slowly into the white dwarf, which compresses and heats the gas to very high temperatures. The accumulating hydrogen is a potential bomb that remains harmless until detonated.

A thin layer of hydrogen slowly builds up on the white dwarf as its companion keeps feeding matter into it. The pressure and temperature rise until hydrogen fusion suddenly is ignited, at about 10 million K, or 10^7 K. A runaway thermonuclear explosion then occurs, and a white dwarf that normally would cool and

fade away if left alone suddenly shines as brightly as 100,000 Suns (Starrfield et al. 1974, 1985). The explosion subsides after a few weeks, but the overflow of the companion continues. Gallagher and Starrfield (1978) have reviewed the theory and observation of classical novae.

The envelope of the white dwarf star is thrown off during the nova explosion, at high speeds of up to several thousand kilometers per second. However, despite the violence, the amount of material ejected is only about 0.000005 of a solar mass. The white dwarf therefore can retain its stability, and potentially generate additional novae as its companion continues to feed matter into it. An example of such a recurrent nova is RS Ophiuchi, which has exploded into a bright nova state at least six times between 1898 and 2006.

Eventually, the hydrogen may build up until it pushes the white dwarf above its limiting mass; the entire star then explodes, not just the thin outer atmosphere. This is a supernova that suddenly and unpredictably brightens with the light of 1 billion Suns.

13.5.3 A Rare and Violent End, the Supernovae

On rare occasions, an entire star is annihilated and suddenly becomes so bright that it can be seen easily in daylight rather than just at night like the novae. The Chinese emperor's astronomers in the Sung dynasty recorded one on July 4, 1054, near the constellation now known as Taurus, the Bull. The Chinese chronicles indicate that the new star initially was brighter than everything in the night sky except the full Moon; could be seen during the daytime for three weeks after its first appearance; and remained visible in the night sky for 22 months, without the aid of telescopes, which had not yet been invented. The "new" star of 1054 was definitely far brighter and longer lasting than any other guest star.

More than four centuries passed before other exceptionally brilliant guest stars were noticed, and this time they shook the very foundations of European thought. As Aristotle taught, heavenly bodies were supposed to be eternal, pure, changeless, incorruptible, and perfect, unlike anything on the Earth. Yet, in a span of just 32 years, two new daytime stars could be seen by almost anyone in the Earth's Northern Hemisphere who happened to look up. Each star remained fixed in the heavens for about a year, and then disappeared from view.

Both events were also discovered at a time before telescopes were invented. The Danish astronomer Tycho Brahe (1546–1601) witnessed the first visitor in 1572, which initially was brighter than the planet Venus. Perhaps because of the excitement caused by his discovery, Brahe built an observatory in which detailed measurements were made of stars and the planets. Johannes Kepler (1571–1630), who used Brahe's observations of planets to determine the laws of their motion, spied another bright new star as it lit up the heavens in 1604.

The exceptionally brilliant guest stars of 1054, 1572, and 1604 were much brighter and longer lasting than the conventional novae known at the time or

Table 13.6 Historical supernovae visible with the unaided eye[a]

Explosion Date	m_{max}	M_{max}	L_{max} ($L_\odot$)	Visible (months)	Type	Remnant name	D (ly)	θ (′)	R (ly)
SN 185	−8.0	−20.2	$10^{10.0}$	8	Ia	RCW 86	9,100	45	56
SN 386	+1.5	–	–	3	–	–	–	–	–
SN 393	−1.0	−11.0	$10^{6.3}$	8	II/Ib	*b*	3,000	70	30
SN 1006	−7.5	−19.2	$10^{9.6}$	21	Ia	PKS 1451−41	7,200	31	32
SN 1054	−6.0	−17.5	$10^{8.9}$	22	II	Crab nebula	6,500	7	6.6
SN 1181	−1.0	–	–	6		*c*	>8,000	–	–
SN 1572[d]	−4.0	−16.4	$10^{8.5}$	16	Ia	Tycho	11,500	8.3	14
SN 1604[e]	−2.5	−16.4	$10^{8.5}$	12	Ia	Kepler	20,000	3.2	9

[a] Maximum apparent visual magnitude, m_{max}, maximum absolute magnitude, M_{max}, maximum luminosity, L_{max}, in units of the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, length of visibility to the unaided eye, supernova type, supernova remnant name, distance D in light-years, abbreviated ly, angular diameter θ of supernova remnant in minutes of arc, denoted ′, and remnant radius, R, in light-years

[b] Supernova remnant RX J1713.7-3946

[c] Radio source 3C 38. The radio and x-ray pulsar J0205+6449 may not be associated with SN 1181

[d] Supernova explosion also known as Tycho's star, supernova remnant 3C 10

[e] Supernova explosion also known as Kepler's star, supernova remnant 3C 358

subsequently. They therefore have been dubbed *supernovae*, a term coined by the Swiss astronomer Fritz Zwicky (1898–1974).

Now that we know the distances, a supernova also refers to a stellar outburst the maximum luminosity of which exceeds by factors of several billions the luminosity of our Sun (Table 13.6). This is millions of times the peak luminosity of a classical nova. Moreover, unlike novae, the new breed of exploding stars has nothing conventional to return to after the explosion. They require so much energy that the mass of an entire star is annihilated.

In 1934, Walter Baade (1893–1960) and Fritz Zwicky communicated to the United States National Academy of Sciences a remarkable pair of papers on supernovae (Baade and Zwicky 1934a, b). In one paper, they showed that the enormous energy emitted in the supernova process corresponds to the total conversion of an appreciable fraction of a star's mass into energy. In the second paper, they predicted that a supernova explosion will accelerate charged particles to very high energies and that supernovae that occur only once in a millennium can account for the energetic cosmic-ray particles that now rain down on the Earth's atmosphere from all directions in outer space. In this more speculative paper, the two Caltech astronomers also said that the collapsing core of the explosion might become a neutron star, of very small radius and extremely high mass density. It took a half-century for astronomers to realize that Baade and Zwicky were correct on all counts.

The initial evidence for supernovae was extraordinarily sparse. In the Milky Way, they are seen at intervals of roughly 100 years, which is about 3,000 times less common than dwarf novae; so it might take centuries before the next supernova could be observed in the Milky Way. Fortunately, it was found that supernovae occur more frequently in the many spiral nebulae outside our Milky Way.

In the early decades of the twentieth century, astronomers discovered numerous faint novae in spiral nebulae, which suggested that the spirals were very distant if these "new" stars were like the classical novae seen in the Milky Way. At that distance, an exceptionally bright nova, observed in the nearest spiral nebula Andromeda in 1885, would have the luminosity of a supernova.

When the enormous distances to the spiral nebulae were confirmed, it was realized that they are not nebulae at all but instead galaxies that each contain about 100 billion stars (Sect. 14.2). Moreover, at maximum, a supernova briefly will outshine 1 billion stars in the same galaxy. Light from hundreds of supernovae in distant galaxies might be on its way to us now. Their light could take many thousands or even billions of years to travel to the Earth.

Because there are many of these extragalactic spirals, now called galaxies, Zwicky realized that a systematic photographic survey quickly would catch at least one star in the act of supernova explosion – and he was right. He detected the first one in 1937, with a camera attached to a modest telescope placed on the roof of a building at Caltech (Zwicky 1937). This supernova occurred in the spiral galaxy NGC 4157, now known to be about 55 million light-years away, so the actual explosion occurred 55 million years before Zwicky saw it, the time it took for its light to travel the vast distance separating the galaxy from us.

By observing the spectra and fading light of supernovae in distant galaxies, astronomers subsequently found that there are two methods for stars to come to such a violent end. The German-American astronomer Rudolph Minkowski (1895–1976), for example, divided the supernovae into two categories, denoted Type I and Type II, distinguished by the absence or presence of hydrogen in their spectra (Minkowski 1941). Type I was subsequently divided into three categories. Type Ia exhibits a strong absorption feature of singly ionized silicon at a wavelength of 615 nm in their spectra near peak light; Types Ib and Ic do not display this spectral feature. Modern identifying characteristics of supernovae of different types are given in Table 13.7, including the important Type Ia and Type II. Filippenko (1997) has reviewed the optical spectra of supernovae.

The peak light output from a Type I supernova typically is one or two orders of magnitude more luminous than that of the fainter Type II supernovae, which fade more slowly. We now know that the decay of radioactive elements produced during high-temperature Type II explosions heats the expanding gas and produces the optically visible light.

Walter Baade used Tycho's observations of the decaying light from the brilliant 1572 supernova to demonstrate that it is consistent with the superluminous emission of a supernova of Type I (Baade 1945; Van Den Bergh 1993); Kepler's meticulous observations of the 1604 event indicated a Type II light variation. Both

Table 13.7 Characteristics of supernova types[a]

Characteristic	Type Ia	Type Ib	Type II
Optical spectrum	No hydrogen Si II at 615.0 nm	No hydrogen He I at 587.6 nm	Hydrogen present at 656.3 nm
Maximum luminosity	$10^{9.8}\ L_\odot$	$\approx 10^{9.1}\ L_\odot$	$10^{9.1}\ L_\odot$
Ejection velocity	$\geq 10^4$ km s^{-1}	$\geq 10^4$ km s^{-1}	$\leq 10^4$ km s^{-1}
Ejected mass	$\approx 1\ M_\odot$	$\approx 1\ M_\odot$	$\approx 5\ M_\odot$
Progenitor star	White dwarf	Wolf-Rayet	Supergiant
Progenitor star mass	$1\ M_\odot$	4–7 $M_\odot$	$\geq 8\ M_\odot$

[a] Maximum luminosity in units of the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, and mass values in units of the Sun's mass $M_\odot = 3.854 \times 10^{30}$ kg

of these supernovae were observed before telescopes were invented and spectra were obtained from cosmic objects.

Although the two types of supernovae release comparable amounts of total energy during their explosion, there is a radical difference in the mass and kinetic energy of their ejected material. The expanding shells of Type Ia contain roughly 1 solar mass of material, whereas those of Type II events are about five times more massive. The large mass difference suggests that the progenitor stars of Type Ia supernovae are less massive than those of Type II. There is comparatively little discussion of the reasons for Type Ib and Ic supernovae in the scientific literature. Like supernovae of Type II, they are probably massive stars that have run out of nuclear fuel at their centers. Type Ib supernovae may be related to the core collapse of massive Wolf-Rayet stars that have lost hydrogen by strong winds.

13.5.4 Why do Supernova Explosions Occur?

Both types of supernovae involve the explosive conversion of a star's entire mass into energy but by different physical mechanisms. A Type Ia stellar explosion is due to external causes. It involves a white dwarf star pushed into nuclear explosion by too much mass overflow from a nearby companion. The other, Type II, supernova is an internal event, which occurs during the gravitational collapse of the iron core of a massive star that has depleted all of its energy. The nuclear supernovae of Type Ia and gravity-powered ones of Type II are believed to occur with about equal likelihood in the Milky Way, at the rate of 1 every 50–100 years.

Weiler and Sramek (1988) and Trimble (1982, 1983) have reviewed our knowledge of supernovae and supernova remnants. Woosley and Weaver (1986) have reviewed the physics of supernova explosions. Bethe (1990) has reviewed our understanding of supernova mechanisms.

Hillebrandt and Niemeyer (2000) review Type Ia supernova explosion models. Smartt (2009) has reviewed the progenitor stars of Type II supernovae. Heger et al.

(2003) describe how massive single stars end their lives. Langer (2012) has reviewed the pre-supernova evolution of both massive single and massive binary stars, and Paczynski (1971) has reviewed evolutionary processes in close binary systems. The next two subsections of the text provide a general description of the two types of supernova explosions.

13.5.5 When a Nearby Star Detonates Its Companion

Like the novae, there is one type of supernova that gets assistance from the outside, being pushed over the edge into explosion, shattering an entire star. Such a supernova – now known as Type Ia and characterized by the absence of emission from hydrogen – occurs in a close binary-star system, with a white dwarf star – the shrunken dense remnant of a former low-mass star – circling a main-sequence star.

The English astrophysicist Fred Hoyle (1915–2001) and his American colleague William A. "Willy" Fowler (1911–1995) introduced the detailed mechanisms for this type of supernova (Hoyle and Fowler 1960; Fowler and Hoyle 1964). When the nearby companion star expands as a result of its normal evolution, hydrogen from its outer atmosphere spills onto the white dwarf. The overflow, for example, might happen when the ordinary visible companion runs out of core hydrogen fuel and swells into a red giant star. As the hydrogen overflow continues, a steady increase in the mass of the white dwarf will compress and heat the star. As the increasing mass approaches the upper mass limit for a white dwarf star at 1.46 solar masses, the rise in internal temperature ignites a nuclear explosion.

Because the white dwarf is supported against gravity by temperature-independent, degenerate electron pressure, adding heat to the star's interior increases the temperature but not its pressure; therefore, the white dwarf does not expand and cool in response. Instead, the increased temperature initiates the fusion of carbon nuclei in a runaway nuclear explosion that obliterates the star in a few seconds and releases about 10^{49} J in energy.

In other words, the added mass detonates a carbon bomb, triggering explosive nuclear reactions that quickly spread throughout the star and completely shatter it. The entire star explodes into a Type Ia supernova that shines with the light of billions of Suns, and there is nothing left.

Because every one of these explosions is triggered at the same mass limit, under similar conditions, and also because the star is completely destroyed, Type Ia supernovae are expected to produce about the same maximum light output every time they occur. The typical absolute visual magnitude is −19.3, or about 5 billion times the luminosity of the Sun, with little variation. Astronomers use this bright, uniform luminosity as a "standard candle" to measure the distances to their host galaxies located far beyond the Milky Way in the remote parts of the observable universe, thereby determining the pace of its expansion. They have shown that a

repulsive force, called dark energy, is making the universe expand at an accelerating rate (Sect. 15.6).

13.5.6 Stars that Blow Themselves Up

There is more than one way to explode a star, and some of the supernovae are gravity-powered, catastrophic outbursts from very old massive stars. This method of shattering a star applies to an isolated star with the right mass – between about 8 and 20 times the Sun's mass – that blows itself apart. This Type II supernova, which exhibits hydrogen in its spectra, follows the creation of an iron core within an evolving, massive supergiant star. Smartt (2009) has reviewed the progenitors of such core-collapse supernovae.

The material in the core of such a massive supergiant star is not degenerate, in the mathematical sense used for white dwarf stars; nuclear reactions proceed in the advanced burning stages at ever-increasing central temperatures until an iron core is produced and all of the available nuclear fuel has been exhausted. Deprived of these resources, the iron core collapses under its own weight into a neutron star or black hole in less than 1 s, and an explosion blows away the rest of the in-falling matter.

When the iron nuclei in the core of such a star are pushed together, no energy is released. The iron does not burn, regardless of how hot the star's core becomes. So, there is no longer any energy being generated to sustain the star's structure. Then, a massive star, having burned brightly for perhaps 10 million years, can no longer support its own crushing weight and the iron core collapses.

The central iron core can be crushed into a ball no bigger than New York City in less than 1 s, accruing energy from its in-fall. Electrons are squeezed inside the iron nuclei, combining with their protons to make neutrons. The material is compacted to nuclear density, and the center collapses to form a neutron star. If the collapsing core is more massive than about 3 solar masses, however, the collapse proceeds to the formation of a black hole.

Having lost the supporting core, the surrounding material first plunges in toward the center. When reaching mass densities approaching that of an atomic nucleus, the collapsing core bounces back and a powerful shock wave pushes out against the rest of the star. With the help of a dense shower of neutrinos produced in the collapsing core, the star's outer layers are torn apart and expelled into deep space at supersonic speeds. The doomed star suddenly increases in brightness 100 million fold, becoming a Type II supernova that briefly outshines up to 1 billion of its neighboring stars combined (Fig. 13.5).

This type of supernova hurls into surrounding space all of the elements synthesized inside the star and residing in shells surrounding the iron core before its collapse. During the high-temperature explosions, the supernova also produces vast amounts of other heavy elements, including radioactive elements such as uranium. Newly formed radioactive nickel, for example, eventually decays into

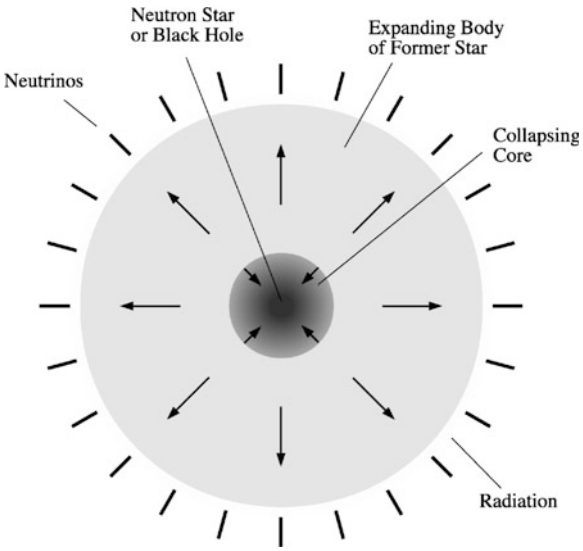


Fig. 13.5 Type II supernova In this type of supernova explosion, an isolated star blows up and its shattered remains are propelled into surrounding space. Radio and x-ray radiation from the expanding supernova remnant can be observed for thousands of years after the explosion. The core of the star is compressed by gravitational contraction into a neutron star or a black hole. Neutrinos emitted from the collapsing core remove most of the supernova energy and assist shock waves in pushing the stellar remains into an expanding remnant. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

iron, producing most of the iron now found in the universe. Such a supernova occurred within a nearby satellite galaxy of the Milky Way in 1987, which led to new insights about how these explosions occur.

13.5.7 Light of a Billion Suns, SN 1987A

For more than three and a half centuries, nobody was fortunate enough to see a supernova with the unaided eye. Then, late in the evening of February 24, 1987, astronomers discovered one (Fig. 13.6), which was designated SN 1987A (SN is for supernova; 1987A denotes the first one discovered that year). The star exploded 168,000 years before that night in the Large Magellanic Cloud, one of the two satellite galaxies of the Milky Way visible from the Earth's Southern Hemisphere. It generated an intense burst of light that peaked in visual brightness at the third magnitude.

Astronomers at the Las Campanas Observatory, a barren mountaintop near La Serena, Chile, were the first to notice the new star when Ian Shelton (1957–)

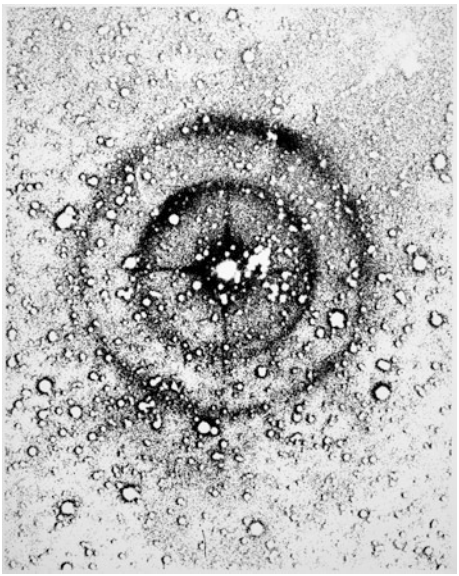


Fig. 13.6 Light echoes from SN 1987A Two complete rings of light surround the exploded star SN 1987A in this negative image taken with the 3.9 m (153.5 inch) Anglo-Australian Telescope on 15 July 1988. The initial flash of light from the supernova explosion has been reflected off clouds of interstellar dust and observed 14 months after the explosion was brightest, somewhat like an echo of sound. These light echoes arise in two thin sheets of microscopic dust grains located about 470 light-years (*inner ring*) and 1,300 light-years (*outer ring*) in front of the supernova. The rings have been made more prominent by photographically subtracting an image taken 3 years before the supernova exploded, canceling much that existed previously. Stars, however, are still visible as faint haloes. (Courtesy of David Malin and the Anglo-Australian Observatory.)

photographed it at 3 o'clock in the morning using a small 0.25 m (10 inch) telescope placed in an unheated shed. The discovery was relayed to the International Astronomical Union's clearinghouse for such events, which relayed the news to astronomers throughout the world. By the next evening, nearly all major radio and optical telescopes south of the Equator were observing the supernova. One month later, it was featured on the cover of *Time* magazine with just one word: "BANG!" Astronomers were still watching its expanding debris years after the exploding star hurled it into space.

McCray (1993), Arnett et al. (1989), and Trimble (1988) have provided us with reviews of Supernova 1987A.

SN 1987A was a gravity-powered, iron-catastrophe Type II supernova, generated during the core collapse of a former blue supergiant star that began life with a mass of about 20 times that of the Sun. The physical properties of this progenitor star and its subsequent supernova explosion are listed in Table 13.8.

Table 13.8 Supernova SN 1987A[a]

Progenitor star
Name: Sanduleak −69°202
Spectral Type: B3 Ia (blue supergiant)
Location: Large Magellanic cloud
Distance: 168,000 light-years
Radius: 3×10^{10} m $\approx 50\ R_\odot$
Effective temperature = 16,000 K
Luminosity = 4.6×10^{31} J s^{-1} $\approx 10^5\ L_\odot$
Mass $\approx 16\ M_\odot$
Neutrino burst
Number of neutrinos detected: 19 anti-neutrinos
Energy of each neutrino: $\approx$ 20 MeV = 3.2×10^{-12} J
Neutrino flux at Earth: 5×10^{14} m^{-2}
Number of neutrinos emitted: $\approx 10^{58}$ neutrinos
Energy released in neutrinos: $\approx 10^{48}$ J
Duration of neutrino burst: $\approx$ 10 s
Neutrino luminosity: $\approx 10^{48}$ J s^{-1} $\approx 10^{22}\ L_\odot$
Visible explosion
Peak visible luminosity: 3.8×10^{33} J s^{-1} $\approx 10^7\ L_\odot$
Velocity of ejected material: $\approx 10^7$ m s^{-1}
Mass of ejected material $\approx 8 \times 10^{30}$ kg $\approx 4\ M_\odot$
Kinetic energy of ejected material $\approx 4 \times 10^{44}$ J

[a] The symbols $R_\odot$, $L_\odot$ and $M_\odot$ respectively denote the radius, luminosity and mass of the Sun

One of the more interesting observations of SN 1987A was made from beneath the Earth's surface, when massive subterranean instruments detected a few neutrinos emitted during the explosion. Although solar neutrinos had been observed coming from nuclear reactions that power the Sun, all other stars are so far away and the number of their neutrinos striking the Earth is so low that they had never been detected coming from any other cosmic object. It had nevertheless been proposed that a supernova might generate a great number of neutrinos (Focus 13.2).

Focus 13.2 Neutrinos generated during a supernova
As first realized by George Gamow (1904–1968) and his Brazilian colleague Mario Schönberg (1914–1990), the temperature of a collapsing stellar core can become high enough to create both neutrinos and antineutrinos, which can easily escape and carry away prodigious amounts of energy (Gamow and Schönberg 1941). This would lead to further loss of supporting pressure, the implosion of the core with a rapid rise in temperature, and an explosion as a supernova. They named the nuclear transformation associated with neutrino energy loss the Urca process, after its similarity to the gambling operations at the Casino de Urca near Rio de Janeiro. There also, no matter how you played the game, you always seemed to lose.

Today we know that supernova neutrinos may be emitted during nuclear processes other than the Urca process, but the basic idea was correct. Vast numbers of neutrinos are released as the crushed iron core of a dying star is broken into its subatomic components. A neutrino is produced each time a proton and an electron combine, and pairs of neutrinos and antineutrinos are created when the core temperature reaches 100 billion K. All of these neutrinos are without electrical charge, have almost no mass, and move unimpeded at nearly the speed of light through nearly any amount of matter, even the entire Earth. But if enough cosmic neutrinos were directed at the Earth, massive, subterranean instruments might detect a small number of them.

For several seconds, the relatively nearby supernova explosion SN 1987A generated such enormous amounts of the elusive neutrinos that a small number were detected. They were recorded in two underground neutrino detectors 3 hours before the first visible sighting of SN 1987A.

Only 19 neutrinos were detected flashing through the underground darkness. Yet, even this small number signaled the presence of an awesome energy, vastly exceeding the amount contained in the radiation and expanding debris of the supernova (also see Table 13.8). They indicated that 10 billion trillion trillion trillion, or 10^{58}, neutrinos were produced by the exploding star. As the iron core imploded, the neutrinos carried away energy of about 5×10^{46} J, or roughly 99 % of the explosion energy, and emitted a total neutrino luminosity comparable to that of the optically visible luminous output of more than ten thousand billion billion stars like the Sun.

Example: Energy of a supernova
During the supernova explosion of SN 1987A, about a solar mass, or $1.0\ M_\odot = 1.989 \times 10^{30}$ kg was ejected with a velocity V of about 10^7 m s^{-1}. The kinetic energy of the ejected mass is equal to $0.5\ M_\odot V^2 \approx 10^{44}$ J. The binding energy released in forming a central neutron star of radius R_{NS} is equal to $E_G = GM_\odot^2/R_{NS} \approx 1.9 \times 10^{46}$ J, where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and $R_{NS} = 14$ km. The energy, E, released in completely destroying a solar-mass star is $E = M_\odot c^2 \approx 1.8 \times 10^{47}$ J, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The energy of each detected neutrino was 20 MeV $= 3.2 \times 10^{-12}$ J, and the neutrino flux at Earth was 5×10^{14} m^{-2}. The number of neutrinos emitted is equal to the product of this flux times the area at the Earth's distance, or $4\pi D^2$, where the distance $D = 168{,}000$ light-years $= 1.59 \times 10^{21}$ m and 1 light-year $= 9.46 \times 10^{15}$ m, so the total number of neutrinos emitted is $5 \times 10^{14} \times 4\pi D^2 = 1.58 \times 10^{58}$. The total energy emitted by the neutrinos, $E_{neutrino}$, is about $1.58 \times 10^{58} \times$ energy per neutrino $\approx 5 \times 10^{46}$ J.

The neutrinos detected from SN 1987A (Bionta et al. 1987; Hirata et al. 1987) solved one of the thornier problems in understanding such a supernova explosion. It was known that a stellar collapse generates tremendous amounts of energy, but there was difficulty explaining how that energy was transferred from the collapsing core into the outer layers of the star in sufficient amounts to produce an explosion. The core might rebound, sending shock waves propagating into the surrounding material, but computer simulations indicated that the shock waves could not blow away the rest of the star. They always became stalled when encountering the in-falling matter from the outer layers.

Unlike the stalled shock waves, the flood of escaping neutrinos carries tremendous amounts of energy far away from the stellar core, a very small fraction of which gets caught in the in-falling outer layers of the collapsing star, heating up the gas to a temperature of more than 10 billion K. This produces a buoyant, convecting bubble of energy that reverses the in-fall and powers the explosion.

Three hours after the initial collapse and generation of neutrinos in SN 1987A, its heated bubble expanded, driving shock waves before it, and burst through the surrounding material, breaking the star apart and hurling its pieces into space, which produced the dazzling light of the supernova. This explains why the neutrinos were detected 3 hours before any light was seen.

13.5.8 Will the Sun Explode?

There is no explosion forecast for the Sun's future. It is going out alone, passing into its final resting state unaccompanied by a close companion. Although it will end up as a dense, high-gravity white dwarf star, even the nearest star still will remain far beyond the dead Sun's gravitational embrace, never orbiting it. Moreover, our Sun is nowhere near massive enough to ever explode by itself. So, no nova or supernova is expected when the Sun runs out of nuclear fuel. It will go quietly into the oblivion of permanent night.

13.6 Expanding Stellar Remnants

In their explosive death, stars that go supernova blast their outer layers into surrounding space, expelling much or all of the stellar material at supersonic speeds of up to 30,000 km s^{-1}, or 1/10th of the speed of light. A strong shock wave forms ahead of the ejected material, colliding with the surrounding interstellar gas and heating it up to temperatures of tens of millions of K. The high-temperature material emits intense x-rays that have been observed with instruments aboard spacecraft located above the Earth's obscuring atmosphere, thereby recording the

debris of cataclysmic stellar explosions that occurred even thousands of years ago, before recorded history.

Like the explosions that cast this material out, there are two types of supernova remnants. They can be the remains of a white dwarf star sent into explosion by a nearby companion in a Type Ia supernova or the explosive debris of a single massive star that has expired in a Type II supernova. Weiler and Sramek (1988) and Trimble (1982, 1983) have reviewed our knowledge of both supernovae and supernova remnants.

The *Chandra X-ray Observatory*, for example, has imaged the *Tycho supernova remnant*, named for Tycho Brahe (1546–1601) who reported observing the original explosion in 1572. A Type Ia supernova formed this remnant, when a white dwarf was sent into annihilation by overflow from a nearby companion star. An arc of x-ray emission in the supernova remnant was attributed to material blown off the companion star, which otherwise survived the destruction of its neighbor and now is moving within the remnant more quickly than its neighbors as the result of the explosion. The properties of the arc and remaining star indicate that the former white dwarf star and its companion once orbited one another in a five-day period at a separation of less than 1/10th of the mean distance between the Earth and the Sun, or 0.1 AU (Lu et al. 2011).

Supernova remnants often emit intense radio radiation. The majority of these radio supernova remnants appear as bright rings, or shells, in projection against the sky. The intense radio radiation cannot be produced by a hot gas, like x-rays, but instead is emitted by electrons accelerated to high speeds by the supernova explosion and spiraling in a magnetic field.

A beautiful example of an expanding shell-like supernova remnant is the *Cassiopeia A supernova remnant*, abbreviated Cas A. It is located roughly 10,000 light-years away in the direction of the constellation Cassiopeia and is the brightest radio source in the sky. The expanding shell of Cas A is also a strong source of x-rays, emitted by a 50-million-K gas (Fig. 13.7).

Despite its radio and x-ray brilliance, the remnant is faint at optically visible wavelengths, which nevertheless indicate that it is rich in oxygen and now expanding at a speed of about 5,000 km s^{-1}. The Type II supernova explosion would have been observed as a daytime star around 1680, but there are no historical records of the event. Thick clouds of interstellar dust, or material ejected from the massive star's outer layers, may have absorbed the light and rendered the explosion optically invisible.

The most spectacular example of a Type II supernova remnant is the *Crab Nebula supernova remnant* (Fig. 13.8, Baade 1957), also one of the brightest radio sources in the sky. The Crab Nebula, also designated M 1, NGC 1952, or Taurus A, is the remnant of a supernova explosion that was observed in 1054. It is about 6,500 light-years away, has a diameter of 11 light-years, and expands at a speed of about 1,500 km s^{-1}. The physical properties of this fascinating object are given in Table 13.9.

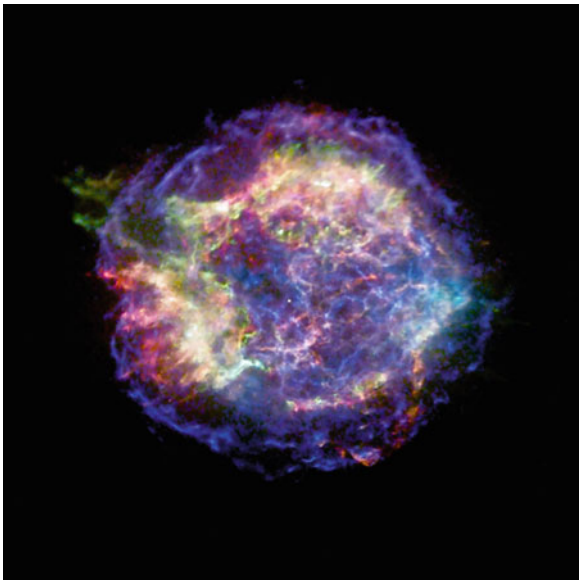


Fig. 13.7 X-ray image of Cassiopeia A supernova remnant The expanding supernova remnant Cassiopeia A has a temperature of about 50 million K and therefore is a luminous x-ray source, seen in this image from the *Chandra x-ray Observatory*. Still visible in x-rays, the tiny point-like source near the center of Cas A is a neutron star, the collapsed core of the star that exploded about 330 years ago, as observed from the Earth. (Courtesy of NASA/CXC/MIT/University of Mass. Amherst/M. S. Stage, et al.)

The English amateur astronomer John Bevis (1695–1771) discovered the nebula in 1731. The French astronomer Charles Messier (1730–1817) independently found it a few decades later (Messier 1781), listing it as the first entry in his famous catalogue. The Irish astronomer William Parsons (1800–1867), the third Earl of Rosse, gave it its present name, the Crab Nebula, after sketching it to resemble a crab (Rosse 1850).

The Crab is the first nebula to be associated with expanding material – by the American astronomer John C. Duncan (1882–1962); and the first to be recognized as the remnant of a stellar explosion – by the American astronomer Edwin Hubble (1889–1953), who identified the Crab Nebula with the guest star recorded by Chinese astronomers in 1054 A.D. in the same region of the sky (Duncan 1921; Hubble 1928). The Crab supernova remnant is also the first cosmic radio source (Bolton et al. 1949) and the first cosmic x-ray source (Bowyer et al. 1964) to be discovered outside the solar system; it was identified as one of the brightest, persistent emitters of gamma rays (Haymes et al. 1968). Hester (2008) has provided a review of the Crab Nebula, as an astrophysical chimera. Reynolds (2008) reviewed supernova remnants at high energy.

Photographs taken by Walter Baade using the Mt. Wilson 2.5 m (100 inch) telescope in the early 1940s indicated that the visible nebula consists of two

Fig. 13.8 The Crab Nebula supernova remnant The optically visible light of the Crab Nebula, designated as M 1 and NGC 1952, consists of two distinct parts: (1) A system of expanding filaments forms an outer envelope in which emission lines occur at well-defined wavelengths, and (2) an inner amorphous region that emits continuum radiation at all wavelengths. Walter Baade (1893−1960) took this photograph of the expanding filaments in the wavelength range of 640 nm to 670 nm using the 5.0 meter (200 inch) telescope on Mount Palomar, California in 1955. A Type II supernova explosion observed nearly 1,000 years ago, in 1054, ejected the filaments. The continuum glow that is concentrated in the inner parts of the nebula is the nonthermal radiation of high-speed electrons spiraling in magnetic fields (see Fig. 13.9). This continuum emission is powered by a spinning neutron star, the southwesternmost (*bottom right*) of the two central stars. The neutron star is the crushed, ultradense core of the exploded star. It also is a radio pulsar that acts like a lighthouse spinning 30 times a second. (Courtesy of Hale Observatories.)

distinct parts that emit radiation differently (Baade 1942). A tangled, oval-shaped network of red and green filaments, seen in the light of bright emission lines from ionized atoms, encases the inner blue and milk-white continuum radiation. The filamentary remnants of the explosion contain about 4 solar masses of material, consisting mostly of ionized helium and hydrogen along with lesser amounts of carbon, oxygen, nitrogen, iron, neon, and sulfur. The progenitor star that was annihilated to make the explosion had an estimated mass of between 9 and 11 solar masses, some of that mass was sent into invisibility within surrounding space and the rest remained within a central neutron star.

In addition to an expanding shell, the radio nebula has a filled center that coincides with the inner visible light. These types of supernova remnants are named *plerions*, from the Greek word *pleres* for “full” or “filled.”

The inner radiation contains practically all of the energy emitted by the Crab Nebula, and it is 1,000 times more intense at radio wavelengths than at optically visible wavelengths. It is impossible to reconcile the observed radio emission with the optical emission through the thermal radiation of a hot gas at any plausible temperature. As pointed out by the Russian astronomer Iosif Shklovskii (1916–1985), both the radio and optical emission of the Crab Nebula come from

Table 13.9 Physical properties of the Crab Nebula supernova remnant

Date of explosion = 1054 AD
M_S = progenitor star mass $\approx 9\ M_\odot$
M_{ejec} = ejected mass = 2–3 $M_\odot \approx 5 \times 10^{30}$ kg
M_N = neutron star mass = 1.4 $M_\odot$
D = distance $\approx$ 6,500 light-years $\approx$ 2.0 kpc $\approx 6.2 \times 10^{19}$ m
θ = angular extent = $4.5' \times 7.0'$
R = radius = $\theta D/2$ = 4.1 light-years × 6.1 light-years = 3.9×10^{16} m × 6.1×10^{16} m
V = velocity of ejected material $\approx$1,450 km $s^{-1} \approx 1.45 \times 10^6$ m s^{-1}
L = total luminosity = $10^{31.14}$ J s^{-1}
S = radio flux density = 1,040 Jy at 1 GHz
L_X = x-ray luminosity = $10^{30.38}$ J s^{-1}
N_e = electron density $\approx 4 \times 10^7$ m^{-3}
B = magnetic field strength $\approx 3 \times 10^{-8}$ tesla
M_{ns} = neutron star mass $\approx$ 1.4 $M_\odot \approx 2.8 \times 10^{30}$ kg
E_b = binding energy released in forming neutron star $\approx GM_\odot^2/R_{ns} \approx 10^{46}$ J
P = radio pulsar period = 0.033326 s
dP/dt = radio pulsar period time derivative = 421.288×10^{-15} s s^{-1}
$T = P/(2dP/dt)$ = approximate age of pulsar $\approx 3.95 \times 10^{10}$ s $\approx$ 1,250 year

synchrotron radiation emitted by high-energy electrons spiraling around magnetic fields at nearly the speed of light (Shklovskii 1953).

The Crab Nebula was the first supernova remnant known to emit nonthermal synchrotron radiation, although the radio emission of the Milky Way had been attributed to the synchrotron radiation of cosmic ray electrons spiraling about the interstellar magnetic field (Sect. 11.3). Electrons of extremely high energy emit optically visible light, whereas electrons of slightly lower energy radiate at radio wavelengths. Because the more energetic electrons lose their energy faster and also radiate at shorter wavelengths, the synchrotron-radiation mechanism provides a natural explanation for the nonthermal spectrum of the Crab's radiation, which is more intense at longer wavelengths. At every wavelength of observation, from x-rays to radio waves, the bulk of radiation from the Crab Nebula is accounted for by the synchrotron-radiation mechanism. It explains the nonthermal spectrum of the optical and radio emission, as well as the polarization of the optical continuum (Fig. 13.9, Baade 1957). The polarized radiation has a preferred orientation or direction due to the high-speed electrons moving in a large-scale, well-ordered magnetic field.

Despite the successes of the synchrotron-radiation theory, explaining the origin of the energetic electrons that gave rise to the radiation remained a fundamental difficulty. The electrons radiating at optically visible wavelengths will dissipate their energy by synchrotron radiation in about 180 years, and the more energetic electrons that produce the short-wavelength x-rays should lose their energy and disappear in less than a year. Because the supernova radiation was emitted more than 900 years ago, the high-speed electrons producing the synchrotron radiation

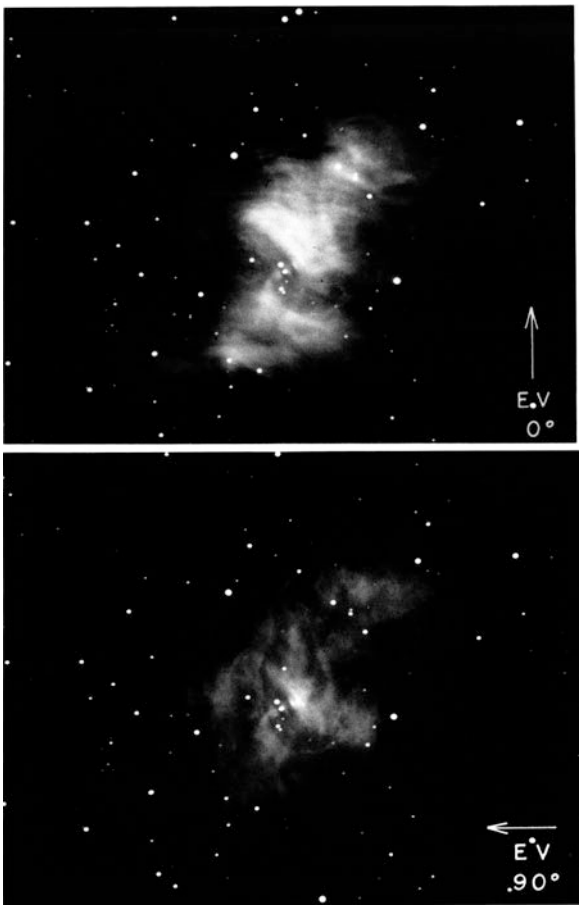


Fig. 13.9 Polarized light of the Crab Nebula The inner continuum emission of the Crab Nebula supernova remnant should be polarized, or emitted in a preferred direction, if it is caused by nonthermal synchrotron radiation. This was confirmed in detail by these visible-light photographs taken in the wavelength range of 540–640 nm through a polarized filter by Walter Baade (1893–1960) in 1955 using the 5.0 meter (200 inch) telescope on Mount Palomar, California. The arrows indicate the direction of the electric vector of the light recorded. The south westernmost of the two central stars is the remnant neutron star and a radio pulsar. This inner amorphous region of the Crab Nebula supernova remnant is a powerful source of radio radiation, while also emitting optically visible light and x-rays. This radiation is most intense at the longer, radio wavelengths, a characteristic of non-thermal radiation. It has been attributed to the synchrotron radiation of relativistic electrons whirling at nearly the speed of light around magnetic fields. Because the more energetic electrons lose their energy faster and also radiate at shorter wavelengths, the synchrotron radiation mechanism provides a natural explanation for the nonthermal emission of the Crab's inner regions and for the fact that its radio emission is a thousand times more intense than its visible light. (Courtesy of Hale Observatories.)

cannot be survivors of the original explosion. Instead, some unknown source must be continuously replenishing these energetic electrons.

Example: Expansion age, pulsar age, and x-ray synchrotron lifetime for the Crab Nebula

The material ejected from the explosion that gave rise to the Crab Nebula supernova remnant is now expanding at a velocity of $V = 1.45 \times 10^6$ m s^{-1}. The largest angular extent of the supernova remnant is $\theta = 7.0' = 420'' \approx 2.035 \times 10^{-3}$ radians, where 1 radian $= 2.063 \times 10^5\,''$. At the Crab Nebula's distance of $D = 6{,}500$ light-years $= 6.15 \times 10^{19}$ m, where 1 light-year $= 9.46 \times 10^{15}$ m, this angular extent corresponds to a radius of $R = \theta D/2 \approx 6.26 \times 10^{16}$ m. If the ejected material has been moving at a constant velocity across this radius, the expansion age is $T = R/V \approx 4.3 \times 10^{10}$ s $\approx$ 1,367 years, where 1 year $= 3.1557 \times 10^7$ s.

The supernova explosion observed by the Chinese in this region of the sky occurred in 1054 AD, or about 955 years ago, which is slightly younger and might indicate that the debris initially expanded at a faster rate than that observed today.

The Crab pulsar, described in the next section, has a period of $P = 0.033326$ s, and that period is increasing at the rate of $dP/dt = 421.288 \times 10^{-15}$ s s^{-1}. The approximate pulsar age is $T = P/(2dP/dt) = 3.95 \times 10^{10}$ s $= 1{,}250$ years. This upper limit to the age is consistent with the historical age of about 950 years, for the periods lengthen with time and the initial pulsar period might have been shorter. Setting $P = 0.025$ s for that initial period would give the correct age.

The Crab Nebula emits synchrotron radiation at radio, visible light, and x-ray wavelengths. It is emitted by energetic electrons spiraling in a magnetic field of strength $B \approx 3 \times 10^{-8}$ tesla. At a soft x-ray photon energy of $E = 10$ keV $= 1.602 \times 10^{-15}$ J, the frequency of the radiation is $\nu_s = E/h = 2.42 \times 10^{18}$ Hz, where the Planck constant $h = 6.626 \times 10^{-34}$ J s. The wavelength of the radiation is $\lambda_s = c/\nu_s \approx 1.23 \times 10^{-10}$ m, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The Lorentz factor γ of high-speed electrons giving rise to synchrotron radiation at this frequency or wavelength is given by $\gamma = ([\nu_s/(2.8 \times 10^{10}\ B)]^{1/2} \approx 5.34 \times 10^7$, and the synchrotron lifetime of the electron radiating this x-ray radiation is $\tau_s = 8.187/(B^2\gamma) \approx 1.7 \times 10^8$ s ≈ 5.4 years. More energetic x-rays correspond to a higher γ and a shorter lifetime for electrons emitting this radiation; synchrotron radiation emitted at gamma-ray wavelengths has even briefer electron lifetimes. The electrons emitting synchrotron x-rays and gamma rays from the Crab Nebula could not have been accelerated in the supernova explosion that occurred nearly 1,000 years ago.

An early clue to the energy source of the Crab Nebula was provided when the American astronomer Carl Lampland (1873–1951) observed moving wisps and knots that originated at the center of the nebula (Lampland 1921). Walter Baade subsequently showed that these features move at up to a tenth of the speed of light, suggesting that the central star was injecting high-speed particles into the nebula. It was eventually realized that a neutron star was born in the crushed center of the explosion.

The Italian astronomer Franco Pacini (1939–) pointed out that gravitational energy released during the collapse of a normal star into a neutron star will be converted into rotational energy (Pacini 1967). It follows from the conservation of angular momentum that the neutron star will be spinning rapidly, with a rotation period of less than a second, and it will also have an intense magnetic field, owing to the conservation of magnetic flux in collapse. Provided that the magnetic axis and rotation axis are not aligned, a rotating dipole magnetic field will convert the rotational energy into electromagnetic energy, thereby providing the luminosity of the Crab Nebula for its entire lifetime.

In the same year as Pacini's prescient publication, the first radio pulsar was discovered. The following year, pulsars were attributed to rapidly rotating neutron stars, and a pulsar was found at the center of the Crab Nebula, spinning 30 times a second. The Crab was the first object whose luminous output was related to a central pulsar.

The nebula is powered by the pulsar wind that is composed of charged particles accelerated to nearly the speed of light by the rapidly rotating, intense magnetic field of the spinning pulsar. Short-lived, flaring bursts of radiation from the Crab Nebula at gamma ray wavelengths may be due to sudden restructuring of the pulsar magnetic field, accelerating electrons to energies more than 100 times greater than can be attained by any particle accelerator on Earth.

Pulsar winds are found inside the shells of other filled supernova remnants, but Cassiopeia A is an exception. Although the x-ray emission of a central neutron star has been located within this supernova remnant, using instruments aboard the *Chandra X-ray Observatory* (Lu et al. 2011), the neutron star is relatively quiet and does not emit any detectable pulsar wind activity. One possible explanation is that the magnetic fields are so extremely strong that they have stifled pulsar wind activity rather than enhanced it. We now turn to early speculations about the possible existence of neutron stars and the subsequent discovery of radio and x-ray pulsars.

13.7 Neutron Stars and Pulsars

13.7.1 Neutron Stars

Walter Baade (1893–1960) and Fritz Zwicky (1898–1974) proposed the possibility that neutron stars might exist just two years after James Chadwick's (1891–1974) discovery of the neutron (Chadwick 1932a, b). They speculated that a supernova

Table 13.10 Physical properties of neutron stars

M_{NS} = mass of neutron star $\approx$ 2 $M_{\odot} = 3.978 \times 10^{30}$ kg
M_{CNS} = critical upper mass limit for neutron star $\approx$3 $M_{\odot} = 5.967 \times 10^{30}$ kg
R_{NS} = radius of neutron star $\approx$ 12 km = 1.2×10^{4} m
ρ_{NS} = mass density of neutron star $\approx 5 \times 10^{17}$ kg m^{-3}
V_{esc} = escape velocity of neutron star = $(2GM_{NS}/R_{NS})^{1/2}$ = 210,000 km s^{-1} = 0.70 c
E_{NS} = binding energy released to form a neutron star = $GM_{NS}^2/R_{NS} \approx 8.8 \times 10^{46}$ J
P = rotation period = 0.001–10 s
B = magnetic field strength = 10^8 tesla

explosion is driven by the gravitational energy released when a massive star runs out of fuel and its core collapses, but the explosion may not completely destroy the stellar core. A dense cinder could remain at the center or, in their prescient words: "With all reserve we advance the view that a super-nova represents the transition of an ordinary star into a *neutron star*, consisting mainly of neutrons. Such a star may possess a very small radius and an extremely high density." (Baade and Zwicky 1934b).

We now know that if a collapsing star is more massive than 1.4 solar masses, the inward force of its gravitation will overcome the outward degenerate electron pressure that halts the collapse at the white-dwarf stage. The young Russian physicist Lev Landau (1908–1968) speculated that dead stars with a mass above this white-dwarf limit might collapse until neutron degeneracy pressure would halt the crush of gravity (Landau 1938). Baym and Pethic (1979) provide an early review of the physics of neutron stars.

The electrons are pushed into direct contact with the atomic nuclei, and thus packed together at nuclear densities. The enormous pressure of the rapid collapse would create neutrons when the electrons merged with the nuclear protons, forming a nuclear gas with a mass density surpassing that of the Sun, and water, by a factor of a million, billion, at 5×10^{17} kg m^{-3} (Table 13.10). At these densities, a normal star like the Sun would have collapsed to a radius of only 12 km. Such a neutron star has a powerful gravity, with an escape velocity of about 0.7 or 70 % of the speed of light.

Within a year of Landau's speculations, George Gamow (1904–1968) pointed out that the high central mass densities of neutron stars can be expected only in those stages of stellar evolution subsequent to the exhaustion of the thermo-nuclear fuel in normal stars, and that only the very massive stars would have evolved rapidly enough to reach this stage during the lifetime of the universe (Gamow 1939).

At about the same time, the American physicist J. Robert Oppenheimer (1904–1967) and a Canadian graduate student George M. Volkoff (1914–2000) found that neutron stars have a limiting mass of their own (Oppenheimer and Volkoff 1939). They showed that at these large mass densities both the degenerate neutron pressure and the effects of gravitation on space–time must be considered. The equation of state of the nuclear material was obtained in the first

approximation by identifying it with a degenerate, relativistic gas of neutrons fulfilling Fermi statistics. The macroscopic structure of the star, its mass, radius, and density distributions, were determined from Einstein' *General Theory* of gravitation. Calculating equilibrium configurations along these lines, Oppenheimer and Volkoff found that a stable neutron star could exist only in a finite range of masses.

Just as degenerate electron pressure supports a white dwarf star, it is degenerate neutron pressure that supports a neutron star. This means that the radius, R_{NS}, of a neutron star of mass, M_{NS}, can be estimated by:

$$R_{NS} = \frac{m_e}{m_n} R_{WD} \approx 10^7 \left(\frac{m_e}{m_n}\right) \left(\frac{Z}{A}\right)^{5/3} \left(\frac{M_{NS}}{M_\odot}\right)^{-1/3} \text{m} \tag{13.21}$$

or

$$R_{NS} \approx 10^4 (M_\odot / M_{NS})^{1/3} \text{ m}, \tag{13.22}$$

where the electron mass $m_e = 9.1094 \times 10^{-31}$ kg, the neutron mass $m_n = 1.6749 \times 10^{-27}$ kg, R_{WD} denotes the radius of a white dwarf star, $Z/A = 1.0$ for a neutron star, and the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. So, for a neutron star of mass $M_{NS} = 1.4\ M_\odot$ and radius $R_{NS} = 10$ km, the star's mass density, $\rho_{NS} = 3\ M_{NS}/(4\pi R_{NS}^3) \approx 7 \times 10^{17}$ kg m^{-3}, which is comparable to the mass density of the nucleus of an atom.

Although the radius of a neutron star shrinks with increasing mass, you can't increase the mass of a neutron star without limit any more than you can for a white dwarf star, and the mass of a neutron star is now believed to be between 1.4 and 3.0 solar masses.

Theoretical considerations of a dense neutron gas suggest that it could be superfluid, with no resistance to flow, and superconducting, without electrical resistance. *Chandra* x-ray observations of the neutron star at the center of the Cassiopeia A supernova remnant indicate a rapid decline in the temperature of the ultra-dense neutron core, of about 4 % over a 10 year period, suggesting that it is made of superfluid and superconducting material (Page et al. 2011; Shternin et al. 2011).

To sum up, when an isolated massive star can no longer support its own crushing weight, the center collapses, and obtains energy from its in-fall. If the collapsing stellar core weighs between 1.4 and 3.0 solar masses, it is compacted to nuclear density, and forms a neutron star. When the core weighs more than 3 solar masses, it collapses into a black hole.

Although important in hindsight, these early considerations about the possibility of neutron stars did not evoke much interest at the time. They would have remained a speculative curiosity if it were not for the serendipitous discovery of radio pulsars.

13.7.2 Radio Pulsars from Isolated Neutron Stars

Before discussing the discovery of pulsars, we provide references to modern reviews for further reading. Lyne and Graham-Smith (2012) have discussed many aspects of pulsar astronomy, and Taylor and Stinebring (1986) provided an earlier review of our understanding of pulsars. Gaensler and Slane (2006) have reviewed the evolution and structure of pulsar wind nebulae; and Phinney and Kulkarni (1994) have reviewed binary and millisecond pulsars.

Pulsars were discovered accidentally during a survey of the *scintillations*, or "twinkling", caused when radio radiation from cosmic sources passes through the Sun's winds. When the radio waves are viewed through the wind-driven material, they blink on and off, varying on time-scales of a few tenths of a second – in much the same way that stars twinkle when seen through the Earth's varying atmosphere. Repeated observations of several scintillating radio sources at different angles in relation to the Sun provide information about the properties of the solar wind and the angular structure of the radio sources. The fluctuations are greatest for the smaller emitters, just as stars twinkle more than the Moon or planets, which have larger angular extents.

To study these effects, Antony Hewish (1924–) and his colleagues at Cambridge University built a large array of 2048 dipole antennas, spread over four and a half acres and operated at a long radio wavelength of 3.7 m, since the scintillating fluctuations were known to be more prominent at longer wavelengths. The combined signals from all the antennas were connected to a radio receiver and chart recorder with a time constant of 0.1 s, the time-scale of the scintillations. When examining the charts in July 1967, graduate student Jocelyn Bell (1943–) found a strong fluctuating signal in the middle of the night, when the array was pointed away from the Sun and the effects of the solar wind should have been small.

Further investigations led to the astonishing detection of periodic radio pulses, with an exceedingly precise repetition period of 1.3372795 s. The first radio pulsar had been detected (Hewish et al. 1968). No one had foreseen its existence, and no other known astronomical object kept time so accurately.

By the time the discovery was ready for publication, in 1968, evidence of other radio pulsars was found in the existing chart recordings; within three weeks, a second paper announced the discovery of three additional radio pulsars (Pilkington et al. 1968). This triggered searches for other previously unknown pulsars with large radio telescopes using rapid time sampling rather than the long integration times formerly used. In less than a year, the list of pulsars was expanded to more than two-dozen, and a pulsar was detected at the position of the very star thought to be the neutron star remnant of the Crab Nebula supernova explosion.

We now know that the term *pulsar* is misleading for the compact stars do not pulsate – they rotate – but the name has stuck. It designates repeating pulses of radio emission rather than a pulsating star.

The physical properties of radio pulsars are given in Table 13.11.

Table 13.11 Physical properties of radio pulsars

P_{RP} = period of radio pulsar = 0.001–4 s
dP_{RP}/dt = rate of increase of radio pulsar period = 10^{-15}–10^{-12} s s^{-1}
$T = P_{RP}/(2dP_{RP}/dt)$ = characteristic radio pulsar age = 10^{3}–10^{10} years (where 1 year = 3.156×10^{7} s)
B_{RP} = magnetic field strength of radio pulsar = 10^{9}–10^{13} G = 10^{5}–10^{9} tesla
L_{RP} = radio luminosity = 10^{18}–10^{24} J s^{-1}
M_{RP} = mass of binary pulsar PSR 1913 + 16 = 1.44 $M_{\odot} \approx 2.85 \times 10^{30}$ kg

The pulsars probably could have been discovered many years earlier, when other large radio antennae were constructed, but radio astronomers were used to adding up signals over long time intervals to detect faint cosmic radio signals. The long time resolutions precluded detection of the pulsars. They are relatively faint radio sources when averaged over their period because there is no emission between the brief radio pulses. If time resolutions comparable to the pulsar burst durations of milliseconds had been used, the intense radio bursts would have been detected easily. It is because Hewish specifically designed a new type of radio telescope for a study of the rapidly changing solar wind effects that the radio pulsars were discovered accidentally.

The extreme regularity of the periodic radio bursts suggested that they are controlled by the rotation of a massive body, and the short duration of the pulsar bursts suggested that their radiation originates in a body that cannot be much larger than the Earth. That is, the size should be smaller than the product of the speed of light and the burst duration; otherwise, it may be violating nature's upper speed limit, the speed of light. A rotating white dwarf was initially suspected, but it could not spin faster than about once per second. Rotation with a shorter period would tear the white dwarf apart, since its outer atmosphere would be rotating faster than the star's escape velocity.

The Austrian-born American astronomer Thomas Gold (1920–2004) proposed that radio pulses are produced by a rapidly rotating neutron star with an intense magnetic field (Gold 1968). He assumed that a pulsar would emit radio radiation in a beam, like a lighthouse, oriented along the magnetic axis (Fig. 13.10). An observer sees a pulse of radio radiation each time the rotating beam flicks across the Earth. Because the neutron star's beam could be oriented at any angle, the beams of many pulsars would miss the Earth and would remain forever unseen.

Gold suggested definitive observational tests of his ideas. He noticed that a spinning neutron star gradually loses its rotational energy and slows down, successfully predicting that this would cause a slow lengthening of the radio pulsar periods with time. He also predicted that radio pulsars with much shorter periods would be found, as they were.

If a slowly rotating star collapses down to a small size, the rate of rotation increases. It's a result of the conservation of angular momentum, which means that the period of rotation is proportional to the square of the radius. When an ordinary visible star runs out of nuclear fuel and is compressed to the size of a neutron star,

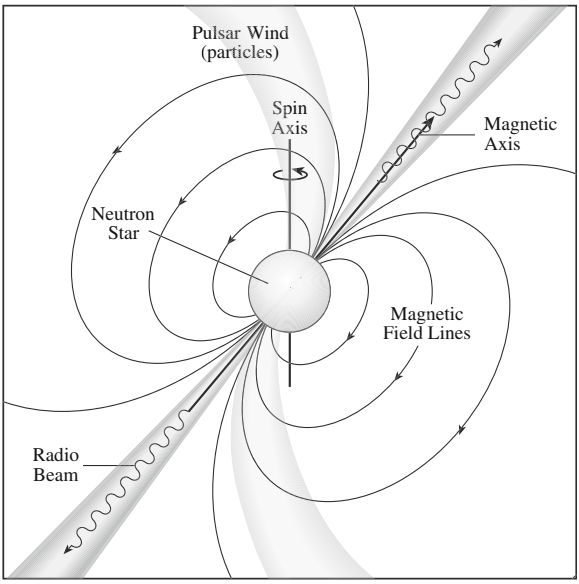


Fig. 13.10 Radio pulsar A spinning neutron star has a powerful magnetic field the axis of which intersects the north and south magnetic poles. The rotating fields generate strong electric currents and accelerate electrons, which emit an intense, narrow beam of radio radiation from each magnetic polar region. Because the magnetic-field axis can be inclined to the neutron star's rotation axis, these beams can wheel around the sky as the neutron star rotates. If one beam sweeps across the Earth, a bright pulse of radio emission, called a *pulsar*, is observed once every rotation of the neutron star

its rotation period would speed up from days to milliseconds. Since the time between successive radio pulses is the same as the rotation period of the neutron star, it has to be initially spinning at a very fast rate.

A neutron star is also a powerful magnet, and the magnetism is attached to the material within the star. When the star collapses it carries its magnetism with it, packing it into a smaller volume and amplifying its strength by factors of billions. The magnetic field strength increases in inverse proportion to the surface area, which is itself proportional to the square of the radius. So a typical surface magnetic field on a visible-light star, with a strength of 0.01 tesla, would be strengthened by a factor of a ten thousand million, or to 10^8 tesla, if collapsing to a neutron star.

Example: Period and magnetic field of a rotating neutron star or white dwarf star

We can infer the rotation period of a neutron star using the conservation of angular momentum. For a sphere of mass, M, radius, R, rotation period, P, and rotation velocity $V = 2\pi R/P$, the conservation of angular momentum in

gravitational collapse requires that $MVR = 2\pi MR^2/P =$ constant, or that the rotation period varies as R^2 if the mass is not changed. For the Sun, $M = M_\odot = 1.989 \times 10^{30}$ kg, $R = R_\odot = 6.955 \times 10^8$ m, $P = 25.67$ days at the solar equator, where one day = 86,400 s, and $V = 1{,}971$ m s^{-1} at the solar equator. If a star of this mass and rotation period collapsed to form a neutron star of radius $R_{NS} = 10$ km, the rotation period would be $P_{NS} \approx 4.6 \times 10^{-4}$ s.

Conservation of magnetic flux in gravitational collapse provides an estimate for the magnetic field of a neutron star. For a sphere with a surface magnetic field strength, B, and radius, R, the conservation of magnetic flux in gravitational collapse requires that the magnetic flux $BR^2 =$ constant, or that the magnetic field strength B varies as R^{-2}. For the Sun, $B = B_\odot = 10^{-2}$ tesla and $R = R_\odot = 6.955 \times 10^8$ m. If a star of this magnetic field strength and radius collapsed to a neutron star of radius $R_{NS} = 10$ km it would have a surface magnetic field strength of $B_{NS} = 4.8 \times 10^7$ tesla.

The same formulae apply for a white dwarf star, which would collapse to a radius of about 0.01 solar radii, to give a rotation period of $P_{WD} = 221$ s and a magnetic field strength of 100 tesla. Although such strong magnetic fields have been detected for white dwarfs, many white dwarfs rotate with periods of days or even years rather than seconds to minutes. The slow rotation of white dwarf stars may be attributed to the powerful winds that created their surrounding planetary nebulae, removing rotational energy and slowing the white dwarf.

The association of radio pulsars with neutron stars in supernova remnants became accepted when Australian radio astronomers found a pulsar with an extremely short period – 89 ms – in the center of the extended radio source Vela X, believed to mark the debris of a supernova explosion. Soon afterward the Crab Nebula was found to contain a radio pulsar with an even shorter period, 33 ms (Reifenstein et al. 1969) and the neutron star theory received further impressive support when optically visible pulsed light was observed (Nather et al. 1969) from the very star that Walter Baade and Rudolph Minkowski had identified in 1942 as the central stellar remnant of the Crab Nebula supernova explosion (Baade 1942; Minkowski 1942). Astronomers have used the powerful *Chandra X-ray Observatory* to trace out the jets, rings, winds, and shimmering shock waves generated by the highly magnetized, rapidly spinning pulsar.

As Franco Pacini previously demonstrated, before the discovery of radio pulsars, a rapidly spinning and highly magnetized neutron star is a powerful source of electromagnetic radiation (Focus 13.13), which could keep the Crab Nebula supernova remnant shining for thousands of years, and ever since the observation of the supernova explosion (Pacini 1967). Moreover, the rotational periods of radio pulsars are increasing, just as Gold predicted. If P denotes the radio pulsar period

and $dP/dt = \dot{P}$ designates the rate of increase of that period with time, then the age, τ, of the radio pulsar can be estimated from:

$$\tau = \frac{P\dot{P}}{2}. \tag{13.23}$$

The ages determined from the observed slow-down rate and periods of radio pulsars range from young pulsars with ages of 1,000–10,000 years to old pulsars with ages up to 100 million years. A supernova remnant will expand and dissipate into interstellar space, becoming undetectable in less than 100 thousand years. Most observed radio pulsars have therefore outlived any observable supernova remnant they might be associated with, and most pulsars are not found in one. Only a few rare, young pulsars, such as the Crab Nebula pulsar are found in a supernova remnant.

Focus 13.3 Luminosity, rotational energy, and magnetic field strength of a radio pulsar

A neutron star with a dipolar magnetic field will behave as a rotating magnetic dipole, with radiation luminosity L_{NS} given by:

$$L_{NS} = \frac{\mu_0 m_\perp^2 \omega^4}{6\pi c^3}, \tag{13.24}$$

where the magnetic constant $\mu_0 = 1.2566 \times 10^{-6}$ N A^{-2}, the symbol $m_\perp$ denotes the component of the magnetic dipole moment perpendicular to the rotation axis, the angular rotation velocity $\omega = 2\pi/P$ for a rotation period P, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. For a uniformly magnetized neutron star of radius R_{NS} and surface magnetic field of strength B_{NS}, the magnetic dipole moment is:

$$m_\perp = \frac{4\pi}{\mu_0} B_{NS} R_{NS}^3 \sin\theta, \tag{13.25}$$

where θ is the angle between the rotation axis and the magnetic axis. We can therefore express the neutron star luminosity by:

$$L_{NS} = \frac{8\pi \left(B_{NS} R_{NS}^3 \sin\theta\right)^2}{3\mu_0 c^3} \left(\frac{2\pi}{P}\right)^4. \tag{13.26}$$

The magnetic dipole radiation extracts rotational energy, E_{rot}, from the neutron star. The rotational energy is related to the moment of inertia $I = 2MR^2/5$ for a uniform rotating sphere of mass M and radius R, and given by:

$$E_{rot} = \frac{1}{2} I\omega^2 = \frac{2\pi^2 I}{P^2} = \frac{4\pi^2}{5} \frac{MR^2}{P^2}. \tag{13.27}$$

For a neutron star $M = M_{NS} \approx 1.4\, M_{\odot}$, where the Sun's mass $M_{\odot} = 1.989 \times 10^{30}$ kg, and the radius $R = R_{NS} = 10$ km $= 10^4$ m.

The rate of change, dE_{rot}/dt of the rotational energy with increasing time, t, is related to the increase $dP/dt = \dot{P}$ of the rotation period P, and given by the expression:

$$\frac{dE_{rot}}{dt} = -I\omega\frac{d\omega}{dt} = -\frac{4\pi^2 I\dot{P}}{P^3} = -\frac{8\pi^2}{5}MR^2\frac{\dot{P}}{P^3}. \tag{13.28}$$

The pulsar period and its rate of change can be combined with this equation to give the power extracted from the rotation of a neutron star, with $M = M_{NS} \approx 1.4\, M_{\odot}$, where the Sun's mass $M_{\odot} = 1.989 \times 10^{30}$ kg, and the radius $R = R_{NS} = 10$ km $= 10^4$ m.

These measurements can also be used to provide a lower limit to the magnetic field strength B at the surface of the pulsar. We do not know the inclination angle θ between the rotation axis and the magnetic axis, but we do know that θ is less than 90° and that sin θ is less than or equal to 1, or that sin $\theta \leq 1$. Setting the loss in rotational energy dE_{rot}/dt equal to the neutron star luminosity L_{NS} and combining terms from the previous equations we obtain:

$$B \geq \left(\frac{3\mu_0 c^3 M}{80\pi^3 R^4}\right)^{1/2}\left(P\dot{P}\right)^{1/2}. \tag{13.29}$$

For a neutron star with $M = M_{NS} \approx 1.4\, M_{\odot}$, where the Sun's mass $M_{\odot} = 1.989 \times 10^{30}$ kg, and a radius $R = R_{NS} = 10$ km $= 10^4$ m, this equation becomes:

$$B \geq 3 \times 10^{15}(P\dot{P})^{1/2} \text{ tesla}. \tag{13.30}$$

As demonstrated in the next example, the loss of rotational energy inferred from the period increase of the pulsar is exactly what is needed to keep the Crab Nebula shining at the present rate for about 1,000 years, ever since the observation of the supernova explosion that was associated with the pulsar's birth. So it is the central pulsar that makes the nebula glow. Moreover, it was soon shown that radio pulsars could efficiently accelerate particles to nearly the speed of light, accounting for both the synchrotron radiation of the Crab Nebula and the beamed radio waves that are observed as pulsars.

Example: Energy loss and magnetic field strength of the Crab Nebula pulsar

The pulsar at the center of the Crab Nebula has a period $P = 0.033326$ s and a period increase of $dP/dt = \dot{P} = 4.213 \times 10^{-13}$ s s^{-1}. Assuming that this pulsar has a mass equal to the mass of a neutron with a mass $M = M_{NS} \approx 1.4\, M_{\odot}$, where the Sun's mass $M_{\odot} = 1.989 \times 10^{30}$ kg, and a radius $R = R_{NS} = 10$ km $= 10^4$ m, the rate of change of the rotational energy is $dE_{rot}/dt = (8\pi^2/5)MR^2\dot{P}/P^3 \approx 5 \times 10^{31}$ J s^{-1}. This rate at which the Crab Nebula pulsar loses its rotational energy is comparable to the energy requirements of the surrounding supernova remnant, including its non-thermal synchrotron radiation and its expansion.

We can also provide a lower limit to the pulsar's surface magnetic field from $B \geq [3\mu_0 c^3 M/(80\pi^3 R^4)]^{1/2}\left(P\dot{P}\right)^{1/2} \approx 3 \times 10^{15}\left(P\dot{P}\right)^{1/2}$ tesla $\approx 4 \times 10^8$ tesla. This magnetic field strength is comparable to that expected from the gravitational collapse of the Sun to a neutron star.

When the observed slow-down rate of most radio pulsars is combined with their periods, typical ages of between 1 million years to 100 million years are obtained. A supernova remnant will expand and dissipate into the vastness of interstellar space, becoming unrecognizable in less than 100 thousand years, removing all signs of the pulsar's birth. Pulsars therefore outlive their supernova remnants and most pulsars are not found in one. Only a few rare, young pulsars, such as the Crab Nebula Pulsar, are rotating fast enough to efficiently accelerate particles to very high energies.

Although most radio pulsars are alone in space without a nearby companion, some binary pulsars have been discovered, the most famous being PSR 1913 + 16 with a period of 0.05898 s; the PSR designates pulsar and 1913 + 16 specifies its position in the sky. Russell A. Hulse (1950–) and Joseph H. Taylor, Jr. (1941–) found it as a result of a deliberate, high-sensitivity, computerized search for new radio pulsars at the Arecibo Observatory in Puerto Rico (Hulse and Taylor 1975). The discovery of a radio pulsar that is a member of a double-star system indirectly suggested the emission of gravitational waves, which had never been seen before (Focus 3.6, Sect. 3.5).

A second binary, millisecond radio pulsar, designated PSR J1614–2230, has now been found, with a period of just 3.15 ms. It is attributed to a neutron star in orbit around a white dwarf star with an orbital period of 8.7 days, and has been used to test aspects of general relativity theory other than gravitational waves. These investigations have shown that the pulsar is the most massive neutron star known so far, with a mass of 1.97 solar masses; the mass of the white dwarf companion is 0.50 solar masses (Demorest et al. 2010).

Kramer and Stairs (2008) have summarized knowledge of the double pulsar; Hughes (2009) has discussed gravitational waves from merging compact binaries; and Joss and Rappaport (1984) have reviewed neutron stars in interacting binary systems.

Individual stars are bound together so tightly that a supernova that leads to the formation of a neutron star in a binary-star system may not disrupt its companion; the two stars can remain together, as evidenced by the binary pulsars. This is important for understanding pulsars that have been detected at x-ray wavelengths. Unlike most radio pulsars, they are members of close binary-star systems rather than single, isolated neutron stars.

13.7.3 X-ray Pulsars from Neutron Stars in Binary Star Systems

Because x-rays are absorbed in our atmosphere, cosmic x-ray sources must be observed with instruments launched above the obscuring air, in rockets or satellites. By the mid-twentieth century brief, 5-minute rocket flights had shown that the Sun radiates detectable x-rays, and it was thought that lunar material also might emit them when illuminated by solar x-rays.

Riccardo Giacconi's (1931–) group at the American Science and Engineering Company (AS&E) concluded that x-rays emitted by conventional stellar objects other than the Sun would be too faint to be detected with existing instruments. They designed the sensitive equipment needed to detect the Moon's x-rays and to search for other unknown sources of x-ray radiation. They unexpectedly found the first known discrete x-ray source outside of the solar system, which led to the discovery of a new class of cosmic objects and new physical processes (Giacconi et al. 1962).

This pioneering rocket flight set the stage for a host of rocket and satellite observations of discrete x-ray sources, including x-ray stars that are 1,000 times brighter in x-rays than the Sun at all wavelengths and that are 1,000 times more luminous in x-rays than in visible light. The x-rays signaled the presence of 1-million-degree gas spiraling from a close companion star into a neutron star or black hole.

Giacconi was awarded the 2002 Nobel Prize in Physics for these pioneering contributions, which led to the discovery of cosmic x-ray sources. He shared the prize with Raymond Davis, Jr. (1914–2006), and Masatoshi Koshiba (1926–), who were recognized for their detection of cosmic neutrinos.

One of the brightest sources in the newly discovered x-ray sky, designated Centaurus X-3, pulses in x-rays every 4.84 s. The rapid pulsating x-ray variations were discovered shortly after the launch of the first dedicated x-ray satellite, on December 12, 1970 from the offshore San Marco platform, near the Coast of Kenya (Giacconi et al. 1971). Because this date coincided with the seventh

Table 13.12 Physical properties of binary x-ray pulsars

P_{XP} = period of x-ray pulsar = 0.7–800 s
dP_{XP}/dt = rate of decrease of x-ray pulsar period = $-(10^{-5}$–$10^{-2})\ P_{XP}$ year^{-1}
P_0 = orbital period = 5–10 days
L_{XP} = x-ray luminosity = (0.1–10) × 10^{30} J s^{-1}
M_{XP} = mass of x-ray pulsar = 1.05–1.87 $M_\odot \approx$ (2.1–3.7) × 10^{30} kg

anniversary of the independence of Kenya, the satellite was given the name *Uhuru*, the Swahili word for freedom.

After analyzing a year of observations of Centaurus X-3, the *Uhuru* scientists found a regular pattern of intensity changes of the x-ray pulses, which increased and decreased in strength with a much longer period of 2.087 days and systematic changes in the timing of pulses in the same period. These effects were attributed to a companion star, which was orbiting the x-ray source and regularly eclipsing its emission.

During the next year, accurate measurements showed that the average pulsation period of Centaurus X-3 was getting shorter, which meant that its rotation was speeding up, not slowing down like radio pulsars. This indicated that the rotational energy of the x-ray pulsar was increasing, rather than decreasing, with time. The physical properties of such binary x-ray pulsars are given in Table 13.12.

The gain in rotational energy is attributed to matter drawn in from a nearby optically visible companion star (Fig. 13.11). Because matter is being pulled toward the surface of an x-ray-emitting neutron star, instead of being expelled from it, the neutron star is knocked up to a faster rotation. The material spirals in at the same direction as the neutron star's rotation. When it lands, it gives the neutron star a sideways kick, increasing its rotational energy, speeding it up, and causing the rotation period to become shorter as time goes on. Because most radio pulsars are not members of binary-star systems, they expel material, lose rotational energy, slow down, and have periods that lengthen.

So a neutron star can be seen in x-rays when it is in very close orbit with a normal star that shines in visible light detected with an optical telescope. This stellar companion does not radiate detectable x-rays, but gas flowing from it fuels the x-ray emitting neutron star.

And there are two ways of looking at the stellar duo. The visible picture portrays only the normal star, and the x-ray image just reveals its compact neighbor, the neutron star. A complete understanding of the double-star system, one visible and the other unseen, can only be understood when the two perspectives are combined.

Because the two stars are rapidly orbiting around one another, the gas from the ordinary star does not fall directly onto the neutron star but instead shoots past the neutron star, swinging around it to form a whirling disk of hot gas, known as an *accretion disk*. It spirals around and down onto the central neutron star. The inner portions of the swirling accretion disk revolve more rapidly than the outer

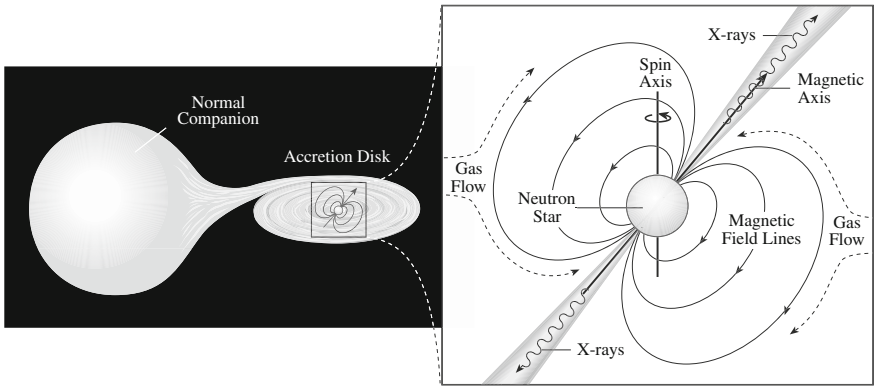


Fig. 13.11 X-ray pulsar The outer atmosphere of an ordinary star, detected in optically visible light, spills onto its companion, an invisible neutron star. The flow of gas is diverted by the powerful magnetic fields of the neutron star, which channel the in-falling material onto the magnetic polar regions. The impact of the gas on the star creates a pair of x-ray hot spots aligned along the magnetic axis at each magnetic cap. Because the magnetic-field axis can be inclined to the neutron star's rotation axis, the x-ray radiation from the hot spots can sweep across the sky once per rotation, which is observed as periodic x-rays if one of the hot spots intersects the observer's line of sight

portions, just as the closer planets orbit the Sun at faster speeds than more remote planets.

The rapidly spinning inner parts of the disk constantly rub against the slower-moving outer parts. This viscous friction heats up the accretion disk and causes the material in it to spiral inward. The closer the material moves toward the central neutron star, the hotter the in-falling gas becomes, eventually reaching temperatures of millions of K and emitting luminous x-rays.

The intense magnetic field of a rotating neutron star acts as a funnel to guide in-falling matter onto a neutron star's magnetic north and south poles, creating an x-ray pulsar. As the accreting material heats up and falls onto the polar surfaces of the neutron star, it emits two beams of x-rays that flash in and out of view as the neutron star rotates and one or two of the beams sweep past the Earth.

But there is a limit to the x-ray emission produced by the accreting material, owing to the radiation pressure it develops. The maximum luminosity is now called the Eddington limit, after Arthur Stanley Eddington (1882–1944) who showed that greater luminosity would blow away any surrounding matter – long before the discovery of any x-ray star (Eddington 1926). The outward force of radiation pressure becomes equal to the inward gravitational force on the accreting material when the radiation luminosity is at the Eddington luminosity. A neutron star that is accreting material at close to the Eddington limit will heat up to a temperature of about 20 million K and emit intense x-ray radiation (Focus 13.4).

Focus 13.4 Accretion luminosity and the Eddington limit

The luminous output resulting from mass falling, or accreting, onto a compact object will depend upon the rate of mass transfer, denoted by $\dot{M} = dM/dt$, as well as the mass, M, and radius, R, of the compact object. From the conservation of energy, half the release in gravitational potential energy will be equal to the gain in kinetic, or thermal, energy of the hot accreting gas, and the remaining half will be converted into heat. The luminous output, L_{acc}, of the thermal radiation from the accreting material will be:

$$L_{acc} = \frac{G\dot{M}M}{2R}. \tag{13.31}$$

The temperature, T, of the radiating gas can be obtained from the Stefan–Boltzmann law, which results in

$$T = \left[\frac{L_{acc}}{4\pi\sigma R^2}\right]^{1/4} = \left[\frac{G\dot{M}M}{8\pi\sigma R^3}\right]^{1/4}, \tag{13.32}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and the Stefan-Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-1}.

When the compact receiving object is a white dwarf star, the mass is $M \approx M_\odot = 1.989 \times 10^{30}$ kg, the radius is $R_{WD} \approx 6 \times 10^6$ m, and the mass accretion rate $\dot{M} = 10^{-9}\, M_\odot$ yr^{-1}, where 1 year $= 3.1557 \times 10^7$ s. The accretion luminosity is $L_{acc} \approx 7 \times 10^{26}$ J s^{-1} $\approx 2\, L_\odot$, and the temperature $T \approx 3 \times 10^5$ K. For a neutron star with the same rate of mass transfer, the luminosity will be about a thousand times that of the Sun and the temperature will be about a million K, but higher rates of mass accretion are possible.

Arthur Stanley Eddington (1882–1944) showed that there is a maximum luminosity, now called the Eddington luminosity, L_{Edd}, for any source of radiation before it blows away the surrounding matter. In effect, the outward force of radiation pressure from the compact accreting star pushes against the inward gravitational force on the accreting material. The two forces become equal when the radiation luminosity is (Eddington 1926a, b Sect. 10.1, Focus 10.1):

$$L_{Edd} = \frac{4\pi G m_p c M}{\sigma_T} \approx 6.3M \text{ J s}^{-1}, \tag{13.33}$$

or

$$L_{Edd} \approx 1.25 \times 10^{31} \frac{M}{M_\odot} \text{ J s}^{-1}, \tag{13.34}$$

where M is the mass of the compact accreting star, the Thomson scattering cross section for the electron is $\sigma_T = 6.65246 \times 10^{-29}$ m^2, the mass of the proton is $m_p = 1.6726 \times 10^{-27}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

The numerical approximation can also be written:

$$L_{Edd} \approx 3.3 \times 10^4 \frac{M}{M_\odot} L_\odot \text{ J s}^{-1}. \tag{13.35}$$

where the Sun's luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1},

The Eddington limit sets an upper limit to the accretion luminosity of a compact object, since for L_{acc} greater than L_{Edd}, the radiation pressure will inhibit further accretion. The maximum accretion rate, $\dot{M}_{Edd}$, is obtained by equating the accretion luminosity to the Eddington limit, or

$$\dot{M}_{Edd} = \frac{8\pi m_p c R}{\sigma_T}, \tag{13.36}$$

or

$$\dot{M}_{Edd} \approx 2 \times 10^{11} R \text{ kg s}^{-1} \approx 3 \times 10^{-12} R\, M_\odot \text{ yr}^{-1}, \tag{13.37}$$

for a compact object of radius R. At the Eddington limit, a white dwarf star with $R = 6 \times 10^6$ m, the accretion rate would be about 2×10^{-5} solar masses per year. It can be substantially less than this amount.

Sometimes a compact, invisible object with a visible companion can be force-fed with more matter than it can consume, and it hurls the in-falling matter out in two oppositely directed jets. This can happen when the invisible star is a black hole, which we discuss next.

Example: Accretion luminosity and temperature from mass transfer to a neutron star

The maximum mass transfer rate $\dot{M}_{Edd}$ onto a neutron star of radius $R_{NS} = 10$ km $= 10^4$ m will be $\dot{M}_{Edd} = 8\pi m_P c R_{NS}/\sigma_T \approx 2 \times 10^{15}$ kg s^{-1}, where the proton mass $m_P = 1.6726 \times 10^{-27}$ kg, the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, and the Thomson scattering cross section $\sigma_T = 6.65246 \times 10^{-29}$ m^2. The accretion luminosity onto a solar-mass neutron star of mass $M = 1.4\, M_\odot = 2.785 \times 10^{30}$ kg will be $L_{acc} = GM_\odot \dot{M}_{Edd}/2R_{NS} \approx 1.25 \times 10^{31}$ J s^{-1}, where the Newtonian constant of gravitation $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}. The temperature, T, of the gas can be estimated from the Stephan-Boltzmann law $L_{acc} = 4\pi\sigma R_{NS}^2 T^4$, or $T = \left[L_{acc}/\left(4\pi\sigma R_{NS}^2\right)\right]^{1/4} \approx 2 \times 10^7$ K, where the Stefan−Boltzmann

constant $\sigma = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}. From the Wien displacement law (Sect. 2.4), a thermal gas at this temperature will emit most of its radiation at a wavelength $\lambda_{max} = 0.0029/\mathrm{T} = 1.4 \times 10^{-10}$ m, or at x-ray wavelengths.

13.8 Stellar Black Holes

13.8.1 Imagining Black Holes

John Michell (1724–1793), a British clergyman and natural philosopher, suggested more than two centuries ago that certain stars could remain forever invisible. He reasoned that a star might be so massive and its gravitational pull so powerful that light could not escape it. He wrote: "All light from such a body would be made to return to it by its own proper gravity" (Michell 1784).

The French astronomer and mathematician Pierre Simon de Laplace (1749–1827) popularized the idea, including it in his *Exposition du système du monde*. He subsequently showed that light could never move fast enough to escape the immense gravitational attraction of some compact stars (Laplace 1796). Their matter might be so concentrated, and the pull of gravity so great, that light could not emerge from them, making these stars forever dark and invisible.

When it was realized that light travels at a definite speed of 2.9979×10^8 m s^{-1}, or roughly 300,000 km s^{-1}, a *black hole* could be defined as any object whose escape velocity, the velocity required for escape from an object's gravitational pull, exceeds the velocity of light. All stars that we can observe have escape velocities smaller than the speed of light, which is why we can observe their light. The Sun, for example, has an escape velocity of 600 km s^{-1}. Compressing a star into a smaller size raises its escape velocity. When a dead stellar core of approximately the Sun's mass collapses into a white dwarf star, its escape velocity increases to about 9,000 km s^{-1}, or 3 % of the speed of light. At a neutron star's radius, the escape velocity becomes about 210,000 km s^{-1}, or 70 percent of the speed of light.

If a massive star has consumed all of its available nuclear fuel and its core mass exceeds the upper limit to a neutron star's mass, at about 3 solar masses, the stellar core's gravity will overcome the degenerate neutron pressure of even nuclear matter. At this point, there is no known force that can halt the collapse. The smaller the collapsing core becomes, the larger is its escape velocity, until it exceeds the speed of light and a stellar black hole is formed. It is black, or rather invisible, because no light can leave it, and it is a hole because nothing that falls into it can escape.

In other words, if a star has used up its thermonuclear fuel and its core is sufficiently massive, there is nothing left to hold back the inexorable force of gravity. The core continues to collapse forever, vanishing from the directly observable universe.

13.8.2 Observing Stellar Black Holes

Since a black hole is invisible, and it does not absorb, emit, or reflect radiation, how do we know it is there? We detect a black hole by its gravitational effect on the motion of a visible star.

With remarkable foresight, the Reverend John Michell also speculated, in 1784, that the unseen star might betray its presence by its gravitational effects on a nearby, luminous star in orbit around it (Michell 1784). In modern extensions of this idea, a black hole may be detected if it is in a tight, close orbit with a visible star whose outer atmosphere spills over into the dominant gravitational influence of the black hole. This material swirls around and down into the black hole, orbiting faster and faster as it gets closer – as a result of the ever-increasing gravitational forces. The rapidly moving particles collide as they are compressed to fit into the hole, heating the material to temperatures of millions of K. At these temperatures, the gas emits almost all of its radiation at x-ray wavelengths. It is analogous to the accretion disks of binary x-ray pulsars, except that the invisible companion is a black hole rather than a neutron star.

Remillard and McClintock (2006) have reviewed the x-ray properties of black-hole binaries, whereas Eardley and Press (1975) provided an earlier review of astrophysical processes near black holes. Mirabel and Rodriguez (1999) review black-hole sources of relativistic jets in our Galaxy.

So, the way to find a stellar black hole is to look for two stars that are in close orbit, one a normal visible star and the other unseen except for its x-rays. The mass, velocity, and orbital period of the visible star can be used to determine the mass of its orbital partner, which emits no visible light. If that mass is noticeably greater that the upper mass limit for a neutron star, set at about 3 solar masses, the unseen star is thought to be a stellar black hole. Any normal star with this mass would be very bright and easily seen through a telescope, but a black hole is dark, emitting no detectable visible light.

The archetype of a stellar black hole is Cygnus X-1, located in the constellation Cygnus and one of the first x-ray sources to be discovered. Rapid, irregular x-ray bursts from this object were detected from the *Uhuru* satellite in 1970. The x-rays flickered on and off as rapidly as a few milliseconds, and because nothing travels faster than the speed of light the emitter had to be less than 300 km across. This meant that it was smaller than a white dwarf star, and had to be either a neutron star or a black hole.

Cygnus X-1 is accompanied by a bright, blue supergiant star of spectral class O, located at a distance of about 6,000 light-years from the Earth. Observations of the continuously and periodically shifting spectral lines of this bright visible star – by the English astronomers B. Louise Webster (1941–1990) and Paul Murdin (1942–), and independently confirmed by the Canadian astronomer Charles Thomas Bolton (1943–) – indicated that it is revolving every 5.60 days about an invisible companion of more than eight times the mass of the Sun for the unseen companion (Webster and Murdin 1972; Bolton 1972, 1975). It emits no light and its mass is greater than that of the upper mass limit for a neutron star, of about 3.0 solar masses; therefore, by elimination, it must be a stellar black hole.

Recent measurements of the Cygnus X-1 binary star system indicate that the visible supergiant star has a mass of 19.2 solar masses and the unseen companion has a mass of 14.8 solar masses, which confirms that its mass is well above the upper mass limit for a neutron star. In addition, the two stars are separated by only 0.2 AU, or 20 percent of the distance from the Earth to the Sun and about half the separation of Mercury from the Sun (Orosz et al. 2011; Reid et al. 2011). Like other supergiant stars of its spectral type, the visible star is thought to be shedding mass in a stellar wind at a rate of about 2.3 solar masses every 1 million years. Due to the proximity of the invisible companion, a significant portion of this wind is being drawn into the black hole to form its x-ray-emitting accretion disk.

We now know of many stellar black holes identified in this way. The orbital properties of visible companions of cosmic x-ray sources indicate masses beyond the neutron-star limit. Like Cygnus X-1, many of these black holes can also exhibit highly luminous, rapid, and irregular x-ray outbursts, showing that they are very small on a cosmic scale. The transient x-ray flickering, sometimes brightening a million-fold in milliseconds, is most likely emitted as in-falling material takes the final plunge and vanishes into the black hole.

13.8.3 Describing Black Holes

The outer edge of a black hole can be defined as the radius at which the escape velocity, required to escape from its gravitational pull, is equal to the speed of light, or when the kinetic energy of an object moving at this speed is equal to the gravitational potential energy of the mass holding it in. This radius is 3,000 m for a stellar black hole with the mass of the Sun. It is known as the *Schwarzschild radius* in recognition of the German astronomer Karl Schwarzschild (1873–1916), who first recognized its mathematical significance, and also sometimes called the *gravitational radius*. While serving as an artillery lieutenant on the Russian front during World War I (1914–1918), Schwarzschild derived the solution to Einstein's *General Theory of Relativity* for a spherical, non-rotating black hole. His publication, titled "On the Field of Gravity of a Point Mass in the Einsteinian Theory" can be used to specify the space time intervals outside a black hole (Schwarzschild 1916).

The radius of a black hole can be defined as the radius, R, at which the escape velocity, V_{esc}, of a particle of mass, m, from a larger mass, M, and radius, R, is equal to the speed of light, c, or when $V_{esc} = (2GM/R)^{1/2} = c$, for a gravitational constant G. That's obtained from classical physics by equating the kinetic energy of a moving object to the gravitational potential energy of the mass holding it in, with $V_{esc} = c$. That is the kinetic energy $mV_{esc}^2/2 = GMm/R$, the gravitational potential energy. Solving for the radius, we obtain the Schwarzschild radius R_g given by

$$R_g = \frac{2GM}{c^2} \approx 2.95 \times 10^3 \left(\frac{M}{M_\odot}\right) \text{ m}, \tag{13.38}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, and the solar mass $M_\odot = 1.989 \times 10^{30}$ kg.

Example: Schwarzschild radius of the Earth and Sun
The Schwarzschild radius, also known as the gravitational radius, of the Earth is $R_{SE} = 2GM_E/c^2 \approx 8.87 \times 10^{-3}$ m, where the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ N m^2 kg^{-2}, the mass of the Earth $M_E = 5.974 \times 10^{24}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The physical radius of the Earth is $R_E = 6.378 \times 10^6$ m, which is almost a billion times larger than the planet's gravitational radius. For the Sun, the gravitational radius $R_{S\odot} \approx 2.95 \times 10^3$ m ≈ 3 km, for a solar mass $M_\odot = 1.989 \times 10^{30}$ kg. The physical radius of the Sun $R_\odot = 6.955 \times 10^8$ m is more than a hundred thousand times larger than its gravitational radius. At the time that Schwarzschild derived the metric involving this term, there was no known physical object whose linear size was smaller than its gravitational, or Schwarzschild, radius.

In the *Special Theory of Relativity*, space and time were combined to define a metric, ds, or space time interval, given by (Minkowski 1908; Hargreaves 1908)

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2, \tag{13.39}$$

for time interval dt and space coordinates x, y, z. The speed of light, c, has been added to give dt the units of distance. This metric applies to Euclidean space without gravity and no curvature of space.

In the *General Theory of Relativity*, for which gravity manifests itself in the curvature of space time, the metric outside a non-rotating mass in a vacuum is the *Schwarzschild metric* given by (Schwarzschild 1916):

$$ds^2 = \left[1 - \frac{2GM}{c^2 r}\right] c^2dt^2 - \frac{dr^2}{\left[1 - \frac{2GM}{c^2 r}\right]} - r^2d\theta^2 - r^2 \sin^2\theta d\phi^2. \tag{13.40}$$

Here r, θ, ϕ are spherical coordinates whose origin is at the center of the massive object, and M is the mass. The trajectories of free-falling particle and radiation are described by a null geodesic, for which $ds = 0$.

A proper time τ is defined by the proper time interval $d\tau$ given by:

$$d\tau = \frac{ds}{c}. \tag{13.41}$$

For a clock at rest, the space coordinates are not changing, so $dr = d\theta = d\phi = 0$, and the proper time interval is:

$$d\tau = \left[1 - \frac{2GM}{c^2 R}\right]^{1/2} dt. \tag{13.42}$$

The proper time interval approaches zero times dt when the radius R approaches the Schwarzschild radius $R_S = 2GM/c^2$. A clock placed on a collapsing star will appear, to a distant observer, to tick more slowly as the star's radius approaches the Schwarzschild radius. This is known as gravitational *time dilation*. The clock will appear to stop, at this critical radius, when time will seem to go on forever and the star has become "frozen".

The decline in the observed radiation due to gravitational time dilation is expressed as a decrease in the number, N_{ph}, of photons observed, with (Oppenheimer and Snyder 1939):

$$N_{ph} \propto \exp\left(\frac{-t}{2R_S}\right), \tag{13.43}$$

or with a characteristic observed free-fall time, τ_{ff}, given by:

$$\tau_{ff} = \frac{2R_S}{c} = \frac{4GM}{c^2} \approx 2.94 \times 10^3 \left(\frac{M}{M_\odot}\right) \text{ s}, \tag{13.44}$$

where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg.

Example: A star's core collapse time to become a black hole

If a sufficiently massive star consumes all the available thermonuclear fuel, and its core is more massive than about 3 solar masses, it will collapse to a black hole in a free-fall time given by $\tau_{ff} = 4GM/c^2$, with the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. For a 5 solar mass star with $M = 5\ M_\odot$, where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, the free-fall time is $\tau_{\text{ff}} \approx 10^4$ s.

During this time, the number of observed photons will have decreased by exp (−1,000), or e^{-1000}, a very small number, so the star will have effectively disappeared, becoming invisible and black.

The Schwarzschild metric has a one over zero term, a singularity, which blows up at the Schwarzschild, or gravitational, radius $R_g = 2GM/c^2$, which is the radius at which the escape velocity becomes equal to the speed of light. In 1963 the New Zealand mathematician Roy Kerr (1934–) described the space–time metric outside a rotating black hole, and it reduces to Schwarzschild's metric when there is no rotation. The gravitational field outside black holes are uniquely determined by their mass and angular momentum and described by Kerr metrics.

The singularity can be moved down to the center of the black hole by a clever change in geometry, but it never disappears. It is a location where matter is compressed into an undefined state of infinite mass density. That is, in 1960 two mathematicians, the American Martin Kruskal (1925–2006) and the Hungarian-Australian George Szekeres (1911–2005), defined a line element that removes the singularity at the Schwarzschild radius through a coordinate transformation, but that didn't mean that the singularity disappeared – you can't form a black hole within the constraints of *General Relativity* without having one; the English astrophysicist Roger Penrose (1931–) demonstrated that (Penrose 1965). The singularity was just moved down to the center of the black hole in the Kruskal-Szekeres transformation.

Nevertheless, the defining notion of a black hole, in terms of an escape velocity that exceeds the speed of light, and the method of inferring its presence in an x-ray emitting binary star system with a close visible companion, do not depend on Einstein's theory and can be determined from classical Newtonian physics without any singularity.

The Schwarzschild radius, which can be located outside the singular center, has not lost its significance. It marks the event horizon – literally, a horizon in the geometry of space–time beyond which no event can be seen, just as the Earth's horizon is the boundary for our vision. Nearby space then is said to curl into a black hole, carrying light and matter and any other form of energy with it. They are so intensely wrapped around a black hole that it becomes a cocoon disconnected from the outside and cut off forever from the rest of the universe.

Black holes are mysterious objects. They cannot be observed directly, because any radiation they might emit cannot escape. A black hole's presence can be inferred only from indirect, circumstantial evidence, using measurements in the accessible parts of the universe − the visible stars − to make inferences about the dark places that cannot be observed.

Chapter 14
A Larger, Expanding Universe

14.1 The Milky Way

14.1.1 A Fathomless Disk of Stars

On a clear, moonless night, we can look up and see a hazy, faintly luminous band of light that stretches across the sky from one horizon to the other; it is known as the Milky Way (Fig. 14.1). According to ancient Greek myth, the goddess Hera, Queen of Heaven, spilled milk from her breast into the sky. The Romans called the spilt milk the *Via Lactea*, or the "Milky Way." It also is designated as our Galaxy, derived from the Greek word *galakt-* for "milk," the celestial milk from Hera's breast.

We are immersed within the *Milky Way*, viewing it edgewise from inside. When gazing directly into the band of starlight, we cannot see through to stars at the center or distant edges of the Milky Way, but if we look up and outside the thin disk of stars, we can look beyond them. It is similar to living in a city: We notice buildings all around us, but none when we look up into the sky.

When Galileo Galilei (1564–1642) turned one of the first telescopes toward the Milky Way, he found that it contains many otherwise unseen stars, which are too faint to be seen by the unaided human eye (Galilei 1610). Astronomers subsequently built increasingly larger telescopes, which collect more starlight and enable us to see the dim, golden beacons of fainter stars. They discovered more dim stars located between or beyond the brighter ones, which make the Milky Way look like a continuously distributed band of light when observed by the unaided eye.

The German-born English astronomer William Herschel (1738–1822) spent much of his life trying to determine the shape and size of the Milky Way; he constructed the biggest telescopes at the time, with the largest mirrors and greatest light-gathering power. By counting the number of stars of different observed brightness in various directions in the night sky, he hoped to determine the places at which the stars disappeared, thereby determining the depths of the Milky Way.

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_14, © Springer-Verlag Berlin Heidelberg 2013

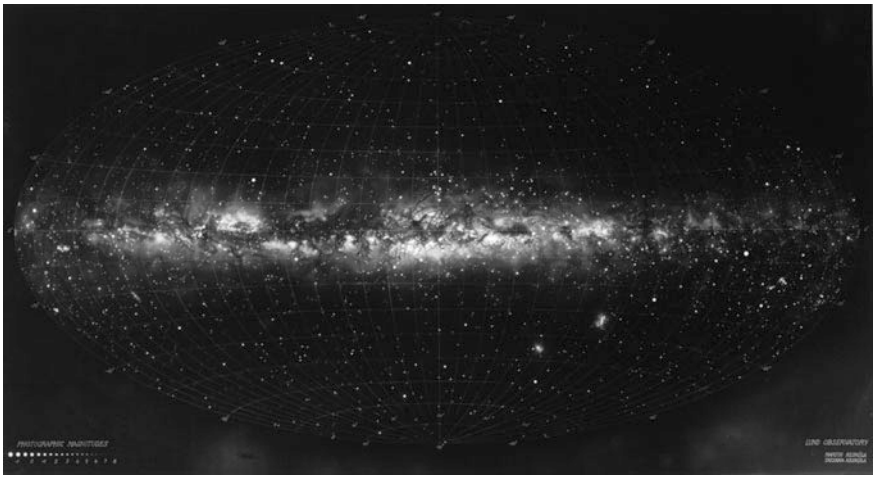


Fig. 14.1 The Milky Way A panoramic telescopic view of the Milky Way, the luminous concentration of bright stars and dark intervening dust clouds that extends in a band across the celestial sphere. We live in this disk and look out through it. Our view is eventually blocked by the buildup of interstellar dust, and the light from more distant regions of the disk cannot get through. The center of the Milky Way is located at the center of the image, in the direction of the constellation Sagittarius. Although the disk appears wider in that direction, the center is not visible through the dust. The large and small magellanic clouds can be seen as bright swirls of light below the plane to the right of center (this map of the Milky Way was hand-drawn from many photographs by Martin and Tatjana Keskula under the direction of Knut Lundmark; courtesy of the Lund Observatory, Sweden)

But the giant telescopes were not big enough to fathom the profundity or depth of the Milky Way. By collecting greater amounts of light than a smaller telescope, Herschel's biggest telescope brought fainter stars into view and pushed the edge of the known universe further into space. These stars were concentrated in a flattened disk with the Sun at the center, and whose greatest extent is in the plane of the Milky Way (Herschel 1785).

Although Herschel concluded that the Sun is in the center of a flattened disk of stars with a disk diameter five time times its thickness, he had no way to determine its size. Early in the twentieth century, the Dutch astronomer Jacobus C. Kapteyn (1851–1922) and his colleagues resumed the star counts, arriving at a similarly flattened, Sun-centered distribution of stars with the greatest extent in the Milky Way. Measurements of the distance of some of these stars, using their parallax, provided a scale to Kapteyn's universe of about 1.5 kpc by 12 kpc, where 1 kpc is 1,000 pc or 3.0857×10^{19} m (Kapteyn 1922).

However, astronomers have never succeeded in deciphering the true extent of the Milky Way by observing its stars, even when looking much farther into it using larger telescopes and photographic or electronic techniques that permitted long exposures. This is because the most powerful telescopes can discern only the

Fig. 14.2 Globular star cluster More than a hundred thousand stars are collected together in this globular star cluster designated NGC 362, which is located about 27,700 light-years away in the southern sky. It is one of many star clusters that are located in an extensive, spherical halo around our Milky Way. They formed in the early evolution of the Milky Way, and contain stars that are more than 10 billion years old. (Courtesy of Royal Observatory, Edinburgh.)

visible parts of our stellar system, not its most distant, invisible parts that lie behind an opaque veil of interstellar dust.

This dust blocks our view when we look deep into the plane of the Milky Way. The total amount of dust through which we are looking builds up with distance and eventually makes an impenetrable barrier; it becomes so thick and dense that it blocks the light of distant stars. We can see only that far; more distant objects are hidden from view. New perspectives were required to look outside and eventually beyond this barrier to the heavens.

14.1.2 The Sun is Not at the Center of Our Stellar System

The true enormity of our stellar system was discovered by using Cepheid variable stars to gauge the distances of globular star clusters (Fig. 14.2) located outside the plane of the Milky Way. These yellow supergiant Cepheids periodically brighten and dim with a period that increases with a star's luminosity. Measurements of this period and, therefore, the stellar luminosity can be combined with observations of a star's brightness to determine its distance (Focus 14.1). Because the Cepheid variable stars are very luminous, they are conspicuous and can be seen to exceptional distances, where conventional parallax methods of determining stellar distance do not work. At large distances, the parallax angles are too small to be reliably measured, even from space.

Focus 14.1 Cepheid variable stars

The luminosity of some stars does not remain constant but instead fluctuates over regular periods. These stars do not only turn on and off, like a switched house light, but instead gradually vary from dimmer to brighter and then back to dimmer again with periods ranging from a few days to a few months.

The very luminous variable stars are know as *Cepheid variable stars*, from their prototype, Delta Cephei. The deaf English astronomer John Goodricke (1764–1786) first noticed its variability (Goodricke 1785). Edward Pigott (1753–1825), another Englishman, discovered the Cepheid variable Eta Aquilae a few months earlier in the same year (Pigott 1785). The North Star, Polaris, is also a Cepheid variable star, the closest one known.

The Cepheids have luminosities up to 100,000 times that of the Sun and masses of 4–20 times the solar mass. Because they are so luminous, these stars can be seen over a wide range of distances, from Delta Cephei, located at only about 272 parsecs, or 887 light-years, from the Earth (Benedict et al. 2002), to galaxies 100 million light-years away (Freedman et al. 2001).

The more luminous a Cepheid variable star is, the more slowly it varies and the longer the period of its luminosity change. This period-luminosity relationship was first discovered from observations of variable stars in the Large and Small Magellanic clouds, which are nearby satellites of the Milky Way. These stellar systems, visible from the Earth's Southern Hemisphere, are named for the Portuguese explorer Ferdinand Magellan (1480–1521), who observed them when his ships were circumnavigating the world for the first time.

At the end of the nineteenth century, Harvard College established an observatory at Arequipa, Peru, with a 0.6 m (24 inch) refractor that was used in a photographic survey of the southern sky, including the Magellanic Clouds. Because of their proximity, the clouds could be resolved into stars. From these photographs, Henrietta Swan Leavitt (1868–1921), a researcher in Cambridge, Massachusetts, found an extraordinary total of 1,777 variable stars. She reported that the brighter stars tended to have the longer cycles of variation (Leavitt 1908). Because the extent of the Magellanic Clouds is small compared to their distance, the relationship of period to apparent brightness also implied a real connection with luminosity. Four years later, Leavitt had obtained precise apparent brightness and period data for 25 variable stars in the Small Magellanic Cloud, thereby establishing the important *period-luminosity relation* for the Cepheid variables (Leavitt 1912).

Once this relation is calibrated suitably by the measurement of a precise distance to one Cepheid variable star using independent methods, observation of the variation period leads to determination of the star's luminosity. Then, using the observed brightness, the distance of the star can be calculated. This technique of measuring distances with Cepheid variable stars has been used to demonstrate the vast extent of the Milky Way – as well as the

Sun's place within it – and, subsequently, to discover the extragalactic nature of spiral nebulae.

The early history of the period-luminosity relation and its calibration has been reviewed by Fernie (1969) and Sandage (1972), and includes the initial calibration. The relation has been used to place constraints on the size and shape of the Milky Way, and to determine slightly incorrect distance estimates for nearby spiral nebulae.

There are the brilliant, younger, and more massive classical Cepheids, of stellar Population I, which are found in the arms of spiral galaxies, and the older, fainter Cepheids of Population II, located in globular star clusters.

Leavitt had no idea why the luminous output of these stars varies, but within a few years the inquisitive English astronomer Arthur Stanley Eddington (1882–1944) showed how (Eddington 1918, 1919). The stars are pulsating with a regular beat, expanding out and contracting in, rising and falling back. And since the star's size is changing, the period-luminosity relation can also be expressed as a period-radius relation. The radial oscillations have been observed using the Doppler shift of spectral lines that arise in the stellar atmospheres.

It follows from Eddington's theory that the pulsation period is inversely proportional to the square root of the star's average mass density, so more rarefied stars pulsate more slowly than dense ones. This also means that the more massive stars, which are also the larger and more luminous stars, possess longer pulsation periods.

Once it was realized that stars are primarily composed of hydrogen and helium, Eddington could show that stellar pulsations might originate in an outer convective zone where hydrogen is alternately ionized and neutral, acting as a valve that absorbs and releases heating radiation during the course of the pulsation.

Over a decade later, the Russian astronomer Sergei A. Zhevakin (1916–2001) showed that the convective ionized hydrogen zone could not maintain the pulsations because it does not absorb sufficient energy during the contraction of the star. He found that an outer region of doubly ionized helium acts as a valve for the heat engine that drives the pulsations of Cepheid variable stars (Zhevakin 1963; Cox 1980). Lang (1999) provides references to individual papers on stellar pulsations. The varying star absorbs the outward flow of energy from the star's center during stellar contraction and repeatedly returns it during expansion.

Because the pulsation depends upon a critical stage of ionization, a star can maintain them only for a specific combination of size and temperature in a narrow range of mass densities. This defines a strip of instability in the Hertzsprung-Russell diagram.

At optically visible wavelengths, astronomers establish the zero point, a, and slope, b, in the *period-luminosity relation* for a variable star:

$$M_V = -a - b \log P, \tag{14.1}$$

where M_V is the absolute visual magnitude and P is the period. A precise calibration of the period-luminosity relation involves an accurate determination of the stellar distances by some independent method. Benedict et al. (2007), for example, used the *Hubble Space Telescope* to obtain the parallax and distance of 10 nearby Cepheid variables, obtaining the relation:

$$M_V = -1.62 - 2.43 \log P, \tag{14.2}$$

where P is the period in days and the uncertainty in the absolute visual magnitude M_V is ± 0.10. This is the period-luminosity relation for classical *Cepheid variables*, also known as *Population I Cepheids*, in the Milky Way.

Example: Distance to Delta Cephei
The prototype Cepheid variable star is Delta Cephei, whose variation period is $P = 5.36634$ days. The period-luminosity relation $M_V = -1.62 - 2.43 \log P$ results in an absolute visual magnitude $M_V \approx -3.40 \pm 0.01$. One could infer the star's distance, D, from the apparent magnitude, m_V, and the relation $m_V - M_V = 5 \log D - 5$, where the distance is in parsecs (Sect. 10.1). Unlike most stars, however, Delta Cephei is a variable star with an apparent magnitude that ranges from $m_V = 3.48$ to 4.37, resulting in uncertainties in the distance determination unless a mean apparent magnitude is used. Instead, the *Hubble Space Telescope* has been use to determine the parallax, $\pi_A = (3.66 \pm 0.15) \times 10^{-3}$ s of arc and a distance of $D = 1/\pi_A = 273 \pm 10$ parsecs and about 887 light-years (Benedict et al. 2002). This distance corresponds to an apparent visual magnitude $m_V \approx 3.8 \pm 0.1$.

It was the American astronomer Harlow Shapley (1885–1972) who observed Cepheids in globular star clusters outside the plane of the Milky Way. He showed that they are distributed within a roughly spherical system, which is centered far from the Sun in the direction of the constellation Sagittarius (Shapley 1918, Fig. 14.3).

These dense stellar clusters orbit the core of the Milky Way in great elongated ellipses, passing through the Milky Way every 100 million years or so. In contrast, the stars within the plane of the Milky Way go around its center in roughly circular orbits with comparable periods for stars about as distant from the core as the Sun is.

Most of the stars and interstellar gas in the Milky Way are located within this flattened, plate-shaped, rotating disk. Its center, known as the *galactic center*, is located at a distance $D_\odot \approx 8.5$ kpc, or about 2.5×10^{20} m and 27,700 light-years, from the Sun. This distance is 1.7 billion times the distance between the Earth and

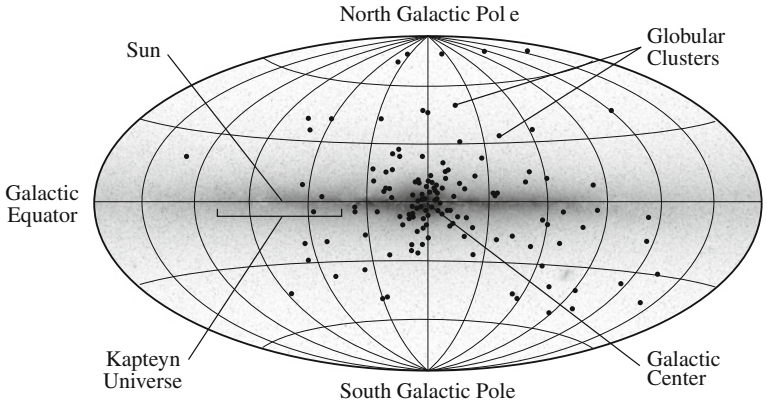


Fig. 14.3 Edge-on view of the Milky Way As shown in 1918 by the American astronomer Harlow Shapley (1885–1972), the globular star clusters are distributed in a roughly spherical system whose center coincides with the core of our Milky Way. The Sun is located in the disk, about 27,700 light-years away from the center. The disk and central bulge are shown edge-on in a negative print of an infrared image taken from the *InfraRed Astronomical Satellite*. The infrared observations can penetrate the obscuring veil of interstellar dust that hides the distant Milky Way from observation at optically visible wavelengths. It is this dust that limited astronomers' view of stars to a much smaller Kapteyn Universe, centered on the Sun

the Sun. Reid (1993) has reviewed determinations of this distance that were available at the time. Even relatively recent estimates indicate that it is quite uncertain with $D_\odot = 8.33 \pm 0.35\,\mathrm{kpc}$ (Gillessen et al. 2009).

The disk has a radius of about 50,000 light-years or 15 kpc, and a thickness around 3,000 light-years or 1 kpc. The bright, massive young stars are found in a disk that is only 120 pc or so thick, despite being over 30,000 pc across. Other older types of stars define thicker disks; some of them form disks that are 2,000 pc thick. Assuming a disk thickness of about 1,000 pc and a disk radius of 15,000 pc, the Milky Way has a volume of about 700 billion (7×10^{11}) cubic parsecs.

How many stars are in the Milky Way? The distribution of stars mapped by the *HIPPARCOS* mission indicates that the mass density of the stars near the Sun is $\rho_{disk} \approx 0.515 \times 10^{-20}\,\mathrm{kg\,m^{-3}} \approx 0.076\,M_\odot\,\mathrm{pc}^{-3}$, which is probably accurate to within a factor of two. The local mass density of main-sequence stars has, for example, been estimated at about $0.031\,M_\odot\,\mathrm{pc}^{-3}$, including those of spectral type M that account for the overwhelming majority of stars in the Milky Way (Reid et al. 2002). Assuming a uniform distribution of stars in the Milky Way disk and multiplying the lower mass density estimate by the disk volume of 700 billion cubic parsecs, we obtain a lower limit of at least 20 billion stars with a mass equal to that of the Sun, which is designated $M_\odot$. Because the stars are more concentrated toward the central region, the Milky Way most likely contains at least 50 billion to 100 billion stars like the Sun, and we can see only about 5,000 of them with the unaided eye.

The physical parameters of the galactic disk are listed in Table 14.1.

Table 14.1 Physical properties of the Milky Way disk

R_{disk} = radius of disk = 50,000 light-years $\approx$ 15,000 pc $\approx 4.6 \times 10^{20}$ m
L_{disk} = thickness of disk = 3,000 light-years $\approx$ 1,000 pc $\approx 3.0857 \times 10^{19}$ m
$D_\odot = R_0$ = Sun's distance from the center = 27,700 light-years = 8.5 kpc = 2.6×10^{20} m
$V_\odot = V_0$ = Sun's orbital velocity about center = 220 km s^{-1}
$P_\odot = P_0$ = Sun's orbital period about center = 7.6×10^{15} s = 2.4×10^8 years
M_{disk} = mass of disk = $10^{11}\, M_\odot \approx 2 \times 10^{41}$ kg
N_{disk} = number of stars in Milky Way = 100 billion or 10^{11} stars like the Sun
L_{Bdisk} = luminosity in blue band = $1.9 \times 10^{10}\, L_{B\odot} \approx 7.2 \times 10^{36}$ J s^{-1}
ρ_{disk} = mass density of disk near the Sun $\approx 0.515 \times 10^{-20}$ kg m^{-3}
$= 0.076\, M_\odot\, pc^{-3} = 0.0022\, M_\odot\, (\text{light-year})^{-3}$.
N_{disk} = number density of stars in disk near Sun = 2.59×10^{-51} m^{-3}
S_{disk} = separation of adjacent stars in disk $\approx$ 6.5 light-years $\approx$ 2 pc $\approx 6.2 \times 10^{16}$ m
Age = oldest disk stars = $(6-13.5) \times 10^9$ year = 6 to 13.5 Gyr, young stars are still forming
Oort's constants $A = (+14.82 \pm 0.84)$ km s^{-1} kpc^{-1} and $B = (-12.7 \pm 0.64)$ km s^{-1} kpc^{-1}.
Center of the Milky Way: right ascension $\alpha(2{,}000)$ = 17 h 45 m 40.04 s,
Declination $\delta(2{,}000) = -29^\circ\ 00'\ 28.1''$

Relatively young stars are found in the Milky Way disk; some are seen even in the earliest stages of formation and they are all less than 10 billion years old. These stars are designated *Population I stars*. In addition to their cosmic youth, they contain a relatively high abundance of the heavier elements, commonly called the metals. The Sun is a Population I star; the open star clusters found in the Milky Way contain Population I stars.

The *Population II stars* are found mainly outside the Milky Way, in globular star clusters. They include the oldest known stars of up to almost 14 billion years old, and these stars have a relatively low abundance of elements heavier than hydrogen or helium. A spherical aggregation of Population II, metal-poor stars also is found near the center of the Milky Way. It is mainly closer to the galactic center than the Sun. Ivezic et al. (2012) have reviewed our recent knowledge of galactic stellar populations. The first stars to be formed in the observable universe have been designated *Population III*; Heger (2012) has provided a review of metal enrichment by Population III.

Freeman and Bland-Hawthom (2002) have reviewed the formation of our Galaxy; Putman et al. (2012) summarized our knowledge of the gaseous galactic halo; and Van den Berg et al. (1996) have reviewed estimates for the age of the galactic globular cluster system.

Physical properties of this galactic spheroid are listed in Table 14.2.

Table 14.2 Physical properties of the globular cluster spheroid

R_{gs} = radius of spheroid $\geq$ 130,000 light-years = 40 kpc $\approx 1.2 \times 10^{21}$ m
M_{gs} = mass of spheroid = $(2\text{–}10) \times 10^9\, M_\odot \approx (4\text{–}20) \times 10^{39}$ kg
ρ_{gs} = mass density of spheroid $\approx 0.000\ 26\, M_\odot\, pc^{-3} \approx 1.9 \times 10^{-23}$ kg m^{-3}
L_{Bgs} = luminosity of spheroid, blue band = (1–2) x $10^9\, L_\odot \approx (4\text{–}8) \times 10^{35}$ J s^{-1}

14.1.3 The Rotating Galactic Disk

As suggested by the German philosopher Immanuel Kant (1724–1804) near the end of the eighteenth century, the flattened shape of the Milky Way can be attributed to its formation from a large, collapsing, rotating nebula, much like the origin of our solar system from a considerably smaller nebula (Kant 1755). Observations of the motions of nearby stars and interstellar gas in the galactic disk indicate that the entire stellar system indeed is rotating around a remote axis that pierces the center of the Milky Way. The enormous mass at this central hub steers stars into circular orbits with an orbital speed that decreases with increasing distance from the center – all in accordance with Kepler's third law.

It is rotation that has flattened the Milky Way. Observations of the motions of nearby stars and interstellar gas indicate that the entire system is whirling about a remote, massive center. The enormous mass at this central hub steers the stars into circular orbits with an orbital speed that decreases with increasing distance from the center, at least as far as the Sun, all in accordance with Kepler's third law (Lindblad 1925).

Stars in orbits inside the solar orbit travel faster than the Sun, thereby forging ahead of it, whereas the stars moving in orbits outside the Sun's orbit are falling behind. When viewed from the Earth, nearby stars that are a little closer than the Sun to the center therefore seem to move in one direction, whereas those a little farther away appear to move in the opposite direction, in two star streams (Kapteyn 1905; Joy 1939).

We can measure the radial, or line-of-sight, component of a star's motion by observing the Doppler shift in the wavelength of its spectral lines. Moreover, radio astronomers can use the same Doppler effect with the spectral line of interstellar hydrogen atoms, emitted at a wavelength of 21 cm, to trace out the motions of interstellar gas. Both techniques indicate that the Sun and nearby gas and stars are revolving about the distant center of the Milky Way at a speed of $V_\odot \approx 220$ km s^{-1}.

The period $P_\odot$ for one rotation of the Sun around the center is $P_\odot = 2\pi R_0/V_\odot \approx 7.6 \times 10^{15}$ s $\approx 2.4 \times 10^8$ years, where 1 year $= 3.1557 \times 10^7$ sec and $R_0 = D_\odot = 8.5$ kpc $= 2.6 \times 10^{20}$ m $= 27{,}700$ light-years.

The stars are moving with speeds that are greater than those of the planets that orbit them. The orbital speed of the Earth around the Sun, for example, is about 30 km s^{-1} and that of Mercury is just 48 km s^{-1}. The orbiting planets accompany their planet star in its faster motion through space.

Stars and interstellar gas rotate differentially, revolving about the galactic center in independent orbits at speeds that vary with distance. Stars that are nearer to the center of the Milky Way than the Sun revolve about the center at faster speeds and take less time to circle it. Astronomers describe this *differential orbital motion* by specifying Oort's constants (Focus 14.2).

Focus 14.2 Differential rotation of the Milky Way

The Dutch astronomer Jan Oort (1900–1992) provided observational evidence for the differential rotation of stars and interstellar matter in the Milky Way, and described their circular motion about a distant galactic center (Oort 1927, 1928). The Doppler shift velocity along the line-of-sight, or the observed radial velocity, V_r, of a galactic object at the distance, R, from the galactic center is given by:

$$V_r = R_0[\omega(R) - \omega(R_0)]\sin l, \tag{14.3}$$

where R_0 is the distance of the Sun from the galactic center, $\omega(R)$ is the circular velocity of the Milky Way at R, and l denotes the galactic longitude (Fig. 14.4). The observed proper motion μ of the celestial object in galactic longitude and at distance D less than R or R_0 is:

$$\mu = \frac{1}{4.74}\left[-\frac{1}{2}R_0\left(\frac{d\omega}{dR}\right)_{R=R_0}\cos 2l - \frac{1}{2}R_0\left(\frac{d\omega}{dR}\right)_{R=R_0} - \omega(R_0)\right] = \frac{V_\perp}{D}, \tag{14.4}$$

where the velocity $V_\perp$ transverse to the line of sight is in units of km s^{-1}, the distance is in parsecs and the proper motion is in seconds of arc (Sect. 4.2).

Oort's constants, A and B, are given by:

$$A = -\frac{1}{2}R_0\left(\frac{d\omega}{dR}\right)_{R=R_0} = \frac{1}{2}\left[\frac{V_0}{R_0} - \left(\frac{dV}{dR}\right)_{R=R_0}\right] \tag{14.5}$$

$$= 14.82 \pm 0.84 \text{ km s}^{-1} \text{ kpc}^{-1}$$

and

$$B = -\frac{1}{2}R_0\left(\frac{d\omega}{dR}\right)_{R=R_0} - \omega(R_0) = -\frac{1}{2}\left[\frac{V_0}{R_0} + \left(\frac{dV}{dR}\right)_{R=R_0}\right] \tag{14.6}$$

$$= -12.7 \pm 0.64 \text{ km s}^{-1} \text{ kpc}^{-1},$$

where the numerical values are derived from *HIPPARCOS* observations (Feast and Whitelock 1997), which also give the distance of the Sun from the galactic center, R_0, as

$$R_0 = 8.5 \pm 0.5 \text{ kpc} \approx 2.6 \times 10^{20} \text{ m} \approx 27{,}700 \text{ light-years}, \tag{14.7}$$

and a rotation velocity of:

$$V_0 = R_0(A - B) \approx 234 \text{ km s}^{-1}. \tag{14.8}$$

Blauuw et al. (1960) reviewed the IAU system of galactic coordinates. In 1986, the International Astronomical Union adopted the standard values

of $R_0 = 8.5$ kpc and a rotational velocity of the Sun about the galactic center, V_0, given by

$$V_0 = R_0(A - B) \approx 220\,\mathrm{km\,s^{-1}}, \tag{14.9}$$

(Kerr and Lynden-Bell 1986), which implies $A - B = 25.9$ km s^{-1} kpc^{-1}, not quite in accord with the *HIPPARCOS* result that would give $V_0 = 231$ km s^{-1} at $R_0 = 8.5$ kpc.

Oort's constants can be used to determine:

$$V_r = -2A(R - R_0)\sin l \;\text{ for } R - R_0 \ll R_0 \tag{14.10}$$

and

$$\mu = \frac{1}{4.74}[B + A\cos 2l]. \tag{14.11}$$

For a nearby object in the galactic plane at a distance D of less than one kiloparsec from the Sun, the radial component of velocity due to differential galactic rotation is

$$V_r = A\,D\sin 2l, \tag{14.12}$$

and its transverse velocity due to differential galactic rotation is:

$$V_\perp = D[A\cos(2l) + B]. \tag{14.13}$$

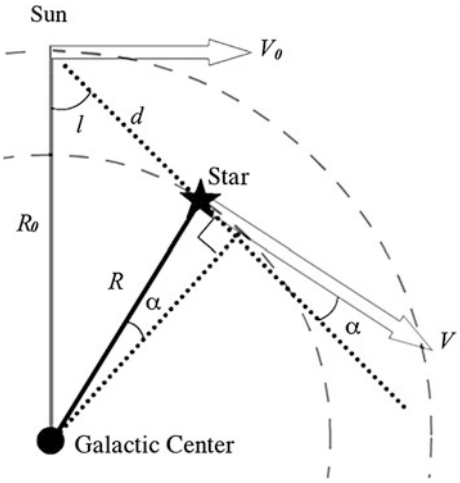


Fig. 14.4 Differential rotation of the Milky Way The Sun rotates about the galactic center, located at a distance R_0 of about 8.5 kpc or 27,700 light-years and a speed of about 220 km s^{-1}. Another star that lies closer to the center revolves about it at a distance R and a different, faster speed. This diagram provides the geometry for deriving Oort's constants $A = 15$ km s^{-1} kpc^{-1} and $B = -10$ km s^{-1} kpc^{-1} that describe this differential rotation about the center of our galaxy

We can use Kepler's third law with measurements of the Sun's distance from and velocity about the remote center of the Milky Way to infer the mass that gravitationally controls solar motion. That is, the galactic disk within the orbit of the Sun has a central mass, M_{cdisk}, of:

$$M_{cdisk} = \frac{4\pi^2 D_\odot^3}{G P_\odot^2} = \frac{V_\odot^2 D_\odot}{G} \approx 1.9 \times 10^{41}\ \text{kg} \tag{14.14}$$

where the gravitational constant $G = 6.674 \times 10^{-11}\ \text{m}^3\ \text{kg}^{-1}\ \text{s}^{-2}$, $P_\odot \approx 7.6 \times 10^{15}$ s, $V_\odot = V_0 \approx 220$ km s^{-1}, and $D_\odot = R_0 \approx 2.6 \times 10^{20}$ m. That is equivalent to about 100 billion solar masses, or $10^{11}\ M_\odot$, where the mass of the Sun is $M_\odot = 1.989 \times 10^{30}$ kg. So there is a mass equivalent to about 100 billion stars like the Sun within the Sun's galactic orbit. Fish and Tremaine (1991) have provided a review of the mass of the Galaxy.

The luminosity of our galactic disk, L_{Bdisk}, in blue light is about 25 billion times that of the Sun, or $L_{Bdisk} = 2.5 \times 10^{10}\ L_{B\odot}$, where the blue luminosity of the Sun is $L_{B\odot} = 3.0 \times 10^{26}$ J s^{-1}. These stars shine with the light of 25 billion Suns. Their combined absolute magnitude is $M_{Bdisk} = -20.5$, and their mass to luminosity ratio, $M_{disk}/L_{disk} \approx 4\ M_\odot/L_\odot$.

14.1.4 Whirling Coils of the Milky Way

The stars do not reside in a uniform whirling disk. They instead are concentrated into arms that coil out from the center of the Milky Way, giving our stellar system a spiral shape. These features are delineated by relatively young, very luminous, and massive stars (Morgan et al. 1952; Georgelin and Georgelin 1976; Paladini et al. 2004), which light up the nearby arms (Fig. 14.5). They coincide with the well-known emission nebulae, or H II regions, which are less than a few million years old and at least a thousand times younger than the oldest stars in the Milky Way. This suggests that recent star formation takes place in the spiral arms of the Milky Way.

Because the Sun is embedded in one of the arms, astronomers must look through that arm to see the rest of the Milky Way. This obscures their distant vision, hiding most of our stellar system from view in optically visible light. However, radio waves pass unimpeded through the obscuring material, permitting the detection of most of the Milky Way. This is because long radio waves are not absorbed by the relatively small particles of interstellar dust.

By observing the radio emission of interstellar hydrogen atoms at a wavelength of 21 cm, radio astronomers constructed a face-on view of the Milky Way, which we might see if we were transported into distant space and looked down on the plane of the Milky Way from above (Fig. 14.6). They delineated extensive, arm-like concentrations that extend out from the short segments defined by young massive stars in the vicinity of the Sun (Oort et al. 1958).

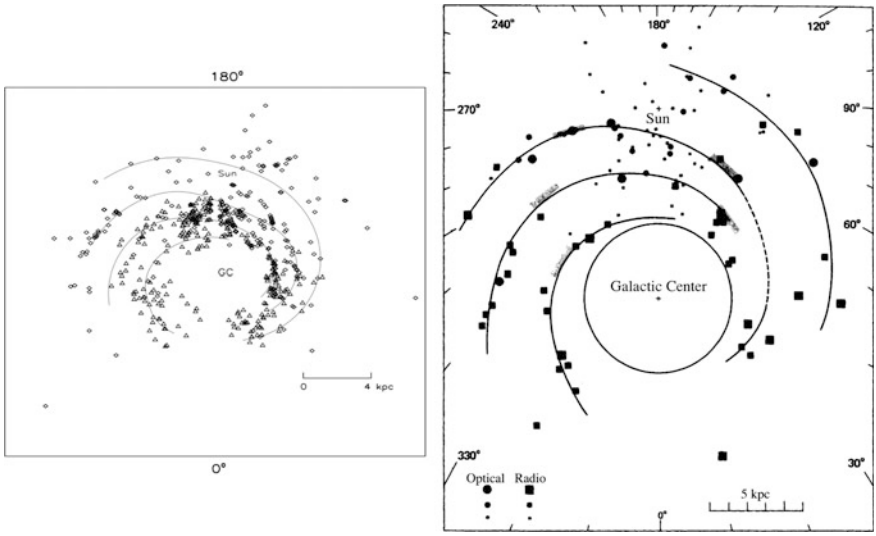


Fig. 14.5 Spiral arms of the Milky Way from H II regions Luminous emission nebulae, known as H II regions, act like beacons that mark out the spiral structure of the Milky Way. The H II regions have lifetimes of just a few million years, which is thousands of times less than the ages of the oldest stars in our Milky Way. This suggests that stars are now formed in the spiral arms of the Milky Way. The center of both diagrams coincides with the center of the Milky Way, labeled as the galactic center or GC, and the galactic longitude is indicated along the figure edges with 180° at center top and 0° at center bottom. The linear scales, shown in the lower right of each diagram, are set at 5 kpc $\approx$ 16,000 light years (*right*) and 4 kpc $\approx$ 13,000 light-years (*left*). The Sun is located at the upper center of both diagrams, and H II regions are denoted by filled circles and squares (*right*) and diamonds and triangles for 550 objects (*left*). [Adapted from (*right*) Y. M. Georgelin and V. P. Georgelin, "The spiral structure of our Galaxy determined from H II regions," *Astronomy and Astrophysics* **49**, 57–69 (1976) and (*left*) R. Paladini R. D. Davies and G. DeZotti, "Spatial Distribution of Galactic H II regions," *Monthly Notices of the Royal Astronomical Society* **347**, 237–245 (2004).]

The Sun has circled the center of the Milky Way more than 19 times during the Sun's 4.6-billion-year lifetime. So, the spiral arms should have wrapped around the massive center many times during the lifetime of the Sun. A persistent dilemma has been why they haven't wound up forming a featureless ball of gas, dust and stars. The explanation seems to be density waves that control the concentrations of stellar and interstellar material (Lin and Shu 1967).

The wave pattern orbits the galactic center at a steady rate and does not wind up; it moves independently of the motions of individual stars, which follow their own orbit around the center. The spiral arms are places where the interstellar material and stars linger – like traffic at a stoplight – and they mark the locations where new stars tend to form and hot, massive, luminous, young stars are found.

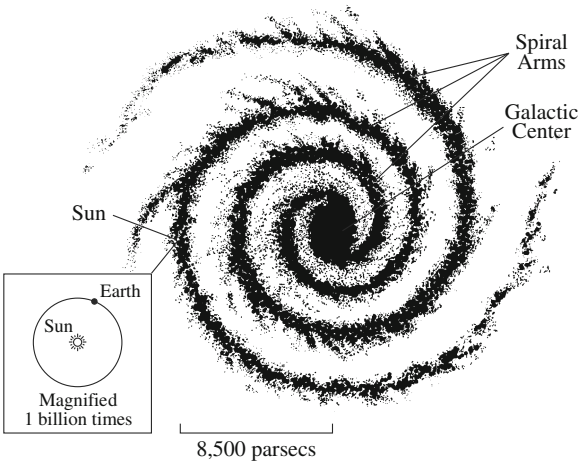


Fig. 14.6 Structure of our stellar system This drawing depicts our Milky Way as viewed from above its plane. The stars and interstellar material are concentrated within spiral arms. The Sun lies within one of these spiral arms at a distance of 27,700 light-years, from the center, designated here as 8,500 pc, or 8.5 kpc. This distance is 1.75 billion times the distance between the Earth and the Sun

14.1.5 A Central Super-Massive Black Hole

Radio astronomers have looked right through interstellar dust and detected an exceptionally powerful and compact radio source at the center of the Milky Way. It is in the direction of the constellation Sagittarius, and therefore been named *Sagittarius A** (abbreviated Sgr A* and pronounced "Sag A" star).

Radio interferometer measurements with very long baselines (VLBI) show that Sgr A* is smaller than our planetary system, with a radius of about half the distance between the Earth and the Sun, which is $1\ \text{AU} = 1.496 \times 10^{11}$ m (Doeleman et al. 2008; Reynolds 2008). It seems likely that an exceptionally massive black hole is energizing the extremely bright, compact radio source.

Because such a black hole is very massive, dense, and compact on a cosmic scale, its formidable gravity can dominate a star's motion if it is close enough. The *super-massive black hole* guides nearby stars into rapid orbital motion, betraying its presence. These stars can be seen at infrared wavelengths that also penetrate interstellar dust.

Astronomers have used large visible-light telescopes in Chile and Hawaii to detect individual stars in the infrared, although they shine too faintly at radio wavelengths to be observed with radio telescopes. By watching the motions of infrared stars that are near the center of the Milky Way and orbit it, a central super-massive black hole has been found.

Melia and Falcke (2001) reviewed evidence, available at the time, for a super-massive black hole at the galactic center, whereas Genzel and Townes (1987) provided an earlier review of the center of our Galaxy.

After an unprecedented study lasting more than a decade, researchers were able to track the full revolution of one infrared star, designated S2, around the invisible black hole. This star moves within 17 light-hours of the unseen center, with speeds of up to 5,000 km s^{-1} (Ghez et al. 2005, 2008). These orbital parameters imply

that the mass of the central black hole is a colossal 4 million times the mass of the Sun – that is, 4 million invisible solar masses not shining but rather gravitationally confining the observed stellar orbit.

Example: Super-massive black hole at the galactic center
After more than a decade of observations, Andrea Ghez (1965–) and her colleagues reported measurements from the W.M. Keck 10 m telescopes describing the elliptical orbit of the star S0-2 about the galactic center (Ghez et al. 2005, 2008). Its orbit passes within a distance of $D = 17$ light-hours $= 1.83 \times 10^{13}$ m of the black hole, at which time it accelerates to a speed of $V = 5 \times 10^6$ m s^{-1}. If we assume that this velocity is equal to the escape velocity, V_{esc}, of the black hole, of mass, M_{BH}, controlling the orbit at that time, then we can estimate the mass from

$$M_{BH} = \frac{DV_{esc}^2}{2G}, \tag{14.15}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. We obtain $M_{BH} \approx 3.4 \times 10^{36}$ kg $\approx 1.7 \times 10^6$ $M_\odot$ where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg.

Detailed comparison with the stellar orbit indicate that $M_{BH} = (4.1 \pm 0.6) \times 10^6$ $M_\odot$. Using this mass, we can determine the Schwarzschild radius, R_S, of the black hole from:

$$R_S = \frac{2GM_{BH}}{c^2} \approx 1.21 \times 10^{10}\,\text{m} \approx 0.08\ \text{AU}, \tag{14.16}$$

where the speed of light $c = 2.9979 \times 10^8$ m s^{-1} and the mean distance between the Earth and the Sun is 1 AU $= 1.496 \times 10^{11}$ m. VLBI observations at 1.3 mm wavelength indicate that the central radio source Sagittarius A* subtends an angle, θ, of $\theta \approx 4 \times 10^{-5}$ s of arc $\approx 2 \times 10^{-10}$ radians, where 1 radian $= 2.063 \times 10^5$ s of arc (Doeleman et al. 2008). The linear radius R corresponding to this angular size is $R = \theta D \approx 1.7 \times 10^{-6}$ pc $\approx 5 \times 10^{10}$ m ≈ 0.34 AU at the $D \approx 8.5$ kpc distance of the galactic center. That size is about 4 times the gravitational radius of 4 million solar masses.

Where did this super-massive black hole come from and why is it located at the center of the Milky Way? Its formation probably coincided with the origin of our stellar system, by the collapse of a huge rotating mass with a nucleus at the center and the flattened Milky Way spinning around it. The globular star clusters probably date back to the beginning of the collapse, about 14 billion years ago.

The super-massive black hole may have originated at the central nucleus, perhaps as the result of the gravitational collapse of an exceptionally massive cloud of gas located there. Alternatively, it may have grown by the coalescence of

smaller stellar black holes, each formed at the end of the lifetime of the first massive stars. The central black hole would have continued to gather in nearby, smaller black holes and surrounding stars and gas, with an active youth and a more sedate old age.

14.1.6 Dark Matter Envelops the Milky Way

As the result of differential rotation, stars and gas near the Sun should revolve about the galactic center at faster speeds than those at greater distances from the center. However, the stars and gas observed near the apparent edges of the Milky Way rotate at speeds that do not decrease with distance. This means that the Milky Way does not end where the light does, and that there are appreciable amounts of dark unseen matter well outside the boundary of the visible Milky Way. That dark invisible matter keeps the fast-spinning visible material connected to our stellar system.

The mass of the Milky Way within the Sun's orbit around the center of the Milky Way is roughly 100 billion, or 10^{11}, Suns. However, the rapid motions of dwarf satellite collections of stars, which revolve about the Milky Way at distances of up to 1 million light-years, indicate that a great reservoir of unseen matter envelops the observed disk of stars. A total mass of roughly 1 trillion, or 10^{12}, times the mass of the Sun and about 10 times the mass of its visible stars, is required to hold onto these dwarf systems. This invisible, massive, outer region is known as the *dark halo*. It surrounds the Milky Way and outweighs it by a factor of about 10.

Example: Mass of the dark halo
We can infer the total mass, M_G, of our stellar system under the assumption that distant, small companions are gravitationally bound to it. (Here we use the subscript G to denote our Galaxy, which is a term that is introduced later in the book.) The dwarf spheroid Leo I is, for example, located at a distance of $D = 230$ kpc, where 1 kpc $= 3.0857 \times 10^{19}$ m, and moves with a radial velocity $V_r = 177$ km s^{-1} $= 1.77 \times 10^5$ m s^{-1}. This velocity must be less than or equal to the escape velocity, V_{esc}, of our stellar system at this distance, so

$$V_r \leq V_{esc} = \left(\frac{2GM_G}{D}\right)^{1/2}, \tag{14.17}$$

or

$$M_G \geq \frac{DV_r^2}{2G}, \tag{14.18}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. Inserting the distance and radial velocity of Leo I into this equation, we obtain $M_G \geq 1.67 \times 10^{42}$ kg $\geq 0.8 \times 10^{12}$ $M_\odot$ where $M_\odot = 1.989 \times 10^{30}$ kg. That is about 20 times the total mass of stars in the Milky Way. The mass-to-luminosity ratio is about one hundred times that of stars like the Sun, or $M_G/L_G \approx 100\ M_\odot/L_\odot$, indicating that most of the mass is dark and unseen (Zaritsky et al. 1989; Kulessa and Lynden-Bell 1992; Peebles 1995).

Thus, our stellar system is held together by the gravity of *dark matter*, which is beyond the range of our vision. A similar darkness pervades and controls much of the universe, giving off neither light nor any other radiation to let us know it is there. Dark matter is studied by its gravitational influence on the motions of the stars that we can see.

Moreover, even the observable universe, the part we can see directly, is not limited to our stellar system, the Milky Way, but instead is populated by more than 100 billion galaxies, each composed of about 100 billion stars and perhaps containing 10 times as much mass in unseen dark matter. These galaxies stretch as far as the largest telescope can see – and perhaps beyond.

14.2 The Discovery of Galaxies

Long before the discovery of dark matter, Edwin Hubble (1889–1953), showed that the Milky Way does not contain everything there is, and settled an ongoing controversy about the nature of spiral nebulae. The issue was presented during the now-famous Shapley-Curtis debate over "The Scale of the Universe" during a meeting of the National Academy of Sciences on 26 April 1920 at the Smithsonian Institution in Washington, DC (Shapley and Curtis 1921). Harlow Shapley (1885–1972), of the Harvard College Observatory, defended his novel conception of a much larger Milky Way than previously had been supposed, with a distant center and the Sun at its periphery, but he supposed that the spiral nebulae are embedded in the Milky Way. In contrast, Heber D. Curtis (1872–1942), of the Lick Observatory, attempted to defend a smaller Sun-centered stellar system but provided cogent arguments that the spiral nebulae are distant stellar systems located far beyond the Milky Way (Curtis 1919). Shapley was correct about the shape and size of the Milky Way, and Curtis was correct in supposing that the spiral nebulae are distant "island universes" composed of numerous stars.

The argument over the location of the spiral nebulae was finally and definitely resolved when Hubble used the 2.5 m (100 inch) Hooker telescope on Mount Wilson to photograph the spiral nebula Andromeda, or M 31 (Fig. 14.7), night after night. He compared hundreds of photographs to find Cepheid variable stars whose brightness waxed and waned like clockwork in a period of several days.

Fig. 14.7 The Andromeda Nebula The nearest spiral galaxy, the Andromeda Nebula, also known as M 31 and NGC 224, is located at a distance of about 800 kpc or 2.6 million light-years, so its light takes about 2.6 million years to reach us. Both the Andromeda Nebula and our Galaxy are spiral galaxies with total masses of about 1 million million, or 10^{12}, solar masses, and roughly 100 billion, or 10^{11}, optically visible stars. The several distinct stars surrounding the diffuse light from Andromeda are stars within our own Galaxy; these stars lie well in front of Andromeda. Two smaller galaxies also are shown in this image: M 32, also designated NGC 221, at the edge of the Andromeda Nebula, and NGC 205, that is located somewhat farther away. These are elliptical systems at about the same distance as M 31 but with only about 1/100th of its mass. (Courtesy of Karl-Schwarzschild Observatorium, Tautenburg.)

On New Year's Day 1925, Hubble's results were read *in absentia* at a meeting of the American Astronomical Society in Washington, DC and caused an overwhelming sensation. The landmark paper, titled *Cepheids in Spiral Nebulae* (Hubble 1925), combined the known period-luminosity relation of Cepheids with observations of the variable stars in M 31 and M 33, another spiral nebula, to derive a distance of 0.275 Mpc = 275 kpc for the two spiral nebulae, where 1 Mpc = 3,260,000 light-years = 1,000 kpc = 3.09857×10^{22} m. Their size or linear extent, determined from this distance and their angular extents, was roughly comparable to that of our Milky Way. They had to be remote objects ablaze with stars, galaxies in their own right and separated from the Milky Way by wide gulfs of apparently empty space.

It took so long to establish the true nature of the spiral nebulae because the method of establishing the distances of the remote, luminous Cepheid variable stars needed to be developed, and a large powerful telescope was needed to detect

the stars in spiral nebulae and collect enough of their faint starlight for a reliable measurement of their periodic brightness variation and, therefore, distance.

The notable Estonian astronomer Ernst Öpik (1893–1985) had already contributed to the notion that spiral nebulae lie outside the confines of the Milky Way (Öpik 1922). He used F. G. Pease's (1881–1938) measurements of the rotation velocities of M 31 to show that its distance has to be about 0.480 Mpc if its mass to luminosity ratio is comparable to that of stars in our the Milky Way. Öpik was much more nearly correct than Hubble in estimating the distance to Andromeda, whose current distance estimate is about 0.788 Mpc. As Walter Baade (1893–1960) showed nearly three decades later, the period-luminosity relation that Hubble used had been incorrectly calibrated.

Example: Öpik's calculation of the distance to M 31
If a star in the Andromeda spiral nebula is located at a distance R from the center of the nebula, with a rotational velocity V_{rot}, then

$$V_{rot}^2 = \frac{2GM}{R} = \frac{4GM}{\theta D} = \frac{4GlD^2}{\theta D}\left(\frac{M}{L}\right), \tag{14.19}$$

where M is the mass of the galaxy within radial distance R, the angular diameter of the nebula is θ, the distance of the Andromeda nebula from the Earth is D, so $\theta = 2R/D$, and its apparent luminosity is $l = L/D^2$ for an absolute luminosity L. Collecting terms, we can specify the distance to Andromeda by

$$D = \frac{\theta V_{rot}^2}{4Gl}\left(\frac{M}{L}\right)^{-1}. \tag{14.20}$$

By assuming the mass to luminosity ratio M/L was the same as that of the Sun, Öpik (1922) was able to derive a distance of $D = 0.480$ Mpc using the angular diameter and rotational velocity of the Andromeda Nebula measured by F. G. Pease (1918), and the apparent magnitude of Andromeda, at $m = 3.44$ to infer the apparent luminosity.

Once the observed periodic variations of the Cepheid variable stars was correctly calibrated, the distance to Andromeda turned out to be 2.54 million light-years, about three times farther away than Hubble initially supposed – but his dramatic conclusion remained unchanged. Hubble broke through the stars, and moved the outer boundaries of the universe far out into space, enlarging our horizons. The universe was no longer limited to the objects our unaided eyes perceive, and our stellar system had become just one of myriad galaxies located far beyond the Milky Way – which became our Galaxy, written with an uppercase G to show that it is special. All of the other galaxies were shown to be extragalactic, or outside of our Galaxy.

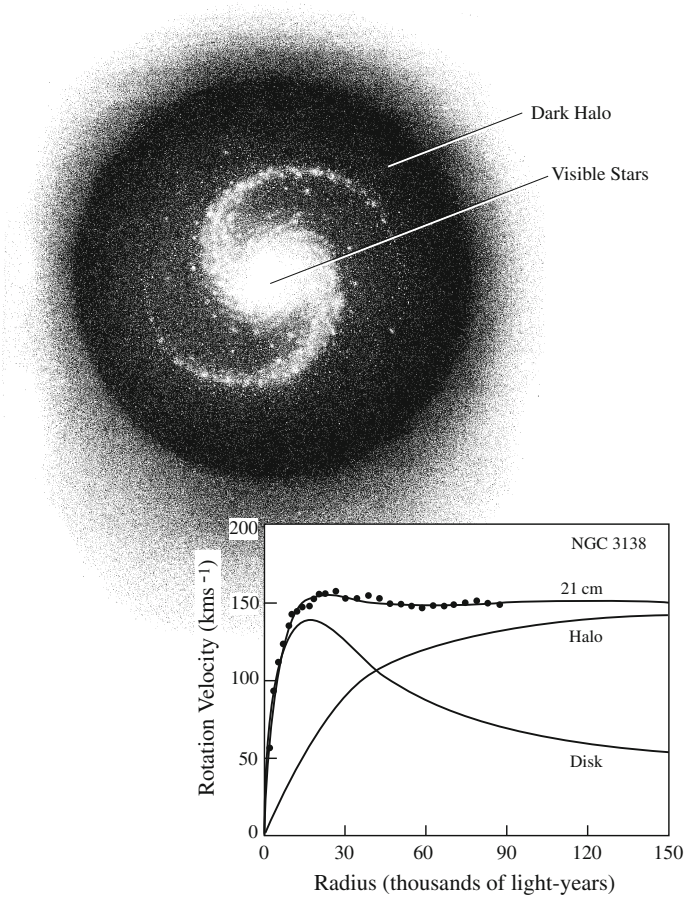


Fig. 14.8 Dark matter envelops a spiral galaxy The rotation velocity of the spinning spiral galaxy NGC 3198 plotted as a function of radius from its center (*bottom*). The observed neutral hydrogen 21 cm data are attributed to an optically luminous disk, containing all the visible stars, and a dark halo that contributes most of the mass at distant regions from the center. The visible stars and surrounding halo of dark matter are illustrated in a hypothetical drawing of a spiral galaxy seen from above (*top*). The fact that the rotational speed of the cool hydrogen gas remains high even at the largest distances indicates that the outermost gas must be constrained and held in by the gravitational pull of dark matter far outside the visible part of the galaxy. [Adapted from T. S. Van Albada et al. "Distribution of dark matter in the spiral galaxy NGC 3198," *Astrophysical Journal* **295**, 305 (1985).]

Hubble called the spirals extragalactic nebulae, but they are not large gaseous nebulae. Hubble (1926, 1936) has described our early knowledge of extra-galactic nebulae, now known as the galaxies.

The spectral characteristics of the light from the spirals are similar to that of our Sun, indicating stellar temperatures of thousands of K. If a spiral were filled with gas at this temperature throughout its enormous dimensions, the nebula would be more

than 1 billion, billion times more luminous than the Sun, rather than the much smaller luminosity of Andromeda at about 100 billion times the solar luminosity. This means that light is not coming from the entire surface of a spiral nebula but instead from individual stars, separated by vast spaces without any stars.

Therefore, we now have dropped the nebula designation and use the term *spiral galaxy* or *elliptical galaxy*, depending on their shape. Each galaxy contains about 100 billion stars, just like our Galaxy. The designation *nebula* now is reserved for cloudy, gaseous material enveloping bright stars. An exception is the famous Andromeda Nebula, the closest spiral galaxy.

Like our Milky Way Galaxy, the other galaxies contain about ten times more dark matter then optically visible matter (Fig. 14.8). Sofue and Rubin (2001) have reviewed rotation curves of spiral galaxies; references to individual papers about dark matter that envelops galaxies are given in Lang (1999).

When looking up at the night sky, we see only stars and the black spaces between them. The galaxies are out there, but we cannot see them without a telescope. They are so far away that their brightness is below the detection threshold of the human eye. This meant that the universe was no longer limited to the things our unaided eyes can focus on, and our stellar system became just one of myriads of galaxies.

Blanton and Moustakas (2009) have summarized our knowledge of the physical properties and environments of nearby galaxies. Faber and Gallagher (1979) provided an earlier review of the masses and mass-to-light ratios of galaxies, while Binggeli et al. (1988) reviewed their luminosity function. Binney (1982) has reviewed the dynamics of elliptical galaxies. Mateo (1998) provided a review of dwarf galaxies in the Local Group of galaxies. Van Der Kruit and Freeman (2011) have reviewed galaxy disks.

Some properties of these galaxies are listed in Table 14.3.

The light we receive from the most distant galaxies was emitted before the Earth and the Sun were formed. Even more fantastic, they are all in flight, rushing away from us at speeds that increase with their distance. Nearly a decade before Hubble's determination of the distance of Andromeda, Vesto M. Slipher (1875–1969) already had helped us move beyond the stars in an entirely unsuspected way.

14.3 The Galaxies are Moving Away from us and from Each Other

At the time of their discovery in enormous numbers, most astronomers thought that the spiral nebulae were nascent planetary systems, not galaxies. The bright center was supposed to be a newborn star, and the spiral arms surrounding it were thought to be developing planets, whirling and rotating around the central star just as the Earth revolves around the Sun.

Table 14.3 Physical properties of galaxies[a]

R_g = radius of a spiral or elliptical galaxy $\approx$ 33 light-years to 326 light-years = 10 kpc to 100 kpc $\approx$ (3–30) $\times$ 10^{20} m

M_S = mass of a spiral galaxy $\approx$ (10^{11}–10^{12}) $M_\odot$ $\approx$ (2–20) $\times$ 10^{41} kg

M_E = mass of elliptical galaxy $\approx$ (10^{12}–10^{13}) $M_\odot$ $\approx$ (2–20) $\times$ 10^{42} kg

M_S/L_B = mass to light ratio of spiral galaxy within radius $R \approx 60\ h\ (R/0.1\ \text{Mpc}) \approx 42\ (R/0.1\ \text{Mpc})$

M_E/L_B = mass to light ratio of elliptical galaxy within radius $R \approx 200\ h\ (R/0.1\ \text{Mpc}) \approx 140\ (R/0.1\ \text{Mpc})$

L_B = mean galaxy luminosity density in blue band = $1.93 \times 10^8\ h\ L_{B\odot}\ \text{Mpc}^{-3} \approx 3 \times 10^{34}\ \text{J s}^{-1}\ \text{Mpc}^{-3}$

L_g = mean galaxy luminosity density = (2–3) $\times$ 10^8 h $L_\odot\ \text{Mpc}^{-3} \approx$ (5–8) $\times$ $10^{34}\ \text{J s}^{-1}\ \text{Mpc}^{-3}$

L_x = x-ray luminosity of spiral or elliptical galaxy = 10^{31}–$10^{35}\ \text{J s}^{-1}$

M_{Sgas} = mass of cool gas of hydrogen atoms and molecules in spiral galaxy $\approx$ 10^9–10^{10} $M_\odot$ $\approx$ (2–20) $\times$ 10^{39} kg

M_{Egas} = mass of hot gas in elliptical galaxy $\approx$ 10^9–10^{10} $M_\odot$ $\approx$ (2–20) $\times$ 10^{30} kg

N_g = average volume density of galaxies = $5.52 \times 10^{-2}\ h^3\ \text{Mpc}^{-3} \approx 1.89 \times 10^{-2}\ \text{Mpc}^{-3}$

ρ_g = mass density of galaxies $\approx 0.37 \times 10^{-26}\ h^3\ \text{kg m}^{-3} \approx 1.5 \times 10^{-27}\ \text{kg m}^{-3}$ (for visible stars and dark matter with a galaxy mass of about 10^{12} solar masses.)

[a] For Hubble constant $H_0 = 100\ h\ \text{km s}^{-1}\ \text{Mpc}^{-1} \approx 75\ \text{km s}^{-1}\ \text{Mpc}^{-1}$ with $h \approx 0.75$. The mass of the Sun is $M_\odot = 1.989 \times 10^{30}$ kg and the total luminosity of the Sun is $L_\odot = 3.828 \times 10^{26}\ \text{J s}^{-1}$ and its luminosity in the blue region of the spectrum is $L_{B\odot} = 1.9 \times 10^{26}\ \text{J s}^{-1}$

Percival Lowell (1855–1916), a wealthy Bostonian, had built an observatory in Tucson, Arizona, primarily to detect canals on Mars supposedly built by parched, industrious Martians. Lowell also believed that the spiral nebulae resemble our solar system in its early formative stages; he therefore instructed his staff astronomer, Vesto M. Slipher (1895–1916), to measure their rotations, hoping to gather insight about our own planetary system.

When using the 0.6 m (24 inch) refractor at the Lowell Observatory in Flagstaff, Arizona, to record the spectra of bright spiral nebulae and measure their rotations, he found that they almost unanimously are moving away from us at high velocities. They were also rotating, and a few were approaching – but at modest speeds in comparison to the outward motion of most spirals. The Andromeda Nebula, for example, was moving toward the Earth at an apparent velocity of 300 km s^{-1} (Slipher 1914). However, the other bright spirals were moving in the opposite direction, usually with higher velocities of up to 1,100 km s^{-1}, much faster than any star in the Milky Way (Slipher 1917).

By 1917, Slipher had accumulated spectra of 25 spiral nebulae, using the Doppler effect to measure their radial velocities, and he showed that none of them are at rest (Slipher 1917). All but three were rushing away from us and from each other, dispersing, moving apart, and occupying an ever-increasing volume.

According to Slipher, the observed motions of the majority of spirals indicated a general fleeing from the Milky Way or us. It certainly was difficult to believe that

objects with such enormous speeds could long remain a part of our stellar system. The combined gravitational pull of the entire 100 billion stars in the Milky Way is not enough to retain any spiral nebula moving at speeds in excess of 1,000 km s^{-1}.

Example: Escape velocity of the Milky Way
The escape velocity $V_{esc} = (2GM/R)^{1/2}$ of the stellar Milky Way can be determined by using $R = R_0 = 8.5$ kpc $\approx 2.6 \times 10^{20}$ m, the distance of the Sun from the center of the Milky Way, and $M = 10^{11} M_\odot$ where the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg, as inferred from the Sun's orbital velocity around the center. The gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. The inferred escape velocity of the stellar Milky Way is about $V_{esc} \approx 3.2 \times 10^5$ m s^{-1} $\approx$ 320 km s^{-1}. The recession velocities measured by Vesto Slipher for spiral nebulae exceeded this escape velocity by up to four times. When the dark matter that envelops the Milky Way is taken into account, the total mass increases to $M \approx 10^{12} M_\odot$ and the escape velocity to about 1,000 km s^{-1}, but most spiral nebulae were soon found to have recessional velocities exceeding this amount.

By 1929, Hubble showed that the measured distances of spirals, which he had established using the superb light-gathering power of the 2.5 m (100 inch) Hooker telescope, were roughly correlated with Slipher's velocities (Hubble 1929). The comparison indicated that the farther a spiral is the faster it is moving away from us. This relationship now is attributed to the expanding universe, which no one had anticipated at the time Slipher made his measurements.

In his publication of these results, titled *A Relation between Distance and Radial Velocity Among Extra-Galactic Nebulae*, Hubble drew a straight line through a plot of the observed data. However, there was a wide dispersion between the plotted points and only a mild tendency for velocity to increase with distance (Fig. 14.9). Nevertheless, his conclusion subsequently was confirmed by more comprehensive observations of a much greater number of galaxies.

The discovery of the expanding universe also explained why the night sky is dark, resolving Olbers' paradox, named for the German astronomer Heinrich Wilhelm Olbers (1758–1840). He realized that the night sky in an infinite, uniform, non-expanding universe should be covered with stars shining as bright as the Sun (Jaki 1969). The expansion of the universe redshifts the most intense light of distant galaxies, and their stars, out of the visible part of the spectrum; therefore, we do not see them and the paradox is resolved.

The connection between velocity and distance is known now as the Hubble law, and the ratio of the velocity of recession of any galaxy and its distance from us is called now the Hubble constant, a fundamental measure of the universe. It is designated by the symbol H_0, in which the H is for Hubble and the zero subscript denotes its current value.

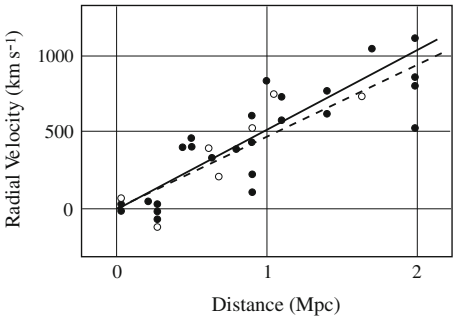


Fig. 14.9 Discovery diagram of the expanding universe A plot of the distance of extragalactic nebulae, or galaxies, versus the radial velocity at which each galaxy is receding from the Earth, published in 1929 by the American astronomer Edwin Hubble (1889−1953). The linear relationship between the distance and radial velocity indicates that the universe is expanding. Vesto M. Slipher (1875−1969) determined most of these velocities more than a decade before this diagram was drawn. Here, the velocity is in units of kilometers per second, abbreviated km s^{-1}, and the distance is in units of millions of parsecs, or Mpc, where 1 Mpc is equivalent to 3.26 million light-years. Hubble underestimated the distances of the spiral nebulae; therefore, the distance scale for modern versions of this diagram is about seven times larger. The filled circles and solid line represent the solution for individual nebulae; the open circles and dashed line are for groups of them

The units of the Hubble constant are given in kilometers per second per Megaparsec, abbreviated km s^{-1} Mpc^{-1}, and Hubble's initial estimate was pegged at 530 in these units. The radial velocity is in units of km s^{-1}, and the distance is in Megaparsecs, abbreviated Mpc, in which 1 Mpc = 10^6 pc = 3.08568×10^{22} m is equivalent to 3.26 million light-years. Galaxies typically are separated by a few Mpc, or about 10 million light-years, which is about 100 galaxy diameters. So, the universe is largely empty space relative to galaxies.

According to the now-famous Hubble law, the radial velocity, V_r, of a galaxy, as measured by the Doppler effect, is given by the linear relation:

$$\text{Hubble's Law} = cz = H_0 \times D, \tag{14.21}$$

where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, and z is the redshift. The Hubble constant, H_0, is the ratio of the speed with which distant galaxies are receding from us to their distance, D. It is as though the expanding universe started with a gigantic explosion, with the fastest-moving parts having traversed the greatest distances. The redshift is an observational parameter defined by:

$$z = \frac{\lambda_{observed} - \lambda_{emitted}}{\lambda_{emitted}}, \tag{14.22}$$

which means that

$$1 + z = \frac{\lambda_{observed}}{\lambda_{emitted}} \tag{14.23}$$

for a spectral line emitted at the wavelength $\lambda_{emitted}$ and observed at $\lambda_{observed}$. For relatively low velocities

$$V_r = cz = H_0 D \text{ for } V_r \ll c, \tag{14.24}$$

and for velocities comparable to the speed of light,

$$\frac{V_r}{c} = \frac{(z+1)^2 - 1}{(z+1)^2 + 1} \text{ for } V_r \approx c, \tag{14.25}$$

or equivalently:

$$1 + z = \left[\frac{1 + \left(\frac{V_r}{c}\right)}{1 - \left(\frac{V_r}{c}\right)}\right]^{1/2}. \tag{14.26}$$

In just 2 years, Hubble extended the velocity-distance relationship to substantially greater distances, with the help of Milton Humason (1891–1972), who made the velocity measurements (Hubble and Humason 1931, 1934). Because of the impossibility of measuring distances to such faint objects, Hubble and Humason simply assumed that all galaxies have the same intrinsic luminosity. They inferred a velocity-distance relationship by comparing the observed velocities of the galaxies to their apparent brightness. Thus, they reformulated Hubble's law and showed that a linear relationship between distance and recession velocity is valid, within the observational uncertainties, to distances as far as 100 Mpc and radial velocities of nearly 20,000 km s^{-1}.

Since distances can only be measured for relatively nearby extragalactic objects, Hubble's law for remote objects is sometimes expressed by the redshift-magnitude relation in which the apparent magnitude, m, is given by:

$$m = 5 \log\left(\frac{cz}{H_0}\right) + M + 25, \tag{14.27}$$

where M is the absolute magnitude, the Hubble constant H_0 is given in units of km s^{-1} Mpc^{-1} and the factor of $+25$ arises because the distance unit is in Megaparsecs, abbreviated Mpc.

Humason then teamed up with Nicholas Mayall (1906–1993) at the Lick Observatory in an ambitious 25 year project of painstakingly measuring the Doppler-effect redshifts of nearly a thousand galaxies visible from the northern hemisphere. Mayall used the venerable 0.9 m (36 inch) Crossley reflector on Mount Hamilton to observe the brighter galaxies, while Humason observed the fainter ones using the 2.5 m (100 inch) Hooker telescope on Mount Wilson.

Although Hubble had initiated the project, he died before the work was finished, and the analysis was left in the hands of his young protégé Allan Sandage (1926–2010). He obtained a value for the Hubble constant of $H_0 =$ 180 km s^{-1} Mpc^{-1}, or about one third the value previous found by Hubble, whose distance scale was in error. His mistake was not discovered until the early 1950s

when red-sensitive photographic plates, developed for military reconnaissance in World War II (1939–1945), became routinely available. Sandage then used them to discover that the brightest stars, which Hubble used to infer distances, are much more luminous emission nebulae.

The resulting redshift-apparent brightness diagram of 474 extra-galactic nebulae, as they preferred to call them, had a large scatter. But a straight line could be drawn through it, out to a radial velocity of 100,000 km s^{-1}, or one-third the speed of light (Humason et al. 1956). So Hubble's law connecting radial velocity and distance still held in every direction as far as one could see. This meant that the entire universe is expanding swiftly and evenly in all directions, with the fastest-moving parts having traversed the greatest distance.

The Hubble constant, H_0, quantifies the current rate of expansion of the universe, and it has often been quantified in the form:

$$H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}. \tag{14.28}$$

There is an ongoing controversy about the exact value of this important constant, with estimates ranging between 50 and 100 km s^{-1} Mpc^{-1}, and current observational constraints of h lying between 0.50 and 0.85, with a favored value of about 0.75 (Fig. 14.10). The systematic uncertainties in H_0 have decreased as a result of observations with the *Hubble Space Telescope* and the *Spitzer Space Telescope* with recent determinations of $H_0 = 73.8 \pm 2.4$ km s^{-1} Mpc^{-1} (Reiss et al. 2011) and $H_0 = 74.3 \pm 2.1$ km s^{-1} Mpc^{-1} (Freedman et al. 2012). Lang (1999) and Freedman and Madore (2010) have reviewed determinations of the Hubble constant, and Feast and Walker (1987) have discussed Cepheids as distance indicators.

The Hubble constant sets the physical scale of the universe, with the distance to any galaxy given by:

$$D = \frac{cz}{H_0} = 2997.9\left(\frac{z}{h}\right)\ \mathrm{Mpc}. \tag{14.29}$$

The most distant objects exhibit larger redshifts, denoted by the lowercase letter z. The largest observed redshift corresponds to the greatest distance and looks the farthest back in time. Astronomers currently are detecting galaxies out to a redshift as great as $z = 8.6$, the radiation of which was emitted about 13.1 billion years ago (Lehnert et al. 2010).

Once the distance to a galaxy is determined from its redshift and Hubble's constant, one can infer the galaxy's luminosity, L, from it apparent brightness or equivalently its absolute magnitude, M, from its apparent magnitude m. When this is done, a wide range of luminosities is determined. The number of galaxies at different luminosities peaks at a blue luminosity, L_*, given by:

$$L_* \approx 10^{10} L_{B\odot} h^{-2} \approx 2 \times 10^{10} L_{B\odot}, \tag{14.30}$$

where $h \approx 0.75$ and the luminosity of the Sun in the blue spectral region is $L_{B\odot} = 1.9 \times 10^{26}$ J s^{-1} corresponding to an absolute blue magnitude for the Sun

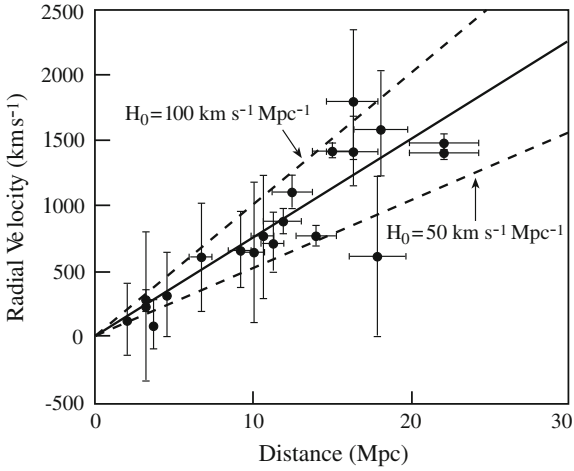


Fig. 14.10 Hubble diagram for Cepheid variable stars This plot of galaxy distance versus recession velocity is analogous to that obtained by Edwin Hubble in his 1929 discovery of the expansion of the universe (see Fig. 14.9). The slope of the linear fit (*solid line*) to the data (*dots*) measures the expansion rate of the universe, a quantity called the Hubble constant, designated H_0. The data shown here summarize 11 years of effort to measure this constant by using the *Hubble Space Telescope* to measure the distances and velocities of Cepheid variable stars in nearby galaxies. The distance is in units of 1 million parsecs, or Mpc, where 1 Mpc is equivalent to 3.26 million light-years; the radial velocity is given in units of kilometers per second, denoted as km s^{-1}. The fit to these data indicate that $H_0 = 75 \pm 10$ km s^{-1} Mpc^{-1} and that this constant lies well within the limits of 50 and 100 in the same units (*dashed lines*). [Adapted from Wendy L. Freedman et al. "Final results from the Hubble Space Telescope Key Project to measure the Hubble constant," *Astrophysical Journal* **553**, 47–72 (2001).]

of $M_{B\odot} = 5.48$. At this luminosity, each galaxy contains at least 20 billion stars like the Sun.

But galaxies can be more or less luminous than L_*, and the range in luminosity is described by a luminosity function $\phi(L)$ that specifies the number of galaxies with luminosities between L and $L + dL$. It is also called the Schechter luminosity function, and it given by (Schechter 1976):

$$\phi(L)dL = \phi_* \left(\frac{L}{L_*}\right)^{\alpha} \exp\left(-\frac{L}{L_*}\right)\frac{dL}{L_*}, \tag{14.31}$$

where ϕ_* is a normalization constant for $L = L_*$ and given by:

$$\phi_* \approx 0.02\, h^3\, \mathrm{Mpc}^{-3} \approx 0.0084\, \mathrm{Mpc}^{-3} \tag{14.32}$$

for the total galaxy population, α is the slope of the luminosity function at low luminosity $L < L_*$ where $\phi(L) \propto L^{\alpha}$ and $\alpha = -0.8$ to -1.3. At high luminosity $L > L_*$ there is an exponential cutoff with $\phi(L) \propto \exp(-L)$. In other words, the number of galaxies falls off on the low side of L_* and on the high side of it, in ways described by the two terms in the luminosity distribution.

The wide range in galaxy luminosity corresponds to a wide range in absolute blue magnitude M_B with $M_B = -7.5$ to -22.5, and a peak value M_{B*} given by:

$$M_{B*} \approx -19.7 + 5 \log (h) \approx -20.3, \tag{14.33}$$

where Hubble's constant $H_0 = 100\ h$ km s^{-1} Mpc^{-1} and $h \approx 0.75$.

The space, or volume, density of galaxies, N_g, their blue luminosity density, L_{Bg}, and their total number N_T out to redshift z can be determined by integrating the observed luminosity distribution function obtained from redshift surveys (Focus 14.3). The results indicate that:

$$N_g \approx 0.02 \text{ Mpc}^{-3} \tag{14.34}$$

$$L_{Bg} \approx 3 \times 10^{34} \text{ J s}^{-1} \text{ Mpc}^{-3} \tag{14.35}$$

and

$$N_T \approx 6 \times 10^9 z^3. \tag{14.36}$$

Altogether, there are least 10 billion galaxies in the volume of space that modern telescopes can detect and there is no end in sight.

Focus 14.3 Density and total number of galaxies

The space density of galaxies, N_g, or their density per unit volume, can be determined by adding up, or integrating, the contribution to the luminosity function $\phi(L)$ at different luminosities, L. Expressed mathematically:

$$N_g = \int_0^\infty \phi(L) dL \approx \phi_* \Gamma(\alpha + 1) \tag{14.37}$$

where the normalization constant $\phi_* \approx 0.02\ h^3$ Mpc^{-1} ≈ 0.0084 Mpc^{-3} at $L_* = 10^{10}\ L_{B_\odot} h^{-2} \approx 2 \times 10^{10}\ L_{B_\odot}$, Γ is a gamma function, and α is low-luminosity slope of the luminosity function.

The blue luminosity density, L_{Bg}, can similarly be defined by:

$$L_{Bg} = \int_0^\infty L\phi(L) dl \approx \phi^* L_* \Gamma(\alpha + 2). \tag{14.38}$$

Redshift surveys can be used to determine these parameters for field galaxies that lie beyond the local concentration of galaxies, known as the Local Group, with the results (Loveday et al. 1992):

$$N_g = 0.0552\, h^3 \text{ Mpc}^{-3} \approx 0.019 \text{ Mpc}^{-3} \tag{14.39}$$

$$L_g \approx 2 \times 10^8 \, h \, L_{B\odot} \, \mathrm{Mpc}^{-3} \approx 3 \times 10^{34} \, \mathrm{J \, s^{-1} \, Mpc^{-3}}, \tag{14.40}$$

where Hubble's constant $H_0 = 100 \, h$ km s^{-1} Mpc^{-1} and $h \approx 0.75$.

With a total galaxy mass of $10^{12} \, M_\odot$, in both visible and unseen dark matter, this corresponds to a galaxy mass density ρ_g of:

$$\rho_g = 10^{12} N_g M_\odot \approx 10^{-27} \mathrm{kg \ m^{-3}}. \tag{14.41}$$

From the Hubble law, the distance of a galaxy at redshift, z, is $D = cz/H_0$, where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. So the total number of galaxies, N_T, out to redshift z is:

$$N_T = \frac{4\pi}{3} \left(\frac{cz}{H_0} \right)^3 N_g \approx 6 \times 10^9 z^3, \tag{14.42}$$

which is independent of h and the exact value of the Hubble constant.

Since galaxies have been observed out to redshifts greater than one, there are at least 10 billion, or 10^{10}, galaxies in the observable universe. The exact shape and form of the universe complicates the precise calculations, since the curvature of space changes the distances at large redshifts. Still, our estimate should be correct to within an order of magnitude, or a factor of ten.

As Hubble realized, astronomers see only as far as their telescopes permit, eventually reaching a limit – a dim boundary to the observable universe where they measure the shadows. Even now, there is no telescope powerful enough to detect the edge where the galaxies might end. There is no edge to the observable universe and it has no detectable center.

Example: Can visible or invisible matter stop the expansion of the universe?

Throughout most of the past decades, it has been assumed that it is the mass of the universe that curves its shape, establishes its geometry, and determines its fate. Under this assumption, which ignores the more recent discovery of dark energy, the mass density of galaxies, ρ_g, determines the ultimate destiny of the universe. If this mass density exceeds a certain critical value, ρ_c, then gravity eventually overcomes expansion. Imagine the most distant galaxy with mass, m_G, distance, D_G, and velocity, $V_G = H_0 D_G$. Gravity will just balance the expansion of this galaxy if its kinetic energy of expansion is equal to the gravitational potential energy of all of the rest of the universe, or if:

$$\frac{m_G V_G^2}{2} = \frac{m_G H_0^2 D_G^2}{2} = \frac{G m_G M_U}{D_G}, \tag{14.43}$$

where M_U is the total mass of all the rest of the universe inside distance D_G and the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. In other words, the velocity, V_G, of the most distant galaxy is just equal to the escape velocity, V_{esc}, of the entire universe, $V_{esc} = (2GM_U/D_G)^{1/2}$.

Collecting terms we obtain a critical mass density of:

$$\rho_c = \frac{3M_U}{4\pi D_G^3} = \frac{3H_0^2}{8\pi G} \approx 1.88 \times 10^{-26}\, h^2\, \mathrm{m}^{-3} \approx 1.06 \times 10^{-26}\, \mathrm{kg\,m}^{-3}, \tag{14.44}$$

where $H_0 = 100\, h$ km s^{-1} Mpc^{-1} $= 3.24 \times 10^{-18}\, h$ s^{-1} $\approx 2.43 \times 10^{-18}$ s^{-1} for $h = 0.75$.

The space density of galaxies is $N_g = 0.055\, h^3$ Mpc^{-3} (see Focus 14.3), where 1 Mpc $= 3.0857 \times 10^{22}$ m. Assuming that each galaxy has a mass in both visible and unseen dark matter of $M_G = 10^{12}\, M_\odot = 1.989 \times 10^{42}$ kg and $h = H_0/100 = 0.75$, then the mass density of visible galaxies, $\rho_G = N_g M_G \approx 1.57 \times 10^{-27}$ kg m^{-3}. This is about a factor of 10 less than the critical mass density needed to stop the expansion of the universe, even when we have assumed there is ten times more dark matter than visible matter in galaxies.

14.4 Galaxies Gather and Stream Together

14.4.1 Clusters of Galaxies

Astronomers have been mapping the distribution of galaxies for about a century, determining the shape and form of the larger universe. The first cosmic maps were two-dimensional, constructed from catalogues giving the celestial positions of the brightest nebulae. Although a foreground and background galaxy might sometimes coincide, the concentrations were too pronounced to be solely due to such a superposition (Charlier 1922).

The galaxies are not placed randomly throughout expanding space. They are not uniformly strewn here and there or isolated from one another but instead knot together in great clusters that are millions of light-years across (Fig. 14.11). They also contain large additional quantities of unseen matter. A rich cluster of galaxies typically spans 10 million light-years to 20 million light-years and contains hundreds and even thousands of individual galaxies. They move within the cluster at velocities of about 1,000 km s^{-1}, on average, and the amount of time, T_C, required for a galaxy to cross the cluster moving at this speed is about a two billion years.

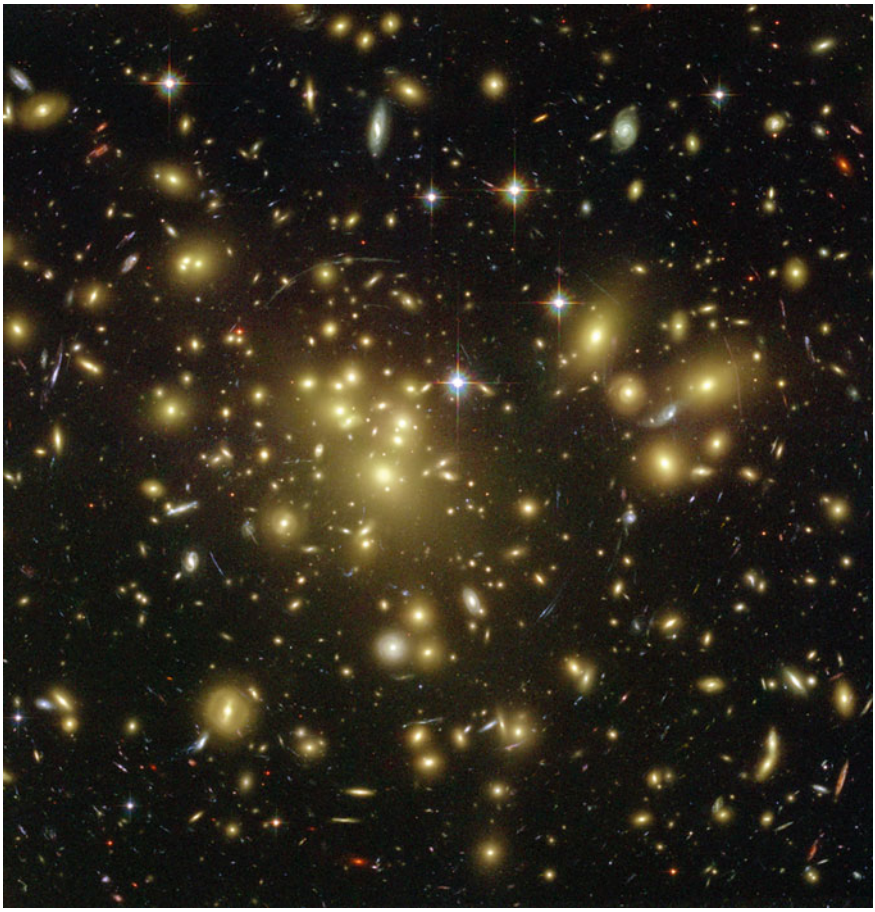


Fig. 14.11 Cluster of galaxies The *Hubble Space Telescope* provided this dramatic view of the center of a massive cluster of galaxies known as Abell 1689, located 2.2 billion light-years away. The gravity of the cluster's million million, or trillion, stars, plus any unseen matter, acts as a gravitational lens in space, bending and magnifying the light of galaxies located far behind it into radiant arcs. (Courtesy of NASA, the ACS Science team of the HST, STScI, and ESA.)

In 1958, George Abell (1927–1983), then a graduate student at Caltech, was able to describe nearly 2,712 rich clusters of galaxies (Table 14.4), using photographic surveys that included both faint and bright galaxies (Abell 1958). Even today, these dense concentrations of galaxies are referred to simply as "Abell clusters" – designated by the word "Abell" or the letter "A" followed by the number in his catalogue.

In addition, as Abell noticed, the clusters gather and congregate together into larger superclusters, which he called second-order clustering. Our Galaxy lies in the outskirts of one, known as the Local Supercluster (de Vaucouleurs 1953),

Table 14.4 Physical properties of rich clusters of galaxies[a]

N_T = total number of galaxies in a rich cluster of galaxies = 30–300
R_{cl} = central radius of galaxy cluster $\approx$ (1–2) h^{-1} Mpc $\approx$ (4.4–8.8) $\times$ 10^{22} m
N_{cl} = volume density of galaxies in galaxy cluster $\approx$ 100 Mpc^{-3}
σ_V = velocity dispersion of galaxy motions $\approx$ 100–1,400 km s^{-1} = 10^5–1.4 $\times$ 10^6 m s^{-1}
M_C = virial mass of galaxy cluster = $\sigma_v^2 R_{cl}/G \approx$ (10^{14}–2 $\times$ 10^{15}) h^{-1}
$M_\odot \approx$ (2.8–57) $\times$ 10^{44} kg
T_C = cluster crossing time $\approx$ 2 $R_{cl}/\sigma_v \approx$ 6 $\times$ 10^{16} s $\approx$ 2 $\times$ 10^9 year
L_B = luminosity of galaxy cluster in blue band = (6 $\times$ 10^{11} to 6 $\times$ 10^{12}) h^{-2}
$L_\odot \approx$ (4.6–46) $\times$ 10^{38} J s^{-1}
M_C/L_B = mass to light ratio of galaxy cluster $\approx$ 300 h $M_\odot/L_\odot \approx$ 210 $M_\odot/L_\odot$
L_X = x-ray luminosity of galaxy cluster = ($10^{35.5}$–10^{38}) h^{-2} J s^{-1} $\approx$ 2.0 ($10^{35.5}$–10^{38}) J s^{-1}
n_{cl} = cluster number density $\approx$ (10^{-5}–10^{-6}) h^3 Mpc^{-3} $\approx$ 0.34 (10^{-5}–10^{-6}) Mpc^{-3}

[a] For Hubble constant H_0 = 100 h km s^{-1} Mpc^{-1}. The Sun's mass $M_\odot$ = 1.989 $\times$ 10^{30} kg and the Sun's luminosity $L_\odot$ = 3.828 $\times$ 10^{26} J s^{-1}

which is oriented perpendicular to the Milky Way and extends all the way to the rich Virgo cluster of galaxies.

Bahcall (1977) has reviewed clusters of galaxies, and Bahcall (1988, 1993) has reviewed the large-scale structure in the universe indicated by galaxy clusters. Oort (1983) reviewed superclusters of galaxies, and Rood (1981, 1988) has discussed clusters of galaxies and voids.

14.4.2 Dark Matter in Clusters of Galaxies

Clusters of galaxies are bound together by gravity, even though the expansion of the universe is pulling the galaxies away from one another. We could think that the combined gravitational pull of the numerous galaxies might be sufficient to hold them together but, in 1937, Fritz Zwicky (1898–1974) showed that there must be substantial amounts of unseen material that is keeping the clusters of galaxies from dispersing (Zwicky 1937). In his extraordinarily prescient paper, titled *On the Masses of Nebulae and of Clusters of Nebulae*, he concluded that there must be noticeable quantities of invisible intergalactic matter in clusters of galaxies or else they would be unstable dynamically. In a German language article discussing many of the same topics years earlier, Zwicky introduced the term dunkle materie, or "dark matter" for the invisible stuff, and he concluded that it might be present with a greater density than luminous matter (Van den Bergh 1999).

Zwicky measured the amount of mass required to keep the Coma cluster of galaxies stable, assuming that the motions of its constituent visible galaxies are balanced by the gravitational pull of their combined mass. He found that the total mass of the Coma cluster must be about 10 times the sum of the masses of the individual galaxies it contains. That is, he inferred the total cluster mass, M_C, from the velocity dispersion, σ_V, of the galaxy motions, above that due to the expansion

of the universe. For a cluster of radius, R_C, and angular diameter θ, the binding mass required to hold the cluster together is:

$$M_C = \frac{R_C \sigma_V^2}{G} = \frac{\theta c z \sigma_V^2}{2GH_0}, \tag{14.45}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, the mean redshift of the cluster due to the expansion of the universe is z, and the Hubble constant $H_0 = 100\ h$ km s^{-1} Mpc^{-1} with $h \approx 0.75$. Zwicky concluded that this binding mass M_C is about 20 times the mass of the visible galaxies in the Coma cluster, and that if this dark matter were not present the Coma cluster would be flying apart.

Example: Binding mass of the Coma cluster of galaxies
The Coma cluster of galaxies has a redshift $z = 0.0231$, which corresponds to a distance of $D = cz/H_0 \approx 92$ Mpc, using the speed of light $c = 2.9979 \times 10^5$ km s^{-1} and the Hubble constant $H_0 = 75$ km s^{-1} Mpc^{-1}. The galaxies within this cluster have a velocity dispersion, σ_V, of $\sigma_V \approx 1{,}000$ km s^{-1} $= 10^6$ m s^{-1}. There are about 800 identified galaxies within an area of $100' \times 100'$ centered on this cluster, where $'$ denotes a minute of arc. For an angular diameter $\theta = 100' = 6{,}000'' = 0.029$ radians, where 1 radian $= 2.06265 \times 10^5$ $''$ and $''$ denotes a second of arc, the cluster radius R_C is $R_C = \theta D \approx 2.7$ Mpc, and the binding mass required to hold the cluster together is $M_C = R_C \sigma_V^2 / G \approx 10^{45}$ kg $\approx 5 \times 10^{14}\ M_\odot$ where 1 Mpc $= 3.0857 \times 10^{22}$ m, the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg. If each galaxy contains about 10^{11} visible stars like the Sun, then the total visible stellar mass of the 800 galaxies in this part of the cluster is about $0.8 \times 10^{14}\ M_\odot$. This means that roughly 4.2×10^{14} or about 85 % of the total mass of the cluster has to be in unseen dark matter to bind the Coma cluster of galaxies together.

Hot, x-ray emitting gas has also been found permeating the space between galaxies in massive clusters. Any hydrogen atoms immersed in these clusters and moving at a similar speed to the galaxies would have to be very hot, with a temperature of about 100 million K. At this temperature, the gas is an intense emitter of x-rays.

Example: How hot is the intergalactic gas in a cluster of galaxies?
The thermal velocity, $V_{thermal}$, of a hydrogen atom of mass m_H at temperature T is (Sect. 5.2), $V_{thermal} = (3kT/m_H)^{1/2}$, where the mass of the hydrogen atoms is $m_H = 1.673 \times 10^{-27}$ kg and the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}. If the thermal velocity is comparable to the galaxy velocity

dispersion $\sigma_V = 1{,}000$ km s^{-1} $= 10^6$ m s^{-1} in the Coma cluster of galaxies, then the temperature $T = m_H\sigma_v^2/(3k) \approx 4.04 \times 10^7$ K. Such a hot gas emits intense x-rays. To put it another way, the hot gas with a temperature of about 10^8 K that is observed through the x-ray emission of clusters of galaxies indicates that their intergalactic hydrogen atoms are moving at a velocity comparable to the cluster galaxies.

Although most of the observable mass of the clusters of galaxies is in the form of hot, x-ray emitting gas, outweighing all the visible-light stars within all the galaxies by a factor of about seven, both the stars and the intergalactic x-ray emitting gas constitute only about 15 % of the total mass of the galaxy clusters. The remaining 85 % of the gravitating material is some kind of mysterious dark matter, emitting no detectable radio, visible-light or x-ray radiation, and no one knows what it is.

In his 1937 paper, Zwicky also proposed that the formidable gravity of dark matter in clusters of galaxies would act as a powerful lens, diverting, and focusing the light of more distant galaxies. Arthur Stanley Eddington (1822–1944) had previously noted that the gravitational effect of a star would produce multiple images and magnification of images of a more remote background star located behind it (Eddington 1920). Later, Albert Einstein (1879–1955) discussed the effect for stars, but stated that there was no hope of observing such a stellar lens (Einstein 1936; Renn et al. 1977). However, it was not until the early 1980s that astronomers began to observe that effect for the much more massive clusters of galaxies. They spread the light of a distant galaxy that lies directly behind them into an array of faint, tangentially stretched arcs.

Hubble Space Telescope (*HST*) images of rich clusters of galaxies reveal the highly stretched, distorted, and magnified images of faint galaxies lying far behind them (Figs. 14.12 and 14.13). As Zwicky proposed, the gravitational-lens effect provides information on both the visible and unseen matter. Moreover, the dark matter can act like a zoom lens, magnifying distant galaxies too faint to be seen and bringing them into view.

The presence of two galaxies along the same line of sight, one more distant than the other, was suggested from the spectroscopic database of the Sloan Digital Sky Survey, and these gravitational-lens candidates were confirmed using the high-resolution imaging capability of the *HST*. Hundreds of these cosmic gravitational lenses have been found. Some exhibit partial or complete "Einstein" rings, which indicate near-perfect alignment of the foreground and background galaxies (Figs. 14.14 and 14.15). In 1924, Orest Chwolson (1852–1934) described the production of such a ring by a gravitational lens (Chwolson 1924); the first observed Einstein ring was a radio source. Even a double ring arising from the light of three aligned visible-light galaxies has been found.

Treu (2010) has reviewed strong lensing by galaxies, and Blandford and Narayan (1992) reviewed cosmological applications of gravitational lensing.

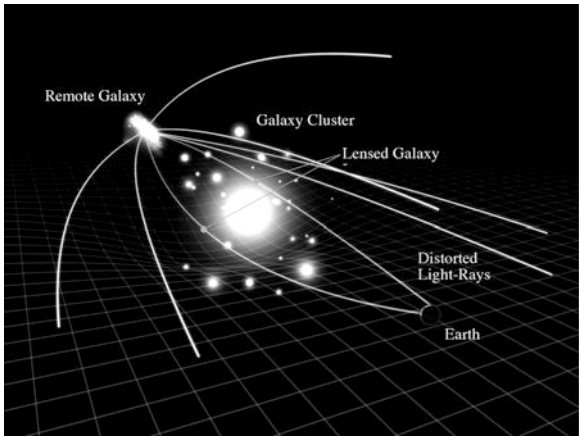


Fig. 14.12 Galaxy cluster lens Very distant and faint galaxies can be investigated by observing them through a cluster of galaxies. The powerful gravitation of the cluster acts like a lens, bending, focusing, and magnifying the light of more distant galaxies that lie behind it (see Fig. 14.14). The gravitational lens action can distort the light from the background galaxies into faint arcs or produce magnified images of individual galaxies that otherwise would remain invisible. (Courtesy of NASA/JPL-Caltech.)

Fig. 14.13 Cluster of galaxies and gravitational lens A *Hubble Space Telescope* image of a rich cluster of galaxies designated Abell 2218. It is about 1,000 Mpc or 3 billion light-years away from the Earth. A typical rich cluster contains hundreds and even thousands of galaxies, each composed of hundreds of billions of stars and possibly up to 10 times more mass within invisible dark matter. The galaxy cluster Abell 2218 is so massive and so compact that its gravity bends and focuses the light from galaxies that lie behind it. Multiple images of these background galaxies are distorted into long faint arcs. Magnified or ring images of individual background galaxies also can be observed. (Courtesy of NASA/STScI/Andrew Fruchter/the ERO team, Sylvia Baggett/STScI, Richard Hook/ST-ECF/Zoltan Levay, STScI.)

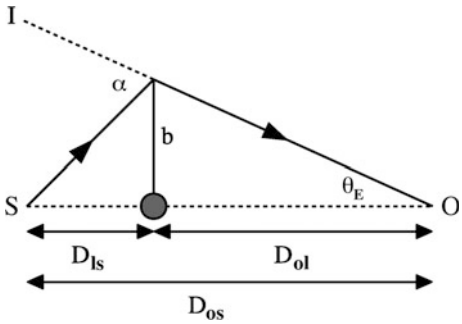


Fig. 14.14 Gravity lens geometry A remote galaxy, labeled S, and an intervening lens galaxy, denoted by L and represented by the *dark filled circle*, lie in a straight line along the observer's line of sight at distances denoted by D and its subscripts. Only rays passing by the lens at the distance b will reach the observer, at O, forming a ring image with an angular radius denoted by θ_E (see Fig. 14.15)

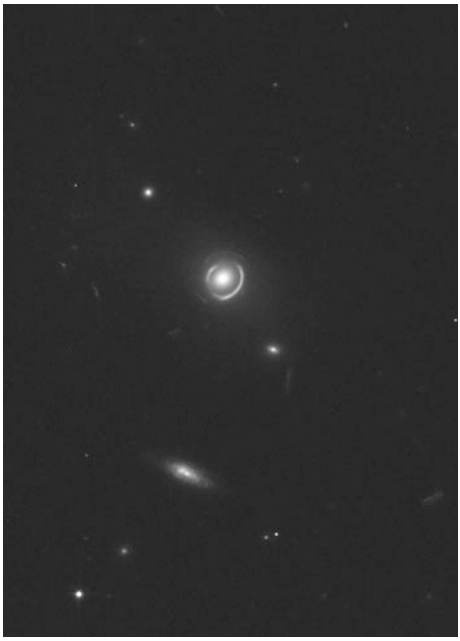


Fig. 14.15 Einstein ring When a background and foreground galaxy are aligned perfectly, the closer galaxy acts as a gravitational lens, bending and magnifying the light of the more distant galaxy and forming a glowing "Einstein" ring. A double ring is captured in this *Hubble Space Telescope* image, indicating an exceptionally rare alignment of a massive foreground galaxy with two background galaxies; the distances of the three galaxies are estimated at 3 billion, 6 billion, and 11 billion light-years from the Earth. (Courtesy of NASA/ESA/Raphael Gavazzi and Tommaso Trea, University of California at Santa Barbara and the SLACS team.)

Example: When a galaxy acts as a gravitational lens
When a light ray from a distant source passes within a distance, b, of an object of mass M, then the bending angle α (see Fig. 14.4), is given by

$$\alpha = \frac{4GM}{c^2 b} \text{ radians} \tag{14.46}$$

where the Newtonian gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

When a light ray passes the limb, or edge, of the Sun's visible disk, the mass is $M = M_\odot = 1.989 \times 10^{30}$ kg, the Sun's mass, and $b = R_\odot = 6.955 \times 10^8$ m, the radius of the Sun. In this case:

$$\alpha = 0.849 \times 10^{-5} \text{ radians } = 1.75'', \tag{14.47}$$

where 1 radian $= 2.06265 \times 10^5\,''$ and the symbol $''$ denotes seconds of arc.

When a lens object, denoted by subscript l and with mass, M, is perfectly aligned along the line of sight to a more distant source, denoted by the subscript s, the light from the distant source will be deformed into a ring (Chwolson 1924) called the Einstein ring. The angular diameter, θ_E, of the ring is given by (Chwolson 1924; Einstein 1936; Lang 1999):

$$\theta_E = \left(\frac{4GM}{c^2 D}\right)^{1/2} = \left(\frac{4GMD_{ls}}{c^2 D_{ol} D_{os}}\right)^{1/2} \text{radians}, \tag{14.48}$$

where the effective distance $D = D_{ol}D_{os}/D_{ls}$, the distance from the observer to the lens mass is D_{ol}, the distance from the observer to the background source is D_{os}, and the distance from the lens to that source is D_{ls} (also see Fig. 14.4).

If the lens mass is a galaxy containing 100 billion stars like the Sun, with a mass $M = 10^{11}\, M_\odot$, that is located halfway between the observer and the source, with $D_{ls} = D_{ol} = D_{os}/2$, then:

$$\theta_E = \left(\frac{2GM}{c^2 D_{ol}}\right)^{1/2} \text{radians} \approx 3.55 \times 10^{12} D_{ol}^{-1/2}\,''. \tag{14.49}$$

For a lens galaxy of redshift $z = 0.5$, the radial velocity $V_r = cz$ and the distance $D_{ol} = cz/H_0 \approx 2.0 \times 10^3$ Mpc $\approx 6.17 \times 10^{25}$ m, where the speed of light $c = 2.9989 \times 10^5$ km s^{-1}, the Hubble constant $H_0 = 100\, h$ km s^{-1} Mpc^{-1}, the $h \approx 0.75$, and 1 Mpc $= 3.0856 \times 10^{22}$ m. For this value of D_{ol} we obtain $\theta_E = 0.45''$, and such rings have been imaged from the *Hubble Space Telescope* (see Fig. 14.15).

14.4.3 Cosmic Streams

The galaxies are not simply flying outward with the expansion of the universe in a smooth and regular manner. Entire groups are streaming together over vast distances in various directions. These so-called peculiar motions are caused by the gravitational pull of huge assemblages of galaxies, and they are not related to the uniform expansion of the galaxies, known as the Hubble flow.

Davis and Peebles (1983) reviewed evidence for local anisotropy in the Hubble flow, and Burstein (1990) reviewed large-scale motions in the universe.

For instance, the nearest large galaxy, Andromeda (M 31), is moving toward the Milky Way at a velocity of about 300 km s^{-1}. A distance of about 778 kpc currently separates the two galaxies, but they are set on an irrevocable collision course. They will meet in a possibly destructive encounter in a few billion years.

Example: When will Andromeda enter the Milky Way?
The distance of the Andromeda nebula from the Earth is $D = 778$ kpc $= 2.40 \times 10^{22}$ m, where 1 kpc $= 3.0857 \times 10^{19}$ m. The observed radial velocity of Andromeda is $V_r = -301$ km s^{-1}, with the negative sign indicating approaching motion. Andromeda will collide with the Milky Way in a time $T = D/V_r \approx 8.0 \times 10^{16}$ s $\approx 2.5 \times 10^9$ years $\approx$ 2.5 billion years, where 1 year $= 3.1557 \times 10^7$ s.

In addition, an entire swarm of galaxies can set off on a trajectory that is independent of the expansion. Because the galaxies are moving together over vast distances of hundreds of millions of light-years, their collective behavior is known as a large-scale streaming motion, which is large in space but not so big in velocity – generally no more than 1,000 km s^{-1}.

All of the galaxies in our region of space, within a volume of 30 Mpc across, are rushing en masse toward the same remote point in space. All of these galaxies are being pulled through space, forced into mass migration by the gravitational pull of "The Great Attractor," located at a distance of about 5 Mpc away. Its mass is equivalent to about 50 million billion (5×10^{16}) stars like the Sun and at least 500,000 galaxies like the Milky Way. The Great Attractor most likely is a rich and massive cluster of galaxies, part of an even larger super-cluster.

Occasionally we observe other galaxies colliding, merging, or passing through each other (Fig. 14.16) when they should be moving farther apart. This mingling also is due to the gravitational attraction of neighboring galaxies, drawing them together and producing local eddies within an outward Hubble flow.

Barnes and Hernquist (1992) have reviewed the dynamics of interacting galaxies.

It is the gravitational interaction of galaxies with one another that distorts the smooth cosmic expansion, producing the peculiar motions superposed on the expanding universe. However, the uniform Hubble flow gathers speed with

Fig. 14.16 Colliding galaxies Gravitational interaction of the Antennae galaxies, catalogued as NGC 4038 and NGC 4039, produces long arms of young stars in their wake. The colliding galaxies are located about 62 million light-years from the Earth and have been merging for the past 800 million years. As the two galaxies continue to churn together, clouds of interstellar gas and dust are shocked and compressed, triggering the birth of new stars. This composite image is from the *Chandra X-ray Observatory* (*blue*), the *Hubble Space Telescope* (*gold and brown*), and the *Spitzer Space Telescope* (*red*). The blue x-rays show huge clouds of hot interstellar gas, the red data show infrared radiation from warm dust clouds that have been heated by newborn stars, and the gold and brown data reveal both star-forming regions and older stars. (Courtesy of NASA/ESA/SAO/CXC/JPL-Caltech/STScI.)

distance. Because the localized streaming motions are limited in velocity, they are relatively slow when compared with the expansion speed of remote galaxies. The very existence of the large-scale streaming motions nevertheless indicates a decidedly uneven and lumpy distribution of galaxies, which eventually was mapped across billions of light-years.

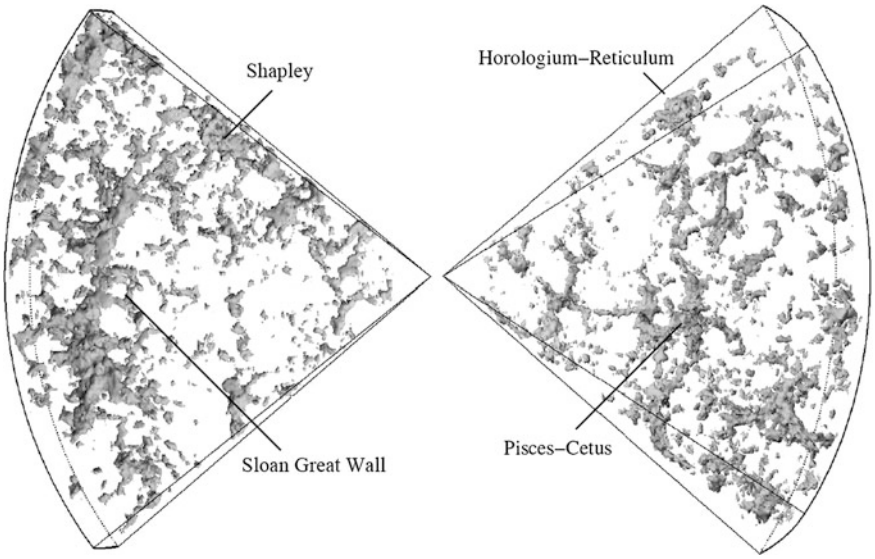


Fig. 14.17 Great walls and voids By measuring the recession velocity, or redshift, of galaxies, astronomers determined their distance and combined it with their location in the sky to obtain the three-dimensional distribution of galaxies. The map shown here is for galaxies within 1 billion light-years (*far left or far right*) from the Earth (*center*). Because galaxies started to form about 12 billion years ago, this is a relatively nearby part of the universe. It includes recession velocities of up to 30,000 km s^{-1}, at a redshift of about $z = 0.1$. The galaxies are concentrated in long, narrow sheet-like walls encircling large empty places known as voids, about 100 million light-years across. The Sloan Great Wall (*left*) spans about 1.4 billion light-years. It may be gravitationally unbound, perhaps beginning to fall apart, but it includes superclusters of galaxies that may stay bound together by their mutual gravitational pull. The Sloan Great Wall was discovered using data from the Sloan Digital Sky Survey in 2004. Other superclusters, or clusters of galaxy clusters, are labeled in the diagram, which is from the Two Degree Field Galaxy Survey. (Courtesy of Willem Schaap, Kapteyn Institute, University of Groningen et al., 2dF Galaxy Redshift Survey.)

14.4.4 Galaxy Walls and Voids

By determining the concentrations of galaxies in different directions and at various redshifts, or depths, astronomers located places where the collective force of gravity pulled galaxies together and locally reversed the uniform expanding motion. The three-dimensional maps reveal fascinating lace-like patterns that connect the galaxies and curving filaments that enclose dark, seemingly vacant places (Lapparent et al. 1986). The galaxies apparently are distributed along the peripheries of gigantic hollow bubbles.

The early redshift surveys also delineated an enormous sheet of galaxies, dubbed "The Great Wall," at distances ranging from 350 million to 500 million light-years away (Geller and Huchra 1989). Giovanelli and Haynes (1991) provided a review of the redshift surveys of galaxies available at that time.

Subsequent three-dimensional maps were obtained using electronic technology that permits the simultaneous measurement of hundreds of galaxy redshifts in a single exposure at a large telescope. The Sloan Digital Sky Survey, for example, revealed the longest sheet of galaxies yet seen (Gott et al. 2005; Einasto et al. 2011). Dubbed the "Sloan Great Wall," it measures 1.37 billion light-years across (Fig. 14.17); and it is the largest observed structure in the universe – at least so far.

Thus, everywhere they look, in whatever direction and near or far, modern telescopes are finding a complex and richly textured universe, filled with luminous concentrations of matter. There is no perceptible end to the lumps and clumps and vacant places, and no one knows where the unevenness will end. Even when looking across 10 % of the observable universe, astronomers continue to find galaxy structures crossing their maps from edge to edge, as well as smaller bubbles, walls, and voids that are nestled together.

All parts of the observable universe are bound within this all-encompassing fabric, glued together by the invisible forces of gravity, suspended in space by motion, and linked by radiation. This all-embracing cosmic web extends throughout the observable universe (Fig. 14.18).

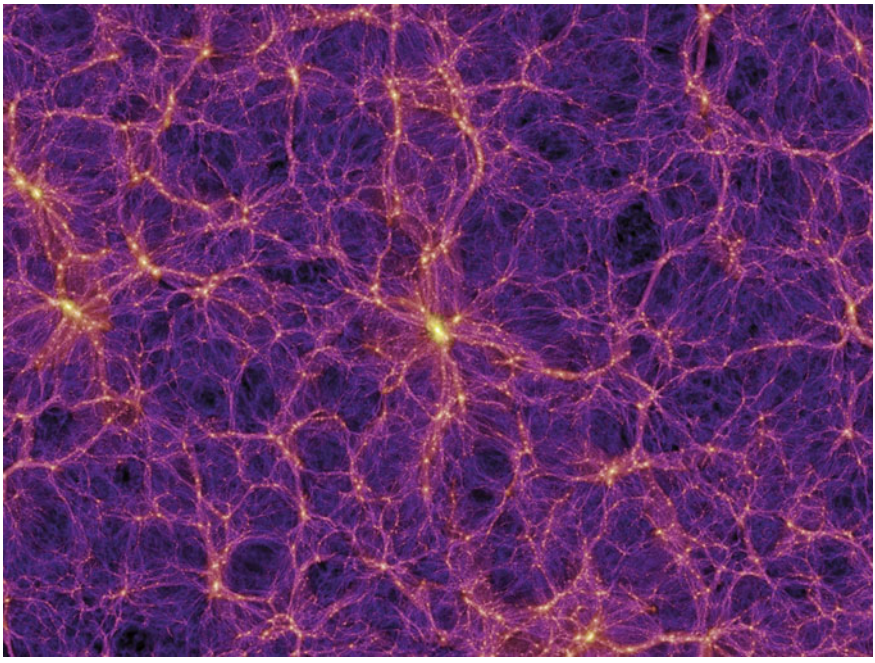


Fig. 14.18 Cosmic web One moment in the ever-changing distribution of galaxies studied using a supercomputer to trace out their formation, evolution, and clustering. The width of this image is about 10 million light-years. (Courtesy of Volker Springel, the Millennium Simulation Project/ Max Planck Institute for Astrophysics, Garching, Germany.)

14.5 Looking Back into Time

Because light travels at a finite speed, to look far into space is to look back into time. When we look farther out into space, we travel back more in time. Large telescopes that detect the faint light of distant objects therefore can be used as time machines to see objects as they were in the past, when their light was emitted, as they were then and not as they are now. In effect, astronomers watch cosmic history race toward us at the speed of light, 299,792 km s^{-1}. The look-back time is simply the amount of time it takes for light to travel from the object to us at that speed.

Moving at the speed of light, it takes 2.3 million years for light to travel from the nearest spiral galaxy, Andromeda, to the Earth. Astronomers have observed radiation from distant galaxies whose light was emitted 13 billion years ago, long before the Sun was formed about 4.6 billion years ago. Therefore, the look-back times for galaxies range from millions to billions of years, spanning an enormous period in which we can watch them evolve.

Some of the most distant galaxies may no longer exist, but they were embryonic galaxies when the light now reaching the Earth began its journey. These galaxies may have perished over time, but their light can survive unchanged, helping us trace out the history of the observable universe from the big bang – about 13.7 billion years ago – to now. As long as a ray of light passes through empty space and encounters no atoms or electrons it will persist forever.

For an object at redshift z, we can specify the look-back time, t_L, at which the radiation was emitted. Looking at objects at larger and larger redshift is looking further and further into the past. The look-back time can be expressed in terms of the Hubble time, t_H, or the reciprocal of the Hubble constant, H_0. For small redshifts $z \ll 1$, we have

$$t_L = z t_H = \frac{z}{H_0} = 9.778 \times 10^9 \frac{z}{h} \text{years} \tag{14.50}$$

where the Hubble constant $H_0 = 100\ h$ km s^{-1} Mpc^{-1}, 1 Mpc $= 3.0857 \times 10^{19}$ km, and 1 year $= 3.1557 \times 10^7$ s. The Hubble time is

$$t_H = \frac{1}{H_0} \approx 3.0857 \times 10^{17} \frac{1}{h} \text{s} \approx 4.1 \times 10^{17} \text{s} \approx 13 \times 10^9 \text{years}. \tag{14.51}$$

At large redshift, the look-back time is given by:

$$t_L = \frac{2}{3 H_0 \Omega_0^{1/2} (1+z)^{3/2}}, \tag{14.52}$$

where the density parameter $\Omega_0 = \rho_0 / \rho_C$, the ratio of the present mass density of the universe, ρ_0, to the critical mass density, $\rho_C = 3H_0^2/(8\pi G) = 1.879 \times 10^{-26}\ h^2$ kg m^{-3}, for a Hubble constant $H_0 = 100\ h$ km s^{-1} Mpc^{-1}, needed to close the universe in the future.

Example: Viewing a distant galaxy
Suppose a galaxy is located at a distance of $D = 100$ Mpc. Assuming a Hubble constant of $H_0 = 75$ km s^{-1} Mpc^{-1}, the radial velocity of the galaxy is $V_r = H_0 \times D = 7{,}500$ km s^{-1} and its redshift is $z = V_r/c = 0.025$. The Hubble time $t_H = 1/H_0 = 4.11 \times 10^{17}$ s $\approx$ 13 billion years, where we use 1 Mpc $= 3.0857 \times 10^{19}$ km and 1 year $= 3.1557 \times 10^7$ s. The look back time $t_L = zt_H = 0.325$ billion years.

If the redshift of a galaxy is $z = 3.0$, then the radial velocity is obtained from $\frac{V_r}{c} = \frac{(z+1)^2-1}{(z+1)^2+1}$, or $V_r = 0.88\,c \approx 2.64 \times 10^5$ km s^{-1}, where the speed of light $c = 2.9979 \times 10^5$ km s^{-1}. Assuming that the density parameter $\Omega_0 \approx 10^{-2}$, then the look-back time for this galaxy is $t_L = 0.83\ t_H = 10.8$ billion years, and about 3 billion years after the big bang that occurred about 13.7 billion years ago.

The Hubble time is the approximate time when the expansion of the universe began. Gravity can slow the expansion to a lesser age, while dark energy can accelerate it, giving a greater age. But the Hubble time provides a pretty good estimate of the age of the expanding universe. Its beginning was not always known with such precision. For a while, it looked as if the Earth was older than the universe itself, leading to a new steady state theory for the universe (Focus 14.3). Kragh (1997) discusses the historical development of the steady state and expanding universe theories.

Focus 14.4 How old is the observable universe?
If the observable universe had a beginning, it must have a finite age, and estimates of that age have lengthened as our astronomical knowledge improved. When the known universe was confined largely to the Earth, an age of 10 million to 100 million years was inferred from the time it would take for the planet to cool from an initially molten state (Kelvin 1862, 1899; Burchfield 1975). The discovery of radioactivity then provided a new source of energy to keep the Earth hot inside (Rutherford 1905), and radioactive elements were used to determine an age of the oldest rocks on the Earth, first at 2 to 3 billion years (Rayleigh 1905; Boltwood 1907; Rutherford 1929; Holmes 1949) and eventually at 4.6 billion years old (Patterson 1956).

When it was realized that subatomic nuclear energy keeps the stars shining, an evaluation of the their nuclear-reaction rates indicated stellar lifetimes of up to 10 billion years (Schönberg and Chandrasekhar 1942). Ages between 10 billion and 20 billion years were inferred from studies of the chemical evolution of our Galaxy (Fowler and Hoyle 1960; Tinsley 1975) as well as the thermonuclear evolution of stars in globular star clusters (Sandage 1970; Hesser et al. 1987; Chaboyer 1995).

Another way of estimating the age of the known universe is to watch the expanding galaxies. They are moving away from a beginning, a cosmic horizon, which is as far away as we ever can see. The distance between there and now, as well as the age of the observable universe, is estimated by calculating how long it has taken for the galaxies to move this far at their observed speeds.

If the expanding universe has been moving along at a steady pace, then its age can be determined from the present rate of expansion. That rate is quantified by the current value of the Hubble constant, the reciprocal of which provides an expansion age. The currently accepted value of 75 km s^{-1} Mpc^{-1} corresponds to an expansion age of 13.0 billion years. A lower value of the Hubble constant would imply that the universe is expanding now at a slower rate and has a greater age. A higher value of the constant implies a more rapid expansion and that less time has elapsed since the expansion started.

The universe may not have been moving apart with an unchanging speed. Its expansion would be slowed by the mutual gravitational pull of all of the dark matter in the universe or accelerated by anti-gravitational forces. Moreover, today's value of the Hubble constant has always been imprecise due to the streaming motions of nearby galaxies, which are not part of their expansion, and by uncertain distances for remote galaxies.

Hubble's measurements of the constant that now bears his name, pegged at 530 in the typical units, caused quite a problem. It corresponds to an expansion age of only 1.8 billion years. At about the same time that Hubble was making his observations, Ernest Rutherford (1871–1937) and his colleagues at the Cavendish Laboratory in England used radioactive dating to estimate an age of at least 3.4 billion years for the oldest terrestrial rocks (Rutherford 1929). How could the Earth be older than the expanding universe?

Arthur Stanley Eddington (1882–1944) was not at all troubled by the discrepancy. He simply extended the beginning of the universe farther back in time than the start of its expansion (Eddington 1933). In this interpretation, the world once existed in an unmoving state described by the cosmological constant that Albert Einstein (1879–1955) previously introduced to stop the eventual collapse of a static, nonexpanding universe (Einstein 1917a, b). An initial nonmoving universe of an indeterminate age could be sent into unrestrained expansion whenever we choose by the slightest disturbance, which upsets the balance between gravitational attraction and cosmological repulsion. For Eddington, a disrupting disturbance about 2 billion years ago caused a slight expansion that thinned out the universe, making it less able to resist the cosmic repulsion, and the runaway expansion of the galaxies began.

Three young scientists at Cambridge University – Hermann Bondi (1919–2005), Thomas Gold (1920–2004), and Fred Hoyle (1915–2001) – noticed that we can adjust the cosmological constant to accommodate almost

any related observation and the theory thereby lost its simplicity and uniqueness. So the trio proposed a universe that had no beginning (Bondi and Gold 1948; Hoyle 1948). The cosmos, they proclaimed, may have always existed, presenting an unchanging steady state on the largest scales of space and time. We would no longer have to attribute the observable universe to a past creation that was inaccessible to scientific scrutiny or understanding.

They acknowledged the inescapable fact that the galaxies are moving apart but supposed that they have always been doing so, while new matter is being created continuously at an average rate of just one hydrogen atom per cubic meter of space per year. That is just sufficient to counteract the dispersal and thinning out of the expanding universe and keep the overall universe unchanged with time.

The age problem that led to serious consideration of the Steady State Theory was partially resolved by Walter Baade (1893–1960) during World War II (1939–1945). As luck would have it, Eastman Kodak had just developed a red-sensitive emulsion for wartime reconnaissance. Pushing the 2.5 m (100 inch) telescope to its very limits, Baade used the red-sensitive plates to resolve the nucleus of the nearby Andromeda galaxy, distinguishing individual red giant stars in the crowded center. His measurements indicated that these red-colored stars were similar to those found in the globular clusters of our own Galaxy, but different from the highly luminous blue-colored stars found in the outer arms of both the Milky Way and Andromeda.

When the 200 inch telescope on nearby Palomar Mountain began operation, in 1948, Baade continued his investigations of the red and blue kinds of stars, and concluded that they obey different period-luminosity relationships. This meant that Hubble had been confused when applying the distance calibration of Cepheid variable stars in globular clusters to the other types of variables in the arms of nearby spiral nebulae. So Baade made the necessary corrections, obtaining a distance of about 2 million light-years for Andromeda, which reduced the Hubble constant by about half and enlarged both the scale and age of the expanding universe by a factor of about two (Baade 1952).

Although a consummate observer, Baade tended to avoid detailed analysis and written accounts of his discoveries. So it was fortunate that Henrietta H. Swope (1902–1980), the daughter of the wealthy president of the General Electric Company, joined Baade to assist with the analysis of his excellent photographs. They used the results to propose a downward revision of the Hubble constant to a value of 100, in the usual units (Baade and Swope 1955). Allan Sandage (1926–2010) then continued to correct for Hubble's mistaken identification of the brightest stars in Andromeda; in 1958, he announced that the elusive constant had a value of 75 (Sandage 1958). Using Sandage's measurements, the expansion age is about 13 billion

years. Therefore, the universe was considerably older than the oldest terrestrial rocks, and scientists were reassured that the observable universe had a definite beginning.

Estimates of the age and size of the observable universe haven't changed all that much over the past decades, but the uncertainties have become smaller. The *Hubble Space Telescope* and the *Spitzer Space Telescope* have been used to refine estimates of Hubble's constant by observing Cepheid variable stars, obtaining values of 72 ± 8 (Freedman et al. 2001), $H_0 = 73.8 \pm 2.4$ km s^{-1} Mpc^{-1} (Reiss et al. 2011) and $H_0 = 74.3 \pm 2.1$ km s^{-1} Mpc^{-1} (Freedman et al. 2012).

Recent estimates indicate that everything we know about, the whole observable universe, is not older than 13.7 billion years. All that we can observe has a history that can be traced back to that beginning, but most of what we see is significantly younger. Perhaps the most interesting consequences are that the observable universe is about three times older than the Earth, and that the light we see from the most distant galaxies was emitted before our solar system existed.

The exact mathematical equations for the look-back time can be complicated, and the uncertainties increase at larger redshifts. Nevertheless, the uncertainties in the distances and look-back times are no larger than those caused by our imprecise knowledge of the Hubble constant. Even at a redshift $z = 5.0$, amongst the largest ones observed, the look-back time ranges from 10 to 15 billion years depending on the choice of the Hubble constant and the mass density of the universe. So we can ignore the effects of space curvature in most practical computations.

Throughout most of the past decades, it has been assumed that it is the mass of the universe that curves its shape, establishes its geometry and determines its fate. Under this assumption, which ignores the more recent discovery of dark energy, the mass density of galaxies, ρ_G, determines the ultimate destiny of the universe. If this mass density exceeds a certain critical value, ρ_c, then gravity will eventually overcome expansion. Imagine the most distant galaxy with mass, m_G, distance D_G, and velocity $V_G = H_0 D_G$. Gravity will just balance the expansion of this galaxy if its kinetic energy of expansion is equal to the gravitational potential energy of all of the rest of the universe, or if:

$$\frac{m_G V_G^2}{2} = \frac{m_G H_0^2 D_G^2}{2} = \frac{G m_G M_U}{D_G}, \tag{14.53}$$

where M_U is the total mass of all the rest of the universe inside distance D_G and the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}. In other words, the velocity, V_G, of the most distant galaxy is just equal to the escape velocity, V_{esc}, of the entire universe, $V_{esc} = (2GM_U/D_G)^{1/2}$.

Collecting terms we obtain a critical mass density of:

$$\rho_c = \frac{3M_U}{4\pi D_G^3} = \frac{3H_0^2}{8\pi G} = 1.879 \times 10^{-26} h^2 \,\mathrm{kg\,m^{-3}} \approx 1.0 \times 10^{-26} \,\mathrm{kg\,m^{-3}} \quad (14.54)$$

where the Hubble constant $H_0 = 100\ h$ km s^{-1} Mpc^{-1}, one parsec = 1 pc = 3.0857×10^{16} m, 1 Mpc = 10^6 pc, and the Newtonian gravitational constant $G = 6.673 \times 10^{-11}$ N m^2 kg^{-2}. For $h = 0.75$, we have $\rho_c \approx 10^{-26}$ kg m^{-3}.

14.6 Using Einstein's *General Theory of Relativity* to Explain the Expansion

Edwin Hubble (1889–1953) relied solely on the power of observation, preferring to avoid what he called "the dreamy realms of speculation" (Hubble 1936), most likely referring to theoretical physicists. Although an expanding universe was a possible consequence of Albert Einstein's (1879–1955) *General Theory of Relativity* (Einstein 1917a, b), Hubble thought that such models were a forced interpretation of the observational results. Even as late as 1953, shortly before his death, Hubble insisted that his law should be formulated as an empirical relation between observed data (Hubble 1953). The Belgian astrophysicist and Catholic priest, Georges Lemaître (1894–1966) had nevertheless already interpreted the radial velocities of spiral nebulae in terms of Einstein's theory in 1927.

After being ordained a priest in 1923, Abbé Lemaître spent a year at Cambridge University as a student of Arthur Stanley Eddington (1882–1944), reviewing Einstein's *General Theory of Relativity*, and during the next 2 years, Lemaître studied at the Massachusetts Institute of Technology and worked with Harlow Shapley at the nearby Harvard College Observatory.

The Belgian cleric also toured the country, meeting Vesto Slipher (1875–1969) at the Lowell Observatory in Flagstaff, Arizona, and Edwin Hubble at the Mount Wilson Observatory, California. As a result, he learned all about the latest measurements of the redshifts, or radial velocities, of spiral nebulae, interpreting them as a cosmic effect of the expansion of the universe, all in accordance with the relativity theory, and additionally derived a theoretical expression for the linear increase of their velocities with distance (Lemaître 1927).

But we now see this in hindsight, and in 1927 most scientists were not even aware of the observational support for an expanding universe. Lemaître also published his interpretation in a fairly obscure journal, the *Annales de la Societé Scientifique de Bruxelles*. So practically no one was aware of his findings, which either went unnoticed or were ignored. At this time, no other astronomer or physicist had used the *General Theory of Relativity* to explain the observed universe, and it wasn't until a few years later that Eddington sponsored an English translation of Lemaître's paper in the *Monthly Notices of the Royal Astronomical Society* (Lemaître 1931, b, c). In English, the title read *A Homogeneous Universe of Constant Mass and Increasing Radius Accounting for the Radial Velocity of*

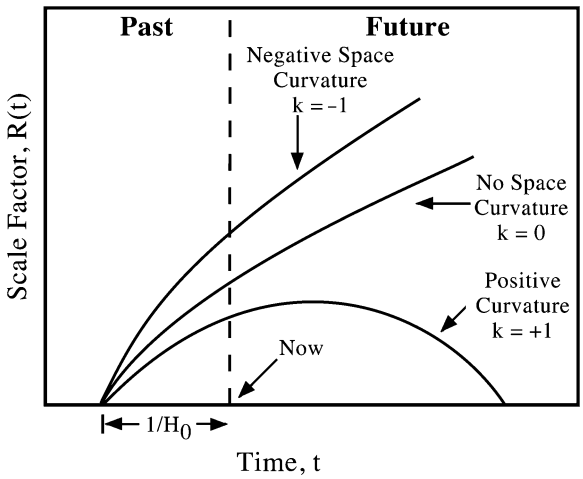


Fig. 14.19 Size of the expanding universe Schematic representation showing the size, or scale factor $R(t)$, of the expanding universe as a function of time, t. The approximate age since the expansion began is given by $1/H_0$ where H_0 is the Hubble constant. Three models describe future possibilities for the universe with no dark energy or cosmological constant. It can become closed with positive space curvature, denoted by space curvature constant $k = +1$, forever open with no space curvature and described by Euclidean space with $k = 0$, or always open with negative space curvature constant $k = -1$

Extra-Galactic Nebulae. Since he was the first to relate the observed galaxy motions to the theory, Lemaître's solution was widely heralded for its novelty and it led to greater appreciation of the significance of Hubble's discovery.

We cannot tell exactly how far away the distant galaxies are, or precisely how long ago their radiation began its journey, until we understand the curvature of space. If space is curved, the path through space might be noticeably longer than that expected in flat, un-curved space, with larger distances and greater look-back times. The detailed expressions for the distance and look-back time therefore depend upon the amount of space curvature, or the mass density of the universe, as well as the redshift.

Schneider (2006) provides a nice textbook of extragalactic astronomy and cosmology, whereas Sandage (1988) and Lang (1999) have reviewed observational tests of world models.

The line element or metric, ds, for a homogenous, isotropic expanding universe is the Robertson-Walker metric (Robertson 1935, 1936; Walker 1936):

$$ds^2 = c^2dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{14.55}$$

where the space curvature constant $k = -1$, 0 and +1, and $R(t)$ is called the radius of curvature or the scale factor of the universe (Fig. 14.19). The r coordinate has zero value for some arbitrary fundamental observer, the surface $r =$ constant has

the geometry of the surface of the sphere, θ, ϕ are polar coordinates, t is the time coordinate and c is the speed of light.

If light was emitted from an extragalactic object at time t_e, then its redshift z is given by:

$$1 + z = \frac{R(t_0)}{R(t_e)}, \tag{14.56}$$

where the subscript zero is used to denote the present epoch, and $R(t_0)$ is the present value of $R(t)$ at time t_0. As previously mentioned, the observed wavelength, λ_0, of a spectral line emitted by an extragalactic object is longer than the emitted wavelength, λ_e, with a redshift z defined by:

$$z = \frac{\lambda_0 - \lambda_e}{\lambda_e}. \tag{14.57}$$

At any cosmic time t, we can define a Hubble expansion parameter, $H(t)$, by:

$$H(t) = \frac{\dot{R}(t)}{R(t)} = \frac{dR(t)/dt}{R(t)}, \tag{14.58}$$

where the $\cdot$ denotes differentiation with respect to time, a deceleration parameter, $q(t)$:

$$q(t) = -\frac{\ddot{R}(t)R(t)}{\dot{R}^2(t)} = -\frac{1}{H_0}\frac{\ddot{R}(t)}{\dot{R}(t)}, \tag{14.59}$$

and a density parameter, $\Omega(t)$, by:

$$\Omega(t) = \frac{8\pi G\rho(t)}{3H^2(t)} \tag{14.60}$$

for mass density $\rho(t)$.

For a homogeneous, isotropic expanding universe of mass-energy density ρ and zero cosmological constant $\Lambda = 0$ (the cosmological constant is discussed in Sect. 15.6.2), we obtain:

$$H^2(t) = \left(\frac{\dot{R}(t)}{R(t)}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{R^2(t)} \quad \text{for } \Lambda = 0, \tag{14.61}$$

where G is the gravitational constant, and:

$$\frac{\ddot{R}(t)}{R(t)} = \frac{-4\pi G}{3}\left(\rho + \frac{3P}{c^2}\right) \quad \text{for } \Lambda = 0, \tag{14.62}$$

where P is the pressure.

These two equations are sometimes called the Friedmann equations, since the Russian mathematician Aleksandr Friedmann (1922–1924) first derived them, but

he had no idea that they might apply to the observable universe (Friedmann 1922, 1924). They can be rearranged to imply a conservation of energy relation:

$$\dot{\rho}(t)c^2 = -3\frac{\ddot{R}(t)}{R(t)}\left(\rho c^2 + P\right). \tag{14.63}$$

At the present time, denoted by t_0, the expansion parameter is the Hubble constant $H(t_0) = H_0 = 100\ h$ km s^{-1} Mpc $^{-1}$, with a value of $h \approx 0.75$. Moreover, in the present matter-dominated era the pressure $P_0 = 0$, the radiation energy density $\rho_r(t_0)$ can be omitted when compared to the mass density $\rho_m(t_0) = \rho_0$, and the deceleration parameter is given by:

$$q_0 = q(t_0) = \frac{4\pi G}{3H_0^2}\rho_0 = \frac{\rho_0}{2\rho_C} \text{ for } \Lambda = 0, \tag{14.64}$$

the density parameter Ω_0 is given by:

$$\Omega_0 = \Omega(t_0) = \frac{\rho_0}{\rho_C} \text{ for } \Lambda = 0, \tag{14.65}$$

and ρ_C, the critical mass density needed to close the universe is given by:

$$\rho_C = \frac{3H_0^2}{8\pi G} = 1.879 \times 10^{-26}\ h^2 \text{ kg m}^{-3} \approx 1.0 \times 10^{-26} \text{ kg m}^{-3}, \tag{14.66}$$

where the Hubble constant $H_0 = 100\ h$ km s^{-1} Mpc^{-1} $= 3.24 \times 10^{-18}\ h$ s^{-1}, one parsec $= 1$ pc $= 3.0857 \times 10^{16}$ m, 1 Mpc $= 10^6$ pc, and the Newtonian gravitational constant $G = 6.693 \times 10^{-11}$ N m^2 kg^{-2}.

We also have the relation:

$$\frac{kc^2}{R^2(t_0)} = H_0^2(2q_0 - 1) \text{ for } \Lambda = 0. \tag{14.67}$$

For model universes with zero cosmological constant, or $\Lambda = 0$, we have three possibilities with three different values of the space curvature constant k, describing a closed, Euclidean or open universe (Fig. 14.19).

If $k = 1$ then $q_0 > 0.5$, $\rho_0 > \rho_C$ and $\Omega_0 > 1$ for elliptical closed space and an oscillating universe in which $R(t)$ reaches a maximum in the future. The universe would then eventually turn back upon itself and reform the dense fireball of its youth.

If $k = 0$ then $q_0 = 0.5$, $\rho_0 = \rho_C$ and $\Omega_0 = 1$ for a flat, Euclidean space without curvature and an ever-expanding Einstein-De Sitter universe (Einstein and De Sitter 1932), in which $R(t)$ continues to forever increase with time, t. The universe is then poised between open and closed, right on the dividing line.

If $k = -1$ then $0.0 < q_0 < 0.5$, $\rho_0 < \rho_C$, and $\Omega_0 < 1$ for a hyperbolic open space and an ever-expanding Milne universe (Milne 1935) in which $R(t)$ also continues to forever increase with time, t. The inward pull of gravity is too weak to ever quell the outward expansion of the universe.

The cosmological constant has been revived with the discovery of mysterious dark energy that is now accelerating the expansion of the universe, and the appropriate adjustments to our equations are given in Sect. 15.6.2.

Chapter 15
Origin, Evolution, and Destiny of the Observable Universe

15.1 Hotter Than Anything Else

Regardless of the direction in which we look out into space, almost all of the distant galaxies are flying apart, dispersing and moving away at speeds that increase with their distance, as if they had been ejected by a cosmic bomb. Astronomers call it the "big bang."

We can envision this early state by putting the observed expansion of the galaxies in reverse and pushing the galaxies back closer together until a time about 13.7 billion years ago, when the universe was incredibly small and all of its mass was compressed to a very high density. It marks the beginning of the observable universe, but no one knows what happened before the big bang propelled the expanding universe into existence (Focus 15.1).

Focus 15.1 Before the Big Bang

How did it all begin? The existing theory falls apart at the first crucial instant, at the beginning of the big bang, and can't be extended to anything that occurred before that. Moreover, there is no observational evidence for prior events.

Einstein's *General Theory of Relativity* cannot be used back then because the equations contain a singularity at the beginning of the big bang, when a non-zero parameter is divided by zero and conditions cannot be defined. So we've just pushed the mystery of the ultimate origin of the universe back about 13.7 billion years, to a point that science cannot penetrate.

An inflation theory does describe what could have happened in the first fraction of a second of the big bang, when the universe was just 10^{-35} s old. Guth (1981) and Linde (1982) describe inflation; Narlikar and Padmanabhan (1991) have described inflation for astronomers.

During inflation the universe was driven by a repulsive gravity, unlike the attracting kind we are used to, and operated on a very small scale in both

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_15, © Springer-Verlag Berlin Heidelberg 2013

space and time, blowing the universe up, enlarging it by an enormous factor. Owing to its inherent instability, the burst of inflation soon decayed away and came to an end, in a time far less than one second, releasing its remaining energy into material particles and creating the heat of the big bang, the primeval fireball.

This accelerated expansion in the first miniscule moments of the big bang, this inflation, supposedly obliterated evidence of previous space, time, energy and matter, erasing previous history. That cosmic forgetfulness closes the door to the very beginning, conveniently avoiding the question of ultimate origins, the original genesis, and removing it from any observational consequences.

In other words, according to this theory the big bang or its immediate consequences destroyed all evidence of what came before. Or the big bang might have initiated time, on a day without a yesterday. So there is no before. Or perhaps the explanation lies outside space and time.

In any event, the existing equations and theories fail to explain how the observed universe began. So we still don't know how the universe came into being, and it remains a captivating mystery.

Because gases become hotter when they are compressed and cool when they expand, the observed universe must have been incredibly hot in its earliest, most compact state. As we look back in time, at the most distant regions, the universe becomes increasingly hot, eventually becoming so exceptionally hot that radiation was the most powerful force, dominating the expansion of the universe.

In the earliest moments of the big bang, there were no stars or galaxies, only intense radiation and subatomic particles from which the material universe subsequently grew. During this hot beginning, matter was then being created by radiation and vice versa. Some of the incredibly energetic radiation was being transformed into electrons and their anti-matter counterparts, the positrons or positive electrons, and just as often an electron would collide with a positron to make radiation again.

If we let γ denote a photon of the energetic radiation, or a gamma ray, then the equilibrium condition can be written in short hand notation as:

$$\gamma + \gamma \rightleftarrows e^- + e^+, \tag{15.1}$$

where e^- denotes an electron and e^+ denotes a positron, the anti-matter particle of the electron. The double arrow means that the reaction goes in both directions at the same rate. The forward process, from left to right, is known as *electron–positron pair creation*, and the reverse one is called *pair annihilation*.

Neutrons and protons were also around, and these subatomic particles would also turn back and forth into each other, through reactions that included electron neutrinos and electron antineutrinos, as well as electrons and positrons. The equilibrium between radiation and these subatomic particles continued as long as it

was hot enough to create positrons; however, the radiation quickly cooled as the result of the expansion of the universe into a greater volume. Electron–positron pair creation continued only as long as the thermal energy of the radiation, $kT_r(t)$, with radiation temperature $T_r(t)$ at time t, was greater than the rest mass energy of the electron, or for $kT_r(t) \geq m_e c^2$. Here the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the mass of the electron is $m_e = 9.1094 \times 10^{-31}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. Using these constants, the equilibrium stopped when the radiation cooled to a temperature of $T_r(t) = 5.93 \times 10^9$ K. After that, no more positrons were made and the leftover positrons were then consumed by interactions with electrons. The neutrons, protons, and electrons that remained eventually gathered together to create the material universe we have today.

In the early stages, the radiation controlled the expansion, with an effective energy-mass density, $\rho_r(t)$, at time, t, given by:

$$\rho_r(t) = \frac{aT_r^4(t)}{c^2} \approx 0.842 \times 10^{-32}\, T_r^4(t) \text{ kg m}^{-3}, \tag{15.2}$$

where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}, and we have divided by the square of the speed of light, $c = 2.9979 \times 10^8$ m s^{-1}, to convert the energy density into an equivalent mass density.

If we denote the effective mass of the radiation as $M_r(t)$, assume the universe is expanding at a constant velocity $V = R(t)/t$ for radius $R(t)$ at time t, and equate the kinetic energy of expansion to the gravitational potential energy of the radiation, then

$$\frac{1}{2} M_r(t) \left[\frac{R(t)}{t}\right]^2 = \frac{GM_r^2(t)}{R(t)}, \tag{15.3}$$

or

$$M_r(t) = \frac{R^3(t)}{2Gt^2}, \tag{15.4}$$

where the gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and

$$M_r(t) = \frac{4\pi}{3} R^3(t) \rho_r(t) = \frac{4\pi}{3} R^3(t) \frac{aT_r^4(t)}{c^2}. \tag{15.5}$$

Collecting terms, we obtain an approximate equation for the radiation temperature $T_r(t)$ at time, t, as long as the radiation is controlling the expansion.

$$T_r(t) = \left[\frac{3c^2 M_r(t)}{4\pi a R^3(t)}\right]^{1/4} = \left[\frac{3c^2}{8\pi G a}\right]^{1/4} \frac{1}{t^{1/2}} \approx 2.15 \times 10^{10} \frac{1}{t^{1/2}} \text{ K}, \tag{15.6}$$

where the time t is in seconds for the numerical approximation, or equivalently

$$t \approx 4.6 \times 10^{20} \frac{1}{[T_r(t)]^2} \text{ s}. \tag{15.7}$$

The radiation temperature has, for example, dropped to 5.93×10^9 K in just about 13 s after the big bang.

In the early stages, the radiation controlled the expansion of the universe because it was incredibly hot. Just 1 s after the big bang, the radiation had a temperature of about 20 billion, or 2×10^{10} K. As the universe continued to expand, the radiation steadily cooled, and the matter eventually took over the expansion. However, the big bang was so intense and so hot that we are still immersed within the radiation.

15.2 Three Degrees Above Absolute Zero

15.2.1 An Unexpected Source of Noise

The discovery of the faint afterglow of the big bang was a serendipitous event, involving a horn-reflector antenna that had been used at the Bell Telephone Laboratories in the first tests of a communication satellite. Arno Penzias (1933–) and Robert Wilson (1936–) were measuring the temperatures of noises in the horn-antenna system so they could make accurate measurements of the intensity of several extragalactic radio sources. A persistent, ubiquitous, and unvarying noise source was detected at a signal frequency of 4,080 MHz $= 4.08 \times 10^9$ Hz, or a wavelength of 7.35 cm, contributing an antenna temperature of only 3 degrees above zero, or about 3 K. It was equally strong in all directions, wherever the antenna was pointed, independent of the time of day and year and with no dependence on the location of any known cosmic radio source.

Penzias and Wilson did not know what they had found and avoided any mention of the cosmological implications in their publication (Penzias and Wilson 1965), which had the modest title, *A Measurement of Excess Antenna Temperature at* 4,090 *MHz*. However, a group at Princeton University, which was attempting to make a similar measurement at the time, drew attention to the implications in a companion paper (Dicke et al. 1965). The unexpected source of noise was the faint, cooled relic of the hot big bang, now known as the *three-degree cosmic microwave background radiation*, because it has a temperature of about 3 K and it originated before the stars and galaxies were formed, lying behind them.

This particular discovery was not entirely unanticipated. In the late 1940s and early 1950s, George Gamow (1904–1968), Ralph A. Alpher (1921–2007), James W. Follin (1919–2007), and Robert C. Herman (1914–1997) had speculated that the 1-billion-degree, or 10^9 K, radiation of the early universe would have cooled to about 5 K during the past billions of years of expansion, but nobody had attempted to observe the relic radiation (Gamow 1948, 1956; Alpher and Herman 1948). Penzias and Wilson were also unaware of this previous calculation until after their discovery. They received the 1976 Nobel Prize in Physics for their discovery of the cosmic microwave background radiation.

15.2.2 Blackbody Spectrum

At the high temperatures during the early history of the expanding universe, the radiation and subatomic particles frequently interacted, achieving thermal equilibrium characterized by single temperature. Later, when the universe thinned out and cooled by expanding into a greater volume, the matter and radiation quit interacting, going their separate ways. However, the radiation would have retained its thermal nature as it cooled and the temperature slowly decreased.

A perfect thermal radiator is known as a *blackbody*, which absorbs all thermal radiation falling on it and reflects none – hence, the term *black*. The distribution of the radiation emitted by the blackbody, its spectrum, peaks at a wavelength that is inversely proportional to the temperature, dropping precipitously at shorter wavelengths and falling off gradually at longer ones (Sect. 2.4).

The expansion of the universe preserves the blackbody spectrum of the radiation for all time. No process can destroy its shape, but the location of maximum intensity will stretch to increasingly longer wavelengths as time goes on and the radiation gets colder. The wavelength of peak intensity, λ_{max}, is inversely proportional to the radiation temperature, T, and the Wien displacement law (Sect. 2.4) specifies that wavelength as $\lambda_{max} = 0.0029/T$ m. In the present epoch, with a temperature of only about 3° above absolute zero, or at 3 K, the blackbody radiation intensity peaks at a wavelength of about 0.001 m or 0.1 cm. Unfortunately, the Earth's atmosphere absorbs cosmic radiation at this short wavelength where the most intense radiation occurs.

The definitive spectral measurements therefore had to be made from above the atmosphere using NASA's *COsmic Background Explorer* (*COBE*), launched on 18 November 1989. Less than two months after *COBE* went into orbit, but a quarter-century after the discovery of the cosmic radiation, John C. Mather (1946–) reported the combined results of millions of *COBE* spectral measurements at an American Astronomical Society meeting near Washington, DC. The spectrum fit the Planck blackbody curve with a precision of 1 part in 10,000 (Fig. 15.1), establishing a temperature of precisely 2.725 K, with an uncertainty of 0.002 K (Mather et al. 1994).

Such a thermal spectrum could not have happened in the universe as it is now. Matter currently has a very different temperature than the background radiation. In other words, the observed spectrum is proof that the observable universe must have expanded from a very hot, dense state in the past, when matter and radiation were in thermal equilibrium and at the same temperature.

Every part of space now is filled with background radiation. From its temperature, of $T = 2.725$ K and we can specify the frequency $\nu_{\max}$ from the Wien displacement law in frequency (Sect. 2.4) as $\nu_{\max} \approx 2.8\,kT/h \approx 5.88 \times 10^{10}\ T \approx 1.60 \times 10^{11}$ Hz, where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and the Planck constant

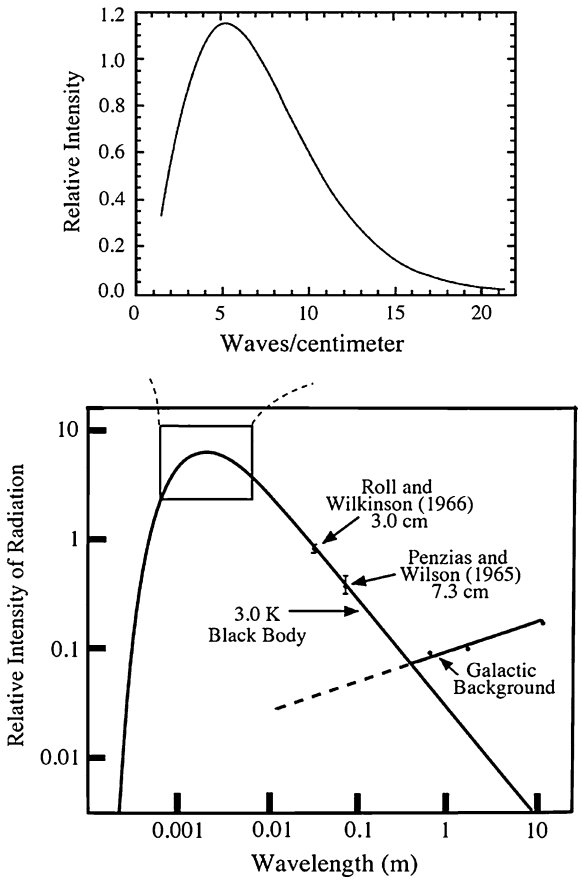


Fig. 15.1 Spectrum of the cosmic microwave background radiation The intensity of the cosmic microwave background radiation plotted as a function of wavelength. This thermal radiation was formed about 390,000 years after the big bang, which occurred about 14 billion years ago. The observed radiation has a nearly perfect blackbody spectrum. Pioneering measurements by Arno A. Penzias (1933–) and Robert W. Wilson (1936–) in 1965 and Peter G. Roll (1935–) and David T. Wilkinson (1935–2002) in 1966, at 7.35 and 3.0 cm wavelength, respectively, are compared to the expected spectrum of a three-degree blackbody and radiation from our Galaxy (*bottom*). The full spectrum at millimeter wavelengths (*top*) was obtained from instruments aboard the *COsmic Background Explorer* (*COBE*) in late 1989. These data are so accurate that the error bars of the individual points all lie within the width of the plot curve. This solid line, which matches the shape and peak location of the observed data, corresponds to a thermal radiator, or blackbody, with a temperature of 2.725 K

$h = 6.6261 \times 10^{-34}$ J s. The number density N_{CMB} of photons in the background radiation can be determined from:

$$N_{CMB} = \frac{aT^4}{h\nu_{max}} \approx 4 \times 10^8 \text{ m}^{-3}, \tag{15.8}$$

where the radiation energy density is aT^4, the radiation constant $a = 7.5656 \times 10^{-16}$ J m^{-3} K^{-4}, and $h\nu_{max}$ is the photon energy. Thus, every cubic meter of space in the observable universe contains about a half billion photons of the cosmic microwave radiation. These are tiny bundles of radiation energy that originated about 14 billion years ago.

15.2.3 As Smooth as Silk

What alerted astronomers to the importance of the background radiation was its equal brightness wherever one looked, indicating that it uniformly fills all of space (Wilson and Penzias 1967). This spatial isotropy satisfied one of the basic tenets of modern cosmology, *the cosmological principle*, which asserts that except for local irregularities, the universe presents the same aspect from every point.

But the radiation seemed too uniform. The *COBE* instruments could detect no regions brighter than others to 0.0003 K, or 1 part in 10,000, on angular scales from minutes of arc to 180°. Yet, the background radiation ought to have concentrations in it, which acted as seeds for the subsequent formation of the material universe. They must have acted as a template or blueprint, encoding the information required to explain the subsequent formation of stars and galaxies.

15.2.4 Cosmic Ripples

In 1992, George Smoot (1945–) and his colleagues announced measurements of the temperature fluctuations in the cosmic microwave background radiation using four years of data gathered by *COBE* (Smoot et al. 1992; Bennett et al. 1993, 1996). Aftersubtracting the known microwave emission of the Milky Way and using mathematical averaging techniques on about 100 million observations, the *COBE* team found that the temperature varies ever so slightly over large angular sizes. The sensitive instrument detected minute temperature differences no larger than a hundred-thousandth, or 10^{-5} K. Mather and Smoot were awarded the 2006 Nobel Prize in Physics for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation.

COBE was pushed to the limits of its sensitivity, with evidence that wasn't quite definitive, mainly because it mapped the cosmic radiation with coarse angular resolution greater than 7°. In the subsequent decade, more than 20 experiments were therefore carried out from the ground and balloon platforms, bringing the temperature fluctuations into shaper focus with angular resolutions as fine as a few minutes of arc. However, the ground-based and balloon experiments only

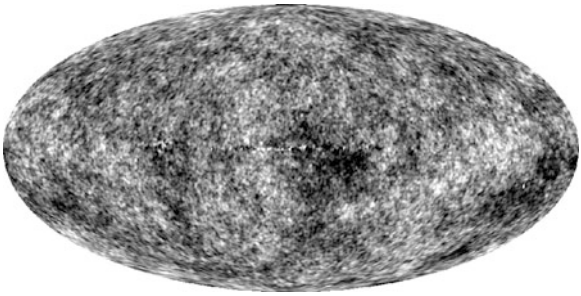


Fig. 15.2 Maps of the infant universe An all-sky view of the three-degree cosmic microwave background radiation emitted from the universe in its infancy, just 390,000 years after the big bang that occurred 13.7 billion years ago. The data, taken in 2003 from the *Wilkinson Microwave Anisotropy Probe* (WMAP) are shown here after seven years of data analysis. The temperature fluctuations range up to 0.0002K above and below the average value. Darker regions are cooler and lighter regions are hotter. These temperature fluctuations provided the seeds from which galaxies subsequently grew. (Courtesy of the NASA/*COBE* and NASA/*WMAP* science teams.)

glimpsed small portions of the sky for a limited time, so there was a possibility that they might not be truly representative of all the background radiation.

It was therefore time for another satellite experiment that would scan the entire sky without the confusion of microwave radiation from the atmosphere and ground. This time, the spacecraft would not only detect the cosmic ripples; it instead would determine their distribution and characteristic sizes, filling in the gaps between the large features seen with *COBE* and the smaller features detected by other instruments.

David T. Wilkinson (1935–2002), of Princeton University, joined Charles L. "Chuck" Bennett (1956–) of the Goddard Space Flight Center to create a small team of experts and design a spacecraft that could accomplish the goal within the modest, for NASA, budget cap of $70 million in 1994 dollars. The resultant *Microwave Anisotropy Probe* (*MAP*) was approved in mid-1996 and launched on 30 June 2001. The name was changed in early 2003 to *Wilkinson Microwave Anisotropy Probe* (*WMAP*), to honor Wilkinson after his death. Instruments aboard *WMAP* provided definitive measurements of the rippling departures from uniformity (Fig. 15.2), with temperature fluctuations of 1 part in 100,000, or at about 0.00003 K (Bennett et al. 2003). This anisotropy, at the level of $\Delta T/T = (1.1 \pm 0.1) \times 10^{-5}$, is given with other physical properties of the background radiation in Table 15.1. Hu and Dodelson (2002) have reviewed cosmic microwave background anisotropies.

When combined with previous measurements, the *WMAP* instruments showed that temperature variations are concentrated within certain angular sizes that are displayed in an angular power spectrum – a plot of the relative strength of the hot and cold spots against their angular sizes (Fig. 15.3). This spectrum is not flat – it is rippled. Gravity explains the ripples, the relative amplitudes of which can be used to infer the gravitational pull that caused them.

Table 15.1 Physical properties of the cosmic microwave background radiation

Parameter	Name	Value
$T_0 = T_{CMB}$	Temperature	2.725 ± 0.002 K
N_{CMB}	Photon density	$(410.4 \pm 0.9) \times 10^6$ m^{-3}
ρ_{CMB}	Mass-energy density of photons	4.648×10^{-31} kg m^{-3}
T_1	Dipole anisotropy	0.003346 ± 0.000017 K
T_1/T_0	Dipole anisotropy/temperature	0.001228
$\Delta T/T_0$	Anisotropy, dipole removed	$(1.1 \pm 0.1) \times 10^{-5}$
$T_2 = Q_{rms}$	Quadrupole moment	$(8 \pm 2) \times 10^{-6}$ K

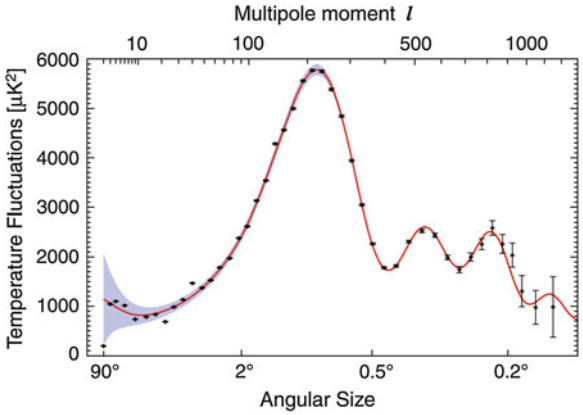


Fig. 15.3 Ripple data The angular fluctuation strength, or power, of the cosmic microwave background radiation in which temperature fluctuations, in units of square micro kelvin (10^{-6} K and designated μK), are displayed as a function of their angular extent in degrees, denoted 0. This plot shows the relative brightness for the all-sky map observed from the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) (see Fig. 15.2) at various sizes. The solid line is the model that best fits the observed data (*solid dots*); the gray band represents uncertainties in the model. Anisotropy data obtained by previous experiments are denoted by dots with error bars. The observed power spectrum has been compared to other astronomical observations and different theoretical models, providing estimates for the amount of dark matter and dark energy in the universe (see Fig. 15.10). (Courtesy of the NASA/*WMAP* Science Team.)

The ratio of the heights of the first and second peak of the angular power spectrum was used to determine the amount of "ordinary" matter with which we are familiar, the baryonic type that comprises atoms. The neutrons and protons found in the nuclei of all atoms are *baryons*.

When the height of the third peak was compared to the other two, scientists estimated the amount of dark, nonbaryonic matter. The comparison indicated that dark matter is five times more abundant than ordinary baryonic matter, and that the combined gravitational pull of both kinds of types of matter is not enough to stop the future expansion of the universe (Table 15.2).

The *COBE* and *WMAP* results have carried cosmology beyond the esoteric realms of theoretical speculation and into precise scientific tests. Definitive new

Table 15.2 Cosmological parameters inferred from *WMAP* observations[a]

Parameter	Name	Value
H_0	Hubble constant[a]	71.0 ± 2.5 km s^{-1} Mpc^{-1}
t_0	Age of expanding universe	$(13.75 \pm 0.13) \times 10^9$ years (after the big bang)
t_{eq}	Equality of matter and radiation (redshift $z_{eq} = 3196 \pm 133$)	76,000 ± 5,000 years
t_{dec}	Decoupling (recombination) (redshift $z_{dec} = 1090.89 \pm 0.69$)	$(3.79 \pm 0.05) \times 10^5$ years (after the big bang)
t_{reion}	Reionization time (redshift $z_{reion} = 10.5 \pm 1.2$)	$(3.5 \pm 1.5) \times 10^8$ years (after the big bang)
$\Omega_b h^2$	Baryonic matter ($\Omega_b = 0.0449 \pm 0.0028$)	0.02258 ± 0.00057
$\Omega_c h^2$	Dark matter density ($\Omega_c = 0.222 \pm 0.026$)	0.1109 ± 0.0056
$\Omega_m h^2$	Total matter density ($\Omega_m = 0.267 \pm 0.026$)	0.1335 ± 0.0056
Ω_Λ	Dark energy density	0.734 ± 0.029
Ω_{tot}	Total density: matter + Energy	1.08 ± 0.09

[a] Parameter values are from *WMAP* only, adapted from Jarosik et al. (2011). The Hubble constant $H_0 = 100\,h$ km s^{-1} Mpc^{-1}, or $h \approx 0.71$ for these estimates. A more accurate value indicates that $H_0 = 75$ km s^{-1} Mpc^{-3}. The parameter Ω is the ratio of the specified quantity to the critical amount required to keep the expansion of the universe on the brink of closure. The matter density parameter, for example, is $\Omega_m = \rho_m(t_0)/\rho_C$. Where $\rho_m(t_0)$ is the total mass density, in visible and invisible form, at the present time t_0, and $\rho_C = 3H_0^2/(8\pi G) \approx 1.0 \times 10^{-26}$ kg m^{-3} for $H_0 = 75$ km s^{-1} Mpc^{-1} is the critical mass density required to stop the expansion of the universe in the future. The total density parameter Ω_{tot} is the sum of the contributions from visible matter, dark matter, and dark energy, and $\Omega_{tot} = 1.00$ is consistent with inflation and a universe that is described by Euclidean geometry without space curvature.

observational descriptions of the background radiation, with refined cosmological consequences, are expected from the *Planck* mission, launched in May 2009. The initial *Planck* results, announced on March 21, 2013, indicate a Hubble constant of $H_0 = 67.15 \pm 1.2$ km s^{-1} Mpc^{-1}, and a universe with a baryonic (normal) matter content of 4.9 percent ($\Omega_b = 0.049$), a dark matter content of 26.8 percent ($\Omega_c = 0.268$), a total matter density of 31.7 percent ($\Omega_m = 0.317$) and a dark energy content of 68.3 percent ($\Omega_\Lambda = 0.683$). The values of these parameters inferred from the *WMAP* results are given in Table 15.2.

15.3 The Beginning of the Material Universe

15.3.1 The First Three Minutes

George Gamow (1904–1968) and his colleagues proposed that the first elements were formed during the big bang that propelled the universe into expansion (Alpher et al.1948, Sect. 10.5). As they supposed, the lightest atomic nuclei were

synthesized before any atoms were created, during the first three minutes following the big bang explosion; the nuclei of less abundant, heavier atoms were manufactured at a later time, within the cores of stars.

The Japanese astrophysicist Chushiro Hayashi (1920–2010) improved their scenario by noting that the temperature was hot enough in the immediate aftermath of the big bang to create neutrinos and positrons, the antimatter particles of the electron, and equilibrium was established between electrons, positrons, neutrons, protons and radiation (Hayashi 1950).

The positrons, denoted e^+, interacted with the neutrons, denoted n, to form protons, p, by the reaction:

$$e^+ + n \rightleftarrows p + \bar{\nu}_e, \tag{15.9}$$

where $\bar{\nu}_e$ denotes an antielectron neutrino. The double arrow in these equations indicates that the reverse reactions happen just as often as the forward reactions, so positrons and neutrons were being produced just as often as they were being consumed. The neutrons also interacted with the electron neutrinos, ν_e to form protons by:

$$\nu_e + n \rightleftarrows p + e^-, \tag{15.10}$$

where e^- denotes an electron.

And in the meantime, the radiation was producing electron–positron pairs, which turned back into radiation just as rapidly, through the reactions:

$$\gamma + \gamma \rightleftarrows e^- + e^+, \tag{15.11}$$

where γ denotes a photon of gamma ray radiation.

During the first few seconds of the expanding universe, when the radiation temperature was greater than 10^9 K, there were no atoms, just radiation and subatomic particles, like neutrons, protons, electrons, positrons, and neutrinos. The equilibrium number density ratio of neutrons, N_n, and protons, N_P, just depended on the mass difference between neutrons and protons, denoted $m_n - m_p$, and the temperature T, according to the relation (Hayashi 1950):

$$\frac{N_n}{N_p} = \exp\left[\frac{-(m_n - m_p)c^2}{kT}\right], \tag{15.12}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the neutron mass $m_n = 1.6749274 \times 10^{-27}$ kg, the proton mass $m_p = 1.6726218 \times 10^{-27}$ kg, and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. The mass difference $\Delta m = m_n - m_p = 2.3056 \times 10^{-30}$ kg, and the energy $\Delta E = (m_n - m_p)\, c^2 = \Delta m c^2 \approx 2.07 \times 10^{-13}$ J, which corresponds to a temperature $T = \Delta E/k \approx 1.5 \times 10^{10}$ K.

As the temperature decreased, the number density ratio N_n/N_p also decreased and the protons outnumbered the heavier neutrons. Eventually, the reactions were no longer in thermodynamic equilibrium, neutrons could no longer be created, and the neutron-to-proton ratio became "frozen-in" at an amount that determined the

abundance of helium synthesized. Modern computations by David Schramm (1945–1977) and others have conclusively demonstrated that all of the hydrogen and most of the helium nuclei, which now are found in the universe, were indeed synthesized in the immediate aftermath of the big bang (Sect. 10.5, Peebles et al. 1991).

Example: Big-Bang nucleosynthesis of helium

In the first moments of the big bang, the neutrons and protons were in thermodynamic equilibrium with a number density ratio $N_n/N_p = \exp[-\Delta mc^2/(kT)]$, where Δm is the difference between the mass of the neutron and the mass of the proton, c is the speed of light, $\Delta mc^2 \approx 2.07\times 10^{-13}$ J, the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1} and T is the temperature. Just a few seconds after the big bang, the temperature had cooled to just above 10^9 K, and the production of both positrons and neutrons stopped. The leftover positrons then were consumed by interactions with electrons, and an equilibrium was established in which the relative amounts of neutrons and protons were governed by their mass difference and the temperature. Hayashi (1950) estimated that this frozen-in abundance ratio was $N_n/N_p \approx 0.25$. Alpher et al. (1953) obtained lower amounts of between 0.17 and 0.22 using detailed calculations that depended on the time it takes a free neutron to decay into a proton; outside a nucleus free neutrons are unstable and have a mean lifetime of 881.5 ± 1.3 s or about 14 min 42 s. A lower limit for the "frozen-in" neutron-proton ratio is obtained when the thermal energy kT equals $m_e c^2$, the rest-mass energy of the electron or $N_n/N_p = \exp\left[-(m_n - m_p)/m_e\right] \approx \exp(-2.53) \approx 0.08$, where the electron mass $m_e = 9.10938 \times 10^{-31}$ kg.

If N_n and N_p respectively denote the number densities of neutrons and protons before helium nuclei were synthesized, then $N_n/2$ helium nuclei will be formed, since each helium nucleus contains two neutrons and two protons, and the number of protons left over is $N_p - N_n$. The relative number densities of helium, $N(^4He)$, nuclei to hydrogen, $N(H)$, nuclei is:

$$\frac{N(^4He)}{N(H)} = \frac{N_n/2}{N_p - N_n} \approx \frac{1}{12} \approx 0.08. \tag{15.13}$$

The helium mass fraction, Y, is:

$$Y = \frac{4N(^4He)}{N(H) + 4(^4He)} \approx 0.25. \tag{15.14}$$

About one quarter of the mass of the material universe, in baryons, was synthesized into helium in the first few minutes of the expanding universe, and this is consistent with the amount of helium that is now observed in the universe.

Once neutron production stopped, the neutrons and protons could begin combining to form the nuclei of deuterium and helium atoms in amounts governed by the frozen-in abundance ratio of neutrons and protons. All of the protons that did not participate in forming these deuterium and helium nuclei eventually became the nuclei of hydrogen atoms. The production of light atomic nuclei was over, with vastly more hydrogen left behind than anything else.

15.3.2 Formation of the First Atoms, and the Amount of Invisible Dark Matter

Whole atoms were not formed until the expanding universe cooled enough for electrons to combine with protons and helium nuclei to form long-lived hydrogen and helium atoms. This recombination occurred about 400,000 years after the big bang, when the temperature had fallen to about 3,000 K. The rate of recombination was then higher than the rate of ionization by the intense radiation. By the end of recombination, all of the nuclei and electrons had been bound up in atoms, and the universe became transparent to the radiation that then could travel through space without scattering off free, unattached electrons. The cosmic microwave background radiation that we observe in the present was released back then, almost 14 billion years ago.

Because there is no stable nucleus of atomic mass 5 or 8, elements heavier than helium (of mass 4) could not be synthesized by successive collisions with protons (of mass 1). Big-bang nucleosynthesis therefore stopped at helium 4. Heavier elements needed to be synthesized inside stars where the densities are high enough for triple collisions of helium to form carbon, rather than the big bang in which the density had become too low by the time helium nuclei were formed for triple collisions to become significant at the prevailing temperature.

The nuclei of the hydrogen and deuterium atoms and most of the nuclei of helium atoms that now are present in the universe were synthesized in the first 3 min of the expansion, in the immediate aftermath of the big bang and about 14 billion years ago. All of the hydrogen found in stars and interstellar space, or in the Earth's water, and in our body, was produced by this big-bang. And every time you buy a floating party balloon, which has been inflated by helium, you are getting atoms made about 14 billion years ago. Deuterium is destroyed inside stars and, although helium is synthesized in main-sequence stars, the amount of helium formed inside stars over the lifetime of the expanding universe is no more than 10 % of what is now observed in cosmic objects (Hoyle and Tayler 1964).

The cosmological implications of big-bang nucleosynthesis are profound! The agreement of light-element abundances and predictions from the primordial nuclear reactions works only if the density of ordinary matter in the universe – in both visible and invisible forms – is less than 10 % of the critical mass density, ρ_C, required to eventually stop the expansion of the universe.

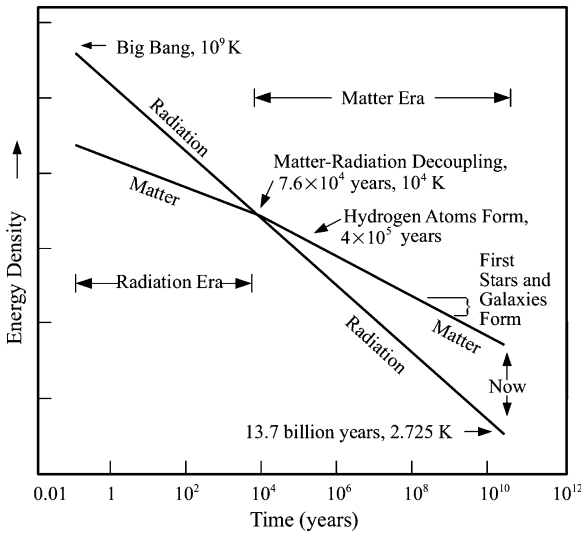


Fig. 15.4 Key events in the expanding universe Shortly after the big bang (*left*) the temperature was about 10^9 K for both radiation and matter, and with its greater energy density radiation dominated the expansion of the universe. With increasing time, the radiation energy density dropped faster than the matter energy density, until the two became equal about 76,000 years after the big bang when the temperature had dropped to about 10^4 K. Thereafter the radiation went its separate way, cooling to the 2.725 K we now detect in the cosmic microwave background radiation 13.7×10^9 years after the big bang (*right*). The matter became cool enough for protons and electrons to recombine and make hydrogen atoms about 400,000 years after the big bang, at a temperature of about 3,000 K, and the first stars and galaxies formed between 10^8 and 10^9 years after the big bang

Observations of cosmic objects with low heavy-element (metals) content indicate that the primordial abundance of helium, or the amount created before the first stars formed and synthesized any helium or heavy elements, is roughly $Y = 0.25$, or 25 % by mass. *WMAP* observations of patterns in the cosmic microwave radiation provide evidence for the presence of helium long before the first stars formed with $Y = 0.326 \pm 0.075$ (Jarosik et al. 2011). The abundance of deuterium, D, relative to hydrogen, H, which is $D/H \geq 10^{-5}$, sets an upper limit to the baryon density parameter of $\Omega_B = \rho_{B0}/\rho_C \leq 0.025\, h^{-2} \approx 0.05$, where the Hubble constant $H_0 = 100\, h \approx 75$ km s^{-1} Mpc^{-1} and the critical mass density required to close the expanding universe in the future is $\rho_C = 3H_0^2/(8\pi G) \approx 10^{-26}$ kg m^{-3} for $H_0 = 75$ km s^{-1} Mpc^{-1}. The number of baryons is equal to the number of neutrons and protons in matter in either seen or unseen forms.

When combined, the big-bang nucleosynthesis constraints of all four of the light nuclei, ^4He, D, ^3He, and ^7Li, provide limits to the baryon to photon number density ratio, $\eta = N_B/N_\gamma \approx 3 \times 10^{-10}$, which has remained unchanged since the epoch of electron–positron annihilation a few seconds after the big bang. Here N_B

is the number density of baryons and N_γ is the number density of photons, so the universe contains less than one baryon (neutrons and protons) per billion photons. During the time of *big-bang nucleosynthesis*, the universe was a dilute gas of radiation photons, contaminated by only trace amounts of baryons. At present the number density of photons in the cosmic microwave background radiation is $N_{\gamma 0} \approx 4 \times 10^8\,\mathrm{m}^{-3}$. The baryon density now, $\rho_{B0} = m_p N_{\gamma 0} \eta \approx 2 \times 10^{-28}\,\mathrm{kg\,m}^{-3}$, where the proton mass $m_p = 1.673 \times 10^{-27}$ kg, and the baryon density parameter $\Omega_B = \rho_{B0}/\rho_C \approx 0.05$ (Copi et al. 1995).

Thus, when measurements of the Hubble constant are considered, the results of big-bang nucleosynthesis and the observations of the abundance of light elements indicate that the baryon density is now 0.05 or 5 %, of the critical mass density required to ever halt the current expansion of the universe in the future.

A completely independent estimate of the baryon density is provided by the power spectrum of the anisotropies in the cosmic microwave background radiation. The relative amplitudes of the peaks in the spectrum constrain it to $\Omega_B = 0.044 \pm 0.004$, in excellent agreement with the conclusions based on big-bang nucleosynthesis.

15.3.3 History of the Expanding Universe

As the universe grew larger, the radiation energy density, $\rho_r(t)$, decreased more rapidly than the mass density, $\rho_m(t)$. Eventually, at the time $t_{eq} \approx 76{,}000$ years after the big bang, the mass-energy density of the radiation had become equal to that of the matter. Thereafter it was mass that dominated the expansion of the universe. This and other critical times in the expansion are illustrated in Fig. 15.4. Loeb and Barkana (2001) have reviewed the re-ionization of the universe by the first stars and quasars.

If $R(t)$ denotes the radius, or scale factor, of the universe at cosmic time, t, the radiation temperature, $T_r(t)$, falls off as $1/R(t)$. This means that the radiation energy density (Tolman 1934):

$$\rho_r(t) = \frac{aT_r^4(t)}{c^2} \propto \frac{1}{R^4(t)}, \tag{15.15}$$

where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}, the speed of light $c = 2.9979 \times 10^8$ m s^{-1}, and the symbol $\propto$ denotes proportional to. The mass density, $\rho_m(t)$, doesn't decrease as rapidly with increasing time and radius, since it thins out with increasing volume, or with

$$\rho_m(t) \propto \frac{1}{R^3(t)}. \tag{15.16}$$

Since

$$T_r(t) \propto \frac{1}{R(t)}, \tag{15.17}$$

the ratio

$$\frac{\rho_m(t)T_r(t)}{\rho_r(t)} = T_r\left(t_{eq}\right) \tag{15.18}$$

is a constant for all times t. Here t_{eq} denotes the time when the radiation energy density is equal to the matter density, or when $\rho_r\left(t_{eq}\right) = \rho_m\left(t_{eq}\right)$

Example: Radiation energy density and mass energy density
As the universe expands, the radiation energy density $\rho_r(t)$ at time t will fall off as the inverse fourth power of the radius of the universe, $R(t)$, since the radiation temperature $T_r(t)$ falls off as $1/R(t)$. The mass energy density $\rho_m(t)$ falls off as the inverse cube of the radius, since the volume increases as the radius cubed. This means that $\rho_m(t)T_r(t)/\rho_r(t) =$ constant for all time t. At the present time t_0 we have $T_r(t_0) = 2.725$ K, the $\rho_r(t_0) = a\,T_r^4(t_0)/c^2 \approx 0.842 \times 10^{-32}\,T_r^4(t_0) \approx 4.64 \times 10^{-31}\,\mathrm{kg\,m^{-3}}$, where the radiation constant $a = 7.5657 \times 10^{-16}\ \mathrm{J\ m^{-3}\ K^{-4}}$ and the speed of light $c = 2.9979 \times 10^8\ \mathrm{m\ s^{-1}}$. *WMAP* observations indicate that the density parameter Ω_b of normal baryonic matter, like protons and neutrons that are found in atoms, is $\Omega_b = \rho_m(t_0)/\rho_c \approx 0.04$ (Table 15.2), where $\rho_m(t_0)$ is the present mass density in both visible and invisible baryons and the critical mass density required to stop the expansion of the universe in the future is $\rho_c = 3H_0^2/(8\pi G) \approx 1.0 \times 10^{-26}\,\mathrm{kg\,m^{-3}}$ (Sect. 14.6), the Hubble constant $H_0 = 100\,h\ \mathrm{km\ s^{-1}\ Mpc^{-1}} = 3.24 \times 10^{-18}\,h\ \mathrm{s^{-1}}$, with $h \approx 0.75$, $1\ \mathrm{Mpc} = 3.0857 \times 10^{19}\ \mathrm{km}$, and the Newtonian gravitational constant $G = 6.674 \times 10^{-11}\ \mathrm{m^3\ kg^{-1}\ s^{-2}}$. This means that $\rho_m(t_0) \approx 0.04\,\rho_c \approx 4 \times 10^{-28}\,\mathrm{kg\,m^{-3}}$, about a thousand times greater than the present radiation energy density, and that $\rho_m(t_0)T_r(t_0)/\rho_r(t_0) \approx 2 \times 10^3$.

The radiation energy density and mass energy density were equal at a time $t_{eq} \approx 76{,}000$ years $\approx 2.40 \times 10^{12}$ s, where one year $= 3.1557 \times 10^7$ s, after the big bang (Table 15.2). At this time, the radiation temperature has dropped to $T_r(t_{eq}) = 2.15 \times 10^{10}\ t_{eq}^{-1/2} \approx 10^4$ K, and $\rho_m\left(t_{eq}\right)T_r\left(t_{eq}\right)/\rho_r\left(t_{eq}\right) = T_r\left(t_{eq}\right) \approx 10^4$, close enough to the value of this ratio now considering the uncertainties.

The radiation dominated era, when the radiation energy density is greater than the matter density, has a radius that scales as the square root of time, t, or

$$R(t) \propto t^{1/2} \quad \text{for } \rho_r(t) > \rho_m(t) \text{ or for } t < t_{eq}, \tag{15.19}$$

and when the matter dominates the expansion of the universe, its radius is:

$$R(t) \propto t^{2/3} \text{ for } \rho_m(t) > \rho_r(t) \text{ or for } t > t_{eq}. \tag{15.20}$$

If radiation was emitted at time t, then its redshift, z, is given by

$$1 + z = \frac{R(t_0)}{R(t)}, \tag{15.21}$$

and the radiation temperature, $T_r(t)$ at that time is given by

$$T_r(t) = (1 + z)\, T_r(t_0) = 2.725\,(1 + z), \tag{15.22}$$

where the temperature of the background radiation $T_r(t_0) = 2.725 \pm 0.002$ K right now, at time t_0.

WMAP observations indicate that the radiation energy density became equal to the matter density at a redshift of $z_{eq} = 3196 \pm 133$ (Jarosik et al. 2011). Since that time, denoted as t_{eq} for the equality of radiation and matter, the matter dominated the expansion and $R(t)$ grew as $t^{2/3}$, so the transition from the radiation dominated era to the matter dominated era occurred at time t_{eq} where:

$$\frac{t_{eq}}{t_0} = \left[\frac{R(t_0)}{R(t_{eq})}\right]^{3/2} = (1 + z)^{3/2}, \tag{15.23}$$

and that

$$t_{eq} = \frac{t_0}{(1 + z_{eq})^{3/2}} \approx 76,000 \text{ years} \tag{15.24}$$

after the big bang. This indicates that matter has dominated the expansion of the universe for all but a relatively small fraction of the age $t_0 \approx 13.7 \times 10^9$ years of the expanding universe.

Another crucial time is the *decoupling time*, or *recombination time*, t_{dec}, when the temperature had fallen to about 3,000 K, which was cool enough for electrons to begin recombining with protons and helium nuclei to form long-lived hydrogen and helium atoms. That is, the rate of recombination to form atoms was then higher than the rate of atomic ionization by the intense radiation. By the end of recombination, all the nuclei had been bound up in atoms, and the universe became transparent to the radiation; so cosmic microwave background radiation that we observe at the present time was released back then, about 13.3 billion years ago.

The decoupling redshift corresponding to this temperature is $z_{dec} \approx 3{,}000/2.725 \approx 1100$, and $t_{dec} = t_0/(1100)^{3/2} \approx 377{,}000$ years. The *WMAP* estimates give $z_{dec} = 1090.89 \pm 0.69$. This also marks the beginning of the dark ages of the expanding universe, for there were no sources of radiation other than the gradually cooling and darkening cosmic background radiation until stars and galaxies

Table 15.3 Crucial times during the expansion of the universe

Time (after the big bang)	Redshift, z	Temperature (K)	Key events
10^{-14} s	10^{27}	10^{27}	Inflation ends, $\Omega_m + \Omega_\Lambda = 1$.
10 s	4×10^9	10^{10}	Neutron and positron production stops.
3 min	4×10^8	10^9	Big-bang nucleosynthesis ends, light elements H, D, He formed
$t_{eq} \approx$ 76,000 year	3,196	10^4	Radiation domination equals matter domination, transfer from $R(t) \propto t^{1/2}$ to $R(t) \propto t^{2/3}$
$t_{dec} \approx$ 377,000 year	1,100	3,000	Decoupling (recombination) time. radiation decouples from matter, hydrogen atoms recombine, universe becomes transparent to background fluctuations, dark ages begin
$t_{reion} \approx 3.5 \times 10^8$ year	10.5 ± 1.2	31	Reionization time, first stars and galaxies form, universe re-ionized by their radiation, and dark ages end
6×10^9 year	1	5	Dark energy begins acceleration of universe expansion
$t_0 = 13.7 \times 10^9$ year	0	2.725	Age of expanding universe, today, present epoch

formed, about 350 million years after the big bang, they provided beacons of bright light that could ionize surrounding matter. However, by then the universe had thinned out enough that the low-density ionized hydrogen remained transparent, ending the dark ages.

A summary of these milestones in the history of the expanding universe is given in Table 15.3.

Example: When did the first stars and galaxies form?
When atoms coalesced to form stars and galaxies, these shining beacons of intense radiation illuminated the former darkness and re-ionized nearby matter. The *WMAP* observations suggest that this began at a redshift $z_{reion} = 10.5 \pm 1.2$. This would happen at a background radiation temperature $T = (1 + z)\, 2.725 = 31$ K. The corresponding time, t_{reion}, is $t_{reion} = t_0/(1 + z_{reion})^{3/2} \approx 3.5 \times 10^8$ years or about 350 million years after the big bang. Since matter decoupled from radiation, and the dark ages began, about 76,000 years after the big bang, the dark ages lasted about 350 million years and stars and galaxies have been around for about $t_0 - t_{reion} \approx 13.4$ billion years, where the big bang occurred about $t_0 = 13.75$ billion years ago. These stars and galaxies have redshifts of 10.5 or less.

15.4 The First Stars and Galaxies

15.4.1 Pulling Primordial Material Together

Immediately after the big bang, the radiation and matter were distributed smoothly, with almost no structure at all; because the subatomic neutrons and protons were then in thermal equilibrium with the radiation, the material universe must have had a smooth beginning. Departures from complete uniformity in the cosmic microwave background radiation were about one part in one hundred thousand, or with fluctuations ΔT in the temperature T of $\Delta T/T \approx 10^{-5}$. Yet, the material universe we see today has gathered together into stars, galaxies, and clusters of galaxies, so something must have given them shape and form. They are composed of baryonic matter, which includes neutrons, protons, and atoms of any sort, and thus consist of ordinary matter that we encounter or experience in everyday life.

When left alone, a spread out distribution of matter will coalesce about small, initial concentrations, as a result of their gravitational pull on surrounding material, so any slight perturbations in otherwise uniform matter will grow and eventually contract. Given enough time, gravity might magnify the extremely slight irregularities in the initial mass distribution, eventually providing the concentrated structures, the observed stars and galaxies, by gravitational collapse.

The problem with this scenario, first recognized by Georges Lemaître (1894–1966) and rigorously derived by Evgeny Lifshitz (1915–1985), is that the initial irregularities could not have grown fast enough to account for the observed stars and galaxies (Lemaître 1934; Lifshitz 1946). Chance density fluctuations in this matter would have grown too slowly to overcome and resist the overall expansion of the universe, which pushes and tears the material apart as soon as it starts to gather together. Even a billion years is not enough time for the fluctuations in ordinary matter density to gravitationally pull the primeval matter together into luminous stars and galaxies.

We can follow the basic argument by supposing that an initial density perturbation $\Delta\rho_m$ in the mass density ρ_m was of order $\Delta\rho_m/\rho_m = 10^{-5}$, since the subatomic matter was in thermal equilibrium with the radiation with temperature fluctuations of $\Delta T/T = 10^{-5}$. In the matter-dominated era the perturbations will slowly grow with increasing time, t, and the scale factor $R(t)$ as $\Delta\rho_m/\rho_m \propto t^{2/3} \propto R(t)$, but even over the past 13 billion years $R(t)$ has only increased by a factor of a thousand, or 10^3, so the initial perturbations would have grown to $\Delta\rho_m/\rho_m = 10^{-2}$, much smaller than the amount of $\Delta\rho_m/\rho_m = 1$ now observed in galaxies.

However, we have considered only ordinary matter, and the paradox could be resolved if the extra gravitational pull of much greater amounts of invisible dark matter helped clump and shape the expanding universe, pulling together

primordial fluctuations in perfectly ordinary matter and forming the seeds from which the first stars and galaxies grew.

According to the *cold dark matter* scenario, galaxies formed first, than gravitationally merged and consolidated into clusters and super-clusters of galaxies as the universe expanded and evolved; for pioneering papers on the cold dark matter hypothesis see Lang (1999).

Understanding how gravity causes the perturbations in ordinary matter to grow in an expanding universe and eventually become galaxies requires studying the interaction between ordinary matter and dark matter. That interaction causes a region of space with more ordinary matter than average to oscillate, sending out waves known as baryonic acoustic oscillations. These sound waves were first predicted in 1970, suggested by *WMAP* fluctuations of the cosmic microwave background in 1999, and measured as rippling imprints in the distribution of galaxies using Sloan Digital Sky Survey data in 2005 (Peebles and Yu 1970; Eisenstein et al. 2005).

15.4.2 When Stars Began to Shine

With the help of cold dark matter, the first stars and galaxies appeared more than 10 billion years ago. We can observe these embryonic galaxies when they were cosmic infants; the light now reaching us began its journey long before the Sun came into existence.

Bromm and Larson (2004) have discussed the first stars; Bromm and Yoshida (2011) provided a review of the first galaxies; Brodie and Stader (2006) discussed extragalactic globular clusters and galaxy formation, and Kravtsov and Borgani (2012) has reviewed the formation of galaxy clusters.

Kennicutt and Evans (2012) have reviewed star formation near and far. Shapley (2011) has reviewed the physical properties of galaxies from redshifts $z = 2$–4, and Giavalisco (2002) has discussed Lyman-break galaxies. Sanders and Mirabel (1996) have reviewed luminous infrared galaxies.

Each galaxy may have formed through the gravitational collapse of a larger, protogalactic cloud, which would become a rotating disk like the Milky Way (Eggen et al. 1962). These flattened, spinning galaxies often show spiral structure, with arms of gas and dust in which new stars are forming.

Not all galaxies have a disk or spiral shape, and the most massive are the giant, rounded, featureless elliptical galaxies. They may result from the collision and subsequent merger of two spiral galaxies. During the encounter, the ordered rotational motions of the stars in the spiral galaxies are transformed by tidal forces, which tear their disks and arms apart and randomize the orbits of their stars. When the merger is complete, a single elliptical galaxy remains, composed of old stars with little or no gas and dust left to form new stars. Many of the giant elliptical galaxies are found in the cores of dense clusters of galaxies where collisions should be frequent on a cosmic time scale.

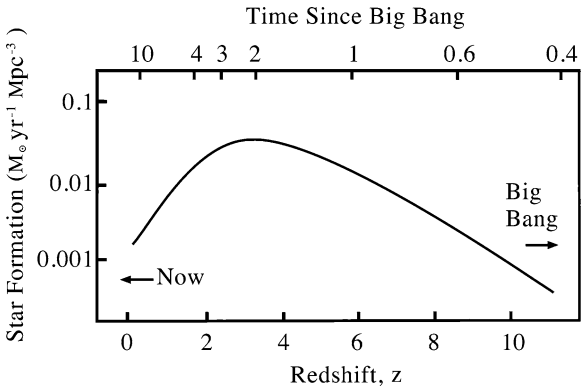


Fig. 15.5 Star–formation rates The star-formation rate, in solar masses per year per cubic Megaparsec, or $M_\odot$ year^{-1} Mpc^{-3}, plotted as a function of redshift, z (*bottom axis*) and time since the beginning of the expanding universe (*top axis*), in units of 10^9 years, or 1 billion years and a G year. The rate of star formation peaked at a redshift of about 3, or roughly 2 billion years after the expansion began, and this rate subsequently has decreased as gravitation pulls more material into stars. (From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

When astronomers use infrared telescopes, aboard the *Herschel* and *Spitzer* spacecraft, they can peer behind veils of local interstellar dust to see infant stars in distant galaxies. Some *starburst galaxies* are very powerful infrared emitters, with an infrared luminosity greater than a million million, or 10^{12}, times that of the Sun, and an infrared output that is 100 times their visible-light emission. An exceptional amount of interstellar dust in these galaxies absorbs the intense ultraviolet radiation produced by enhanced star formation, and the dust reradiates in the infrared part of the electromagnetic spectrum. Their intense infrared emission and implied dust suggests that these galaxies are forming stars more vigorously than our present-day Milky Way (Fig. 15.5).

Starburst galaxies at large redshift of about $z = 3$ are found by a clever comparison of the ultraviolet and visible radiation of galaxies. The technique is related to the main ultraviolet spectral transitions of hydrogen, the Lyman alpha line at 121.7 nm, which occurs between electron orbits with quantum numbers $m = 2$ and $n = 1$, and all the other Lyman transitions at larger m and the same n, culminating at the Lyman limit at 91.2 nm at very large m. The observed Lyman limit at large redshifts is Doppler shifted into visible wavelengths at $\lambda = 91.2\ (1 + z)$, which is at 360 nm for $z = 3$. Radiation at wavelengths lower than the Lyman limit is almost completely absorbed by the alpha transition of neutral hydrogen in the star-forming regions. At large redshifts the sharp decrease, or break, of the emitted spectrum has been Doppler shifted into the visible region. The starburst galaxies are therefore also known as *Lyman-break galaxies*.

About 2 billion years after the big bang and roughly 12 billion years ago, some starburst galaxies had an exceptionally high rate of star formation, which exceeded

100 stars per year for hundreds of millions of years – much greater than the rate in most galaxies and currently in the Milky Way, which is about one new star every year. Scarcely any stars were being formed during the first half billion years of the expanding universe, perhaps because the gravitational forces were still pulling the galaxies together, and after the bursts of star formation there may have been less material available for forming new stars, since some of it already had been used up in creating other stars.

The fast pace of star formation in galaxies in the early universe could not be continued for long times. It would use up the interstellar gas from which stars are formed in much less time than the age of the universe. So the bursts of star formation could be associated with rare circumstances, perhaps feeding off gas stirred up as a result of collisions or close encounters between galaxies.

The mergers of galaxies may not be the dominant method of high star growth. It could be associated with a voracious consumption of hydrogen gas, which has been observed in greater abundance back then when compared to more recent times. A steady supply of gas may have streamed in from filaments of dark matter.

When the first stars formed out of collapsing clouds of gas and ignited the nuclear reactions that make them shine, the early universe consisted of nothing more than the light elements, hydrogen and helium. These young stars must have been uncontaminated by heavier elements. Some of these "infant" stars most likely were very massive, perhaps with about 100 times as much mass as the Sun; therefore, they would have a relatively short lifetime on the cosmic time-scale. The first massive stars would have exploded as supernovae, spewing out ashes of dust made of heavy elements synthesized within them and spawning the next generation of stars.

The interstellar medium would have become steadily enriched in heavy elements as subsequent generations of massive stars were formed, lived, and expired explosively. They would have seeded their surroundings with elements such as carbon, oxygen, and iron, which were needed for the formation of Earth-like planets and life.

Whether a galaxy is young or old, there will always be many more stars of low mass than there are massive stars (see Sect. 10.1). The stellar mass distribution will also depend on the evolution of stars, which varies with mass. The initial mass distribution can be inferred from the observed stellar luminosity function, or the number of stars of different absolute luminosities, by using the stellar mass-luminosity relation together with a model of how the star formation rate varies with time.

The initial mass distribution for stars more massive than the Sun was quantified by the Cornell astronomer Edwin E. Salpeter (1924–2008), who showed that the number of stars with masses in the range M to $M + dM$ within a specified volume of space, is proportional to $M^{-2.35}$ (Salpeter 1955). In other words, the number of stars in each mass range decreases rapidly with increasing mass.

The oldest stars in our Milky Way Galaxy, which were formed when the universe was only about 1 billion years old, are deficient in heavy elements when compared to stars that are now forming in the Milky Way. No one has yet found a

completely pure star that formed out of uncontaminated hydrogen. Perhaps such stars now are inaccessible to direct observation, awaiting the next space telescope that can peer deeper into the remote past.

15.5 The Evolution of Galaxies

15.5.1 Active Galactic Nuclei

By peering out at galaxies that are located at vastly different distances or redshifts, astronomers have shown that the entire observable universe evolves and has a history, and that the properties of galaxies change over vast time scales. The light we detect from a remote galaxy has traveled for a very long time, and was emitted in the galaxy's infancy many billions of years ago. And when our telescopes observe a nearby galaxy, its light may have been generated a few million years ago, after the galaxy has aged for billions of years.

Because no significant change in the equilibrium of galaxies can be produced without a substantial change in the distribution of mass and angular momentum, it was long believed that no significant departures from a stable equilibrium in their shape, form, mass, or luminosity would be produced during most of their lifetimes. Nevertheless, it now is known that the centers of galaxies are locations of pronounced activity that disrupts the expected equilibrium and that galaxies tend to be more active in their youth.

The American astronomer Carl K. Seyfert (1911–1960) provided early observational evidence that the central regions of some galaxies are not in equilibrium when he examined the intense blue centers of certain spiral nebulae – a type subsequently named *Seyfert galaxies* (Seyfert 1943). Although most spirals exhibit spectral lines in absorption, similar to the absorption spectra of stars, the central regions of Seyfert galaxies exhibit intense emission lines of the type produced by ionized emission nebulae (also see Sect. 11.1). They are the emission lines of oxygen [O II], [O III], nitrogen [N II], neon [Ne III] and sulphur [S II], as well as the permitted emission transitions of unionized hydrogen, H, and ionized helium, He II (Table 15.4).

The emission lines of Seyfert galaxies are unexpectedly wide. High-speed motions of the ions and hydrogen atoms have widened the emission lines, and the velocities implied from their widths, when interpreted by the Doppler effect, are up to 8,500 km s^{-1}. Because the central masses derived from the rotation curves of spiral nebulae are no more than 10^{11} solar masses, the escape velocities of the central regions are only a few hundred km s^{-1}. The observed motions at the centers of Seyfert galaxies are therefore far in excess of the expected escape velocities, and they provide the first evidence for violent explosive events in the nuclei of galaxies. Their matter could be flowing out into intergalactic space; some of the Seyfert galaxies exhibit bright filaments that suggest the ejection of gas.

Table 15.4 Intense emission lines found in Seyfert galaxies[a]

Element	Wavelength (nm)	Element	Wavelength (nm)
[O II]	372.62	He I	447.25
[O II]	372.89	He II	468.57
He II	376.89	Hβ	486.1332
Hθ	379.86	[O III]	495.891
He II + Hη	383.56	[O III]	500.684
[Ne III]	386.875	He I	587.56
Hζ + He I	388.89	He II	597.7
[Ne III]	396.746	[O I]	630.0304
Hε	397.01	[O I]	636.3776
[S II]	406.85	[N II]	644.803
[S II]	407.65	Hα	656.281
Hδ	410.17	[N II]	658.341
He II + Hγ	433.86	[S II]	671.647
	+ 434.047	[S II]	673.085
[O III]	436.32		

[a] Adapted from Seyfert (1943)

Weedman (1977) has summarized our then current knowledge of Seyfert galaxies. Early considerations of violent activity in the nuclei of galaxies can be found in Ambartsumian (1958), Burbidge et al. (1963) and Lynden-Bell (1969).

Powerful cosmic radio sources provide additional evidence for intense activity in the central regions of young galaxies. But since there was no radio technique for establishing distances, the optical wavelength counterparts of the radio sources had to be used to determine how far away they were. Research groups led by Joseph L. Pawsey (1908–1962) in Australia and by Martin Ryle (1918–1984) in Cambridge, England built interferometers that were used to obtain accurate positions and identify the optical counterparts of the brightest radio sources, named by the constellation they appeared in. The Australian group identified the source named Virgo A – also numbered 3C 274 in the third Cambridge catalogue of bright radio sources – with giant elliptical galaxy, M 87 (Bolton et al. 1949), and an accurate position established with the Cambridge radio interferometer was used to identify the radio source Cygnus A, numbered 3C 405, with another elliptical galaxy (Baade and Minkowski 1954). Like the nuclei of Seyfert galaxies, the optical counterpart of Cygnus A emits strong "forbidden" emission lines of [O II], [O III], [N II] and [Ne III] with widths corresponding to velocities of a least 1,000 km s^{-1}.

Moreover, the Doppler shift of the central wavelengths of these spectral lines exhibited a redshift of 0.0561, and a recession velocity of 16,820 km s^{-1}. Using the Hubble law with this redshift, the radio galaxy Cygnus A lies at a distance of about 224 Mpc, or 731 million light-years, for a Hubble constant $H_0 = 75$ km s^{-1} Mpc^{-1}, and the apparent radio luminosity and distance can be combined to infer an enormous absolute radio luminosity of about 10^{38} J s^{-1}. It is emitting as much power at radio wavelengths as the visible luminosity of a million million stars like the Sun, of absolute luminosity $L_\odot = 3.828 \times 10^{26}$ J s^{-1}, whose

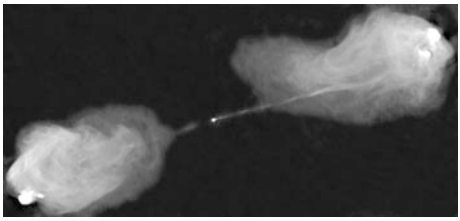


Fig. 15.6 Radio galaxy Cygnus A The radio galaxy Cygnus A, listed as 3C 405 in the third Cambridge catalogue of bright radio sources, which has a radio output 1 million times more powerful than the radio emission of a normal galaxy like the Milky Way. This radio image, taken with the Very Large Array at a wavelength of 6 cm with a field of view of 0.038 × 0.022 degrees, shows two narrow, straight radio-emitting jets of particles that protrude in opposite directions from a giant elliptical galaxy at the center. The redshift of the optically visible elliptical is $z = 0.056075$, indicating a distance of about 224 Mpc or 780 million light-years, and a linear extent for the radio galaxy of about 1 million light-years from end to end. The radio jets probably were ejected along the rotation axis of a super-massive black hole located within a central elliptical galaxy. It had to be active for tens of millions of years to produce the two radio lobes. (Courtesy of NRAO/AUI/NSF.)

most intense radiation is also at the visible, optical wavelengths with relatively dim radio emission.

When the optical image of the elliptical galaxy associated with Cygnus A is combined with maps of the radio signals, it is found that the radio emission is not confined to its visible counterpart; instead it is concentrated in two radio lobes that are separated from the central visible galaxy by hundreds of thousands of light-years. It is as if the radio-emitting clouds were expelled from the central elliptical galaxy, which is detectable only at optically visible wavelengths. The astonishing radio power is attributed to the nonthermal synchrotron radiation of high-speed electrons supplied from the visible center along two oppositely directed jets that feed the radio lobes (Fig. 15.6). These dual jets remain extraordinarily straight and surprisingly stable, energizing the radio lobes and pushing them farther and farther apart.

Harris and Krawcynski (2006) discussed x-ray emission from extragalactic jets; Bridle and Perley (1984) have reviewed extragalactic radio jets; Kellermann and Pauliny-Toth (1981) have reviewed our knowledge of compact radio sources, and Miley (1980) has discussed the structure of extended radio sources.

If the *radio galaxy* has been sending out radio power at the present rate at an estimated million-year lifetime, then it has emitted radio energy equivalent to the complete annihilation of about a 100,000 stars.

Example: Feeding the radio lobes of Cygnus A

The radio galaxy Cygnus A has a redshift of $z = 0.056$, which from the Hubble law $cz = H_0 D$ provides a distance, D, of 224 Mpc $= 6.91 \times 10^{24}$ m and about 730 million light-years for a Hubble constant of $H_0 =$ 75 km s^{-1} Mpc^{-1} where the speed of light is $c = 2.9979 \times 10^5$ km s^{-1} and

1 Mpc $= 3.0857 \times 10^{22}$ m. Its two lobes are separated by an angle of $\theta \approx 0.038$ degrees $\approx 137''$ and 6.63×10^{-4} radians, using the conversion factor of 1 radian $= 2.06265 \times 10^5$ s of arc. The minimum amount of time, t, that the central elliptical galaxy has been feeding the radio lobes is $t = \theta D/(2c)$, where the distance between the visible galaxy and one radio lobe is $\theta D/2$. That time is $t = 7.6 \times 10^{12}$ s $= 0.24 \times 10^6$ years, where 1 year $= 3.1557 \times 10^7$ s.

The absolute radio luminosity of Cygnus A is about $L_R \approx 1.3 \times 10^{38}$ J s^{-1} $\approx 3.4 \times 10^{11}$ $L_\odot$, where $L_\odot = 3.828 \times 10^{26}$ J s^{-1} is the visible absolute luminosity of the Sun. Thus, the radio power emitted by Cygnus A is comparable to the visible-light power of 340 billion stars like the Sun, exceeding that of our Milky Way Galaxy of about 100 billion stars. Over its lifetime of a quarter of a million years, Cygnus A emits about 10^{51} J of radio energy, which is comparable to the rest mass energy of 5,594 solar masses, or 5,594 $M_\odot c^2$ for a solar mass $M_\odot = 1.989 \times 10^{30}$ kg and the speed of light $c = 2.9979 \times 10^8$ m s^{-1}. Finding a radiation mechanism comparable to the complete annihilation of 5,594 stars was initially problematic, until super-massive black holes provided the answer.

Even more dramatic sources of energy were found still deeper in space and generated a longer time ago. As with radio galaxies, the discovery of the first quasar resulted from the accurate location of a bright radio source, which was determined when the Moon happened to pass in front of it. As the radio astronomer Cyril Hazard (1928–), then at the University of Sydney in Australia, realized, a careful timing of the disappearance and reappearance of the occulted source would establish a precise position, since the location of the Moon's edge is known accurately for any time.

In 1962 Hazard and his colleagues used the occultation method to show that 3C 273 is a double radio source, one component of which apparently coincided with a blue stellar object. This coincidence prompted Maarten Schmidt (1929–) to obtain an optical spectrum of the blue object using the 5 m (200 inch) Palomar telescope, which indicated an exceptionally high recession velocity of 0.16 % of the speed of light. When he told his colleague Jesse Greenstein (1909–2002) about the discovery, Greenstein produced a list of emission line wavelengths for the optically visible counterpart of another radio source 3C 48, and within minutes they had found that it is rushing away with an even faster motion at 37 % of the speed of light.

When these velocities are used to infer distances using the Hubble law, it is found that 3C 48 and 3C 273 are located at distances of billions of light-years. And when their observed luminosities are combined with these distances, it was found that they are shining with the visible blue light of 10 million million, or 10^{13}, Sun-like stars.

The sequence leading to the discovery of these hitherto unknown objects of tremendous velocity, distance, and luminosity happened so quickly that the article reporting the major discoveries appeared together in a six-page sequence in the journal *Nature* (Hazard et al. 1963; Schmidt 1963; Greenstein and Matthews 1963).

Because the bright objects appeared star-like in visible light, they became known as *quasistellar radio sources*, a term that soon was shortened to *quasars*. The quasars had, in fact, been ignored as stars on optical photographs for years. Once quasars were known, astronomers located others by obtaining optical spectra of bright, blue-colored, star-like objects that are located well outside the plane of the Milky Way, where stars are not supposed to be, and measuring the large redshifts characteristic of remote quasars. Thousands of quasars have now been discovered in this way, some of them emitting intense radio signals and many more silent ones with their radios turned off.

Astronomers gradually came to realize that quasars are brilliant, tiny cores, sometimes smaller than the solar system, embedded in much larger, extremely active galaxies, whose outer parts are difficult to detect in the intense quasar glare. From its vantage point in space, the *Hubble Space Telescope* resolved the core quasar light and removed it from the computerized images to detect the faint, fuzzy halo of a host galaxy that is as large as the elliptical galaxies found at the centers of many intense radio sources.

Quasars are believed to be very luminous versions of the same blue nuclei that Seyfert observed in the center of nearby spiral galaxies. The visible-light emission of quasars exhibits the same emission lines as both the Seyfert galaxies and the central elliptical galaxies of radio galaxies (Lynden-Bell 1969). Lang et al. (1975) present a composite Hubble diagram that includes normal galaxies, radio galaxies and quasi-stellar objects in the context of the evolution of the universe.

Seyfert galaxies, radio galaxies, and quasars all belong to a common class, known collectively as *active galactic nuclei*. Modern astronomers are now investigating active galactic nuclei using the *Hubble Space Telescope*, the *Spitzer Space Telescope* and large ground based telescopes operating at visible, infrared, millimeter and radio wavelengths. Fabian (2012) has discussed observational evidence for active galactic nuclei feedback. Ho (2008) reviewed nuclear activity in nearby galaxies; Crenshaw et al. (2003) have reviewed evidence for mass loss from the nuclei of active galaxies; Osterbrock (1991) provided a review of active galactic nuclei; Sulentic et al. (2000) have described broad emission lines in active galactic nuclei; Osterbrock and Mathews (1986) reviewed emission-line regions of active galaxies and QSOs; Weymann et al. (1981) have reviewed absorption lines in the spectra of quasi-stellar objects; and Ulrich et al. (1997) have discussed the variability of active galactic nuclei.

The active galactic nuclei radiate so powerfully over the entire range of the electromagnetic spectrum that they cannot possibly consist of ordinary stars, which emit most of their luminous output in a narrow band of wavelengths grouped around visible light. However, super-massive black holes can account for the prodigious energy output, violent activity, and rapid variations of active galactic

nuclei, as well as jets of material that moves out of them at extremely high relativistic speeds that approach the speed of light.

15.5.2 Super-Massive Black Holes

As independently proposed by the astrophysicists Edwin E. Salpeter (1924–), at Cornell University, and Yakov B. Zeldovich (1914–1987) in Moscow, the tremendous luminosity of every radio galaxy and quasar most likely is supplied by a *super-massive black hole*, which emits luminous radiation as its powerful gravity pulls in surrounding stars and gas (Salpeter 1964; Zeldovich 1964). The gravitational pull of a mass equivalent to 100 million Suns is needed to balance the visible quasar luminosity; otherwise its radiation pressure would blow away the quasar. Such a super-massive black hole would be sufficiently small and powerful enough to explain the tiny sizes and the colossal brightness of quasars. The super-massive black hole's rotational energy is used to accelerate charged particles and spew them out in diametrically opposite directions along its rotation axis at about the speed of light, continuously feeding the two radio lobes commonly found symmetrically placed from the center of radio galaxies and quasars.

Rees (1984) has reviewed black hole models for active galactic nuclei, and Begelman et al. (1984) and Lang (1999) have reviewed the theory of extragalactic radio sources. Longair (2011) provides a detailed treatment of the high-energy astrophysics used to describe cosmic radio and x-ray emission.

Kormendy and Richstone (1995) have reviewed the search for super-massive black holes in galactic nuclei. The classic example is M 87, a giant elliptical galaxy whose central spinning disk of hot gas indicates that a super-massive black hole resides at its center. M 87 is close enough to measure the motions of stars, and their increasing velocities toward the center indicate that billions of solar masses must be crammed within a very small, unseen volume to keep the high-velocity stars from flying into space (Sargent et al. 1978; Macchetto et al. 1997; Gebhardt and Thomas 2009).

A one-sided jet of gas emerges from the center of M 87 and stretches out into one of the two lobes of the radio galaxy Virgo A – numbered 3C 274 in the Cambridge survey (Fig. 15.7). The motions of bright concentrations in the jet indicate that they are traveling outward at about half the speed of light. And Very Long Baseline Interferometry observations with widely separated radio telescopes reveal that the M 87 jet emerges from a region at most 6 light-years across, most likely harboring the super-massive black hole that produces the jet.

Monstrous, super-massive black holes seem to inhabit the centers of all galaxies. They are massive, scaled-up versions of stellar black holes, with millions if not billions of times the mass of the Sun packed into a region only a few light-years across. Like their stellar counterparts, the super-massive black holes cannot be observed directly. Their presence is inferred from the orbital motion of nearby

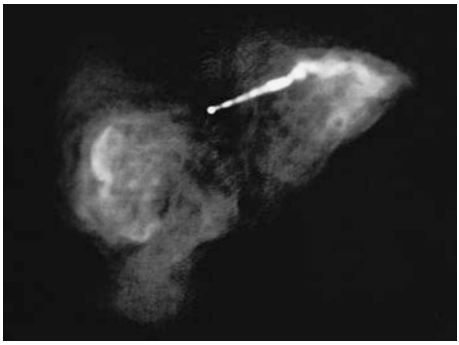


Fig. 15.7 Radio jet from M 87 The bright radio source Virgo A, also designated 3C 274, coincides with M 87, a giant elliptical galaxy with a redshift of $z = 0.00436$ and a distance of about 17.4 Mpc or 56.8 million light-years. M 87 is the largest and brightest galaxy within the Virgo cluster of galaxies. The core of M 87 contains a super-massive black hole of about 6.3 billion solar masses, or $6.3 \times 10^9\, M_\odot$, which is 1,500 times more massive than the black hole at the center of our Milky Way Galaxy. This radio map, made with the Very Large Array, shows two elongated lobes, one on either side of the center of M 87, apparently fed by the super-massive black hole. The most intense radio emission comes from a jet that emerges from the core of the galaxy and extends about 5,000 light-years into one of the two lobes. The observed high-speed motion of bright knots in the jet implies that its radio-emitting electrons are traveling at nearly the speed of light. (Courtesy of NRAO/AUI/NSF.)

visible stars the trajectories of which are guided by the otherwise invisible black holes.

The faster the stars are moving, the more gravity – and therefore mass – is needed to hold the stars in their orbits. By measuring the sharp rise in orbital velocity at close distances from galaxy centers, astronomers have weighed unseen super-massive black holes in nearby giant elliptical galaxies, which are the brightest galaxies in clusters of galaxies. The central black-hole powerhouse in relatively nearby galaxies, designated M 87, NGC 3842, and NGC 4889, tips the scales at 6.3, 9.7 and 21 billion solar masses, respectively. Without a gravitational pull equivalent to about 10 billion Sun-like stars, the close, fast-moving stars would fly away from the galaxies (McConnell et al. 2011). Such central, super-massive black holes most likely reside in more distant galaxies that are too far away to resolve central stars and measure their motions.

Quasars and active galactic nuclei become increasingly numerous as we look deeper into space, at larger redshifts. The number density of unobscured quasars peaks at a redshift between $z = 2$ and $z = 3$, which indicates that the hot, luminous spurt of activity happened in the distant past, about 10 billion years ago and shortly after the first galaxies were born. At smaller redshifts and closer distances, corresponding to an old age, there are relatively few quasars. The spurt of activity apparently became worn out and used up as the galaxies grew older.

To power the youthful activity of a quasar, there has to be about 1 solar mass per year of gas flowing into the super-massive black hole. Therefore, billions of

stars or the equivalent amount of gas must be consumed as its active nucleus evolves during the course of billions of years. The supply dwindles away over time and the activity dies down, but the black hole does not disappear.

Most galaxies probably contain super-massive black holes at their center. Those in the older, nearby galaxies are the starving remains of former quasars, with a dwindling supply of material that once fed a higher rate of activity. They are found in ordinary nearby galaxies, such as Andromeda, whose cores are surviving fossils of former quasars. Our Galaxy, the Milky Way, is almost as old as the observable universe, and it contains a central super-massive black hole. However, its mass is equivalent to only about 1 million stars like the Sun rather than the billions in some super-massive black holes (see Sect. 14.1).

15.5.3 Gamma-Ray Bursts

The brightest sources found in the universe, at least so far, are the gamma-ray bursts whose duration is measured in seconds or less and which never reappear in exactly the same part of the sky. They emit energy at a gamma-ray wavelength shorter than 10^{-11} m, so each photon of a gamma ray is about 100,000 times more energetic than a photon of visible light. When it was found that they originate in remote galaxies, it was realized that the observed gamma ray bursts might radiate, for a few seconds, gamma-ray energy far in excess of the visible-light energy emitted by galaxies. These gamma-ray bursts can briefly become the brightest electromagnetic events in the universe.

The initial burst is usually followed by a longer-lived "afterglow" emitted at longer wavelengths, from x-rays to optical and radio of galaxies. Gamma-ray bursts are attributed to intense radiation emitted during a supernova explosion when a rapidly rotating, high mass star collapses to form a neutron star or black hole.

The discovery of the gamma ray bursts was the unexpected result of defense satellite observations designed to detect clandestine nuclear bomb explosions in the Earth's atmosphere, on the Moon, or in outer space, but brief, intense gamma-ray bursts were instead found to be coming from the distant Cosmos.

The secret *Vela* satellites were launched and operated in identical pairs on opposite sides of a circular orbit around the Sun, eventually with sufficient time resolution to determine the direction of the source from the difference in arrival time of its radiation at two or more satellites. Although a nuclear bomb was never detected, unexpected flashes of gamma rays were discovered coming from different places in deep space, with an existence measured in seconds. Moreover, the lag between the burst arrival times at the two defense satellites indicated that they originated far beyond the solar system, and were therefore of cosmic origin.

The gamma-ray bursts were kept secret for about five years, until the Los Alamos scientists, Ray W. Klebesadel (1932–), Ian B. Strong (1930–) and Roy A. Olson (1924–), described the discovery in the *Astrophysical Journal*

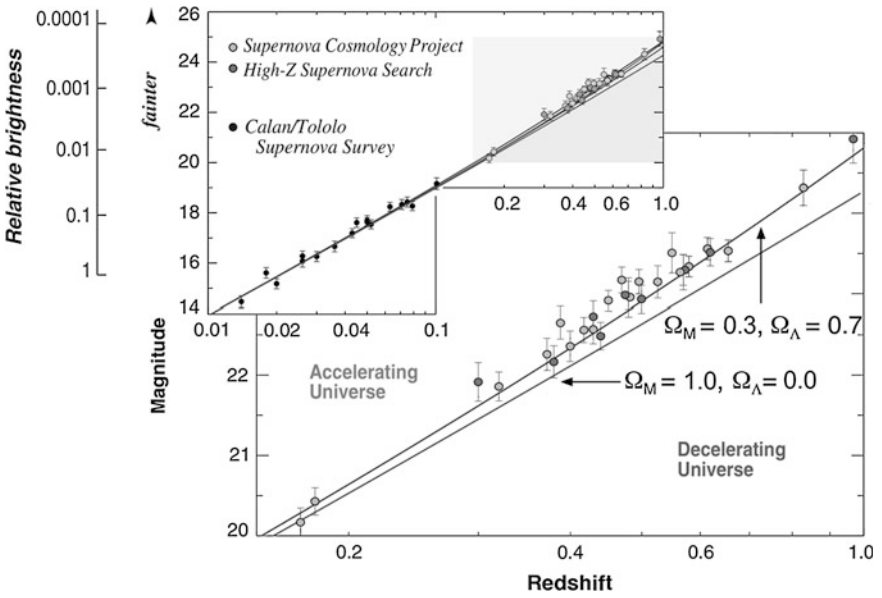


Fig. 15.8 The accelerating expansion of the universe The Hubble diagram plot of the apparent magnitude of Type Ia supernovae plotted as a function of their redshift. At a redshift below about $z = 0.1$, there is a linear fit to the data, but at larger redshifts, the observations begin to diverge from a straight line. The curved departures for distant supernovae at high redshift indicate an acceleration in which the speed of expansion is increasing. The observed data can be compared to cosmological models with different values of the omega parameter, Ω. It is the ratio of the inferred density to the critical mass density needed to stop the expansion of the universe in the future. The subscript Λ denotes the cosmological constant, a possible form of dark energy, and the subscript m denotes matter. (Adapted from Saul Perlmutter, *Physics Today* April 2003. From "The Life and Death of Stars" by Kenneth R. Lang, published by Cambridge University Press, 2013. Reprinted with permission.)

(Klebesadel et al. 1973). Civilian astronomers might blame the military for keeping such an important discovery hidden for years, but it is also likely that the gamma-ray bursts would have never been discovered if it wasn't for the defense satellites. No other government agency was likely to fund a speculative program designed to look for such totally unknown, unsuspected, and unprecedented events in outer space.

The long lasting afterglow of the cosmic gamma-ray bursts was eventually detected at visible wavelengths, enabling the distances to be determined by spectroscopy and the Hubble law, and the enormous energies realized (Van Paradijs et al. 1997; Frail et al. 1997; Metzger et al. 1997; Kulkarni et al. 1998). If these bursts emit radiation in all directions, then the total gamma ray luminosity of the burst can briefly exceed the combined visible-light luminosity of millions of galaxies. A lower gamma ray luminosity is inferred if the bursts are beamed, like a pulsar's radio emission. At least some of the gamma ray bursts are attributed to

powerful extragalactic supernovae explosions of very massive stars whose centers collapse into black holes.

Gehrels et al. (2009) have reviewed gamma-ray bursts in the *Swift* era; Fishman and Meegan (1995) discussed earlier observations of gamma ray bursts. Woosley and Bloom (2006) have presented the supernova-gamma-ray burst connection, and Mészáros (2002) has provided a review of theories of gamma-ray bursts. Paradijs et al. (2000) have reviewed gamma-ray burst afterglows; and Weiler, Panagia, and Sramek (2002) have summarized our knowledge of radio emission from supernovae and gamma-ray bursts. Fender and Belloni (2004) have discussed GRS 1915 + 105 and the disc-jet coupling in accreting black hole systems.

15.6 Dark Energy, the Cosmological Constant, and How it All Ends

15.6.1 Discovery of Dark Energy

The redshift, z, of a galaxy increases with its distance, D, according to the Hubble law $cz = H_0 D$, where c is the speed of light and H_0 is the Hubble constant. Since distances can only be independently measured for nearby galaxies, the law is expressed as a redshift – magnitude relation applicable at larger redshifts. This relation is given by

$$m = 5 \log(cz/H_0) + M + 25, \tag{15.25}$$

where the apparent magnitude is denoted by m and the absolute magnitude by M. If all galaxies have the same absolute magnitude, then a plot of m against $\log z$ will describe a straight line whose slope is related to the Hubble constant.

At redshifts greater than one, observable departures from linearity in the apparent magnitude – log redshift diagram could occur due to changes in the expansion velocity. If the expansion of the universe is slowing down and decelerating, for example, then the apparent brightness would be greater than expected from uniform expansion, and the apparent magnitude less than expected since greater brightness means a smaller magnitude.

Supernova explosions of Type Ia, denoted by SNe Ia, provide a useful method for measuring such possible effects (Reiss et al. 1995, 1996; Branch, 1988). Since they are the result of a thermonuclear explosion of a white dwarf star that has grown above the Chandrasekhar limit (Sect. 13.4), there is little variation in their absolute magnitude of $M_B = 19.6 \pm 0.2$ in blue light,and they are so luminous that they can be detected at relatively large redshifts of $z = 0.5$–1.5.

In the late 1980s, a group led by Saul Perlmutter (1959–) at the Lawrence Berkeley National Laboratory began a dedicated search for Type Ia supernovae to measure how fast the expansion of the universe was slowing down, due to the gravitational pull of its combined matter. By 1998, his *Supernova Cosmology*

Project, and a rival group, dubbed the *High-z Supernova Search*, announced that the light from the distant supernovae was fainter than predicted, which meant that the galaxies are speeding up, expanding at a quickening pace, and accelerating instead of slowing down by gravity (Perlmutter et al. 1998, 1999; Riess et al. 1988; Schmidt et al. 1988). In other words, the distant galaxies were not where they were supposed to be, and the space they are in seems to be expanding at a faster rate as time goes on.

The redshift – magnitude diagram of SNe Ia has a straight-line, linear shape at low redshifts, as expected from the Hubble law, but unexpected nonlinear effects appear at large redshifts (Fig. 15.8). The line describing the data indicates that the galaxies are expanding at a quickening pace.

Saul Perlmutter, Brian P. Schmidt (1967–), and Adam G. Riess (1969–) were awarded the 2011 Nobel Prize in Physics for the discovery of the accelerating expansion of the universe through observations of distant supernovae.

Distant galaxies are being accelerated by the anti-gravity push of a mysterious *dark energy*. So the fate of the universe is no longer supposed to depend on its mass, but rather on its energy. If dark energy retains its vigor, the universe will not stop expanding.

Dark energy, which pushes matter apart, is not the same as dark matter, which encourages attraction. But the discovery of dark energy did do away with the need for overwhelming amounts of dark matter to keep the universe poised on the edge of future collapse, but never quite pulling it there. No more than one quarter of the critical mass density is now imagined to reside in mass of any kind, and astronomer's observations of such a low-density universe are now widely accepted. Dark energy has taken over; perhaps keeping the universe at the brink of closure within "flat" space, which is described by Euclidean geometry.

Frieman et al. (2008) have reviewed dark energy and the acceleration of the expansion of the universe. Ratra and Vogeley (2008) have reviewed the current standard model for evolution of the universe, including big-bang cosmology, inflation, dark matter and dark energy, as well as the formation and observations of galaxies and stars.

15.6.2 Using the Cosmological Constant to Describe Dark Energy

The trouble is, nobody understands this mysterious something, this dark energy that permeates space and eventually overwhelms the gravitational self-attraction of the entire material universe. But an old idea, termed the cosmological constant, has been revived to give dark energy another name and couch it in mathematical terms. That's the anti-gravity fudge factor that Einstein introduced to stabilize a non-moving universe against collapse.

When the expanding universe had not yet been discovered, it looked as if the unrelenting, universal attraction of gravity would cause the eventual collapse of an unmoving universe, so Einstein (1917a, b) inserted a mathematical fix, the cosmological constant, into his relativity equations. The extra term represented the repulsive force of an unknown and undetected form of energy that permeated space and exerted a sort of outward pressure that opposed gravity and kept the universe from collapsing.

In a little more than a decade, it was discovered that the galaxies are moving away from us in a cosmic expansion, so the universe wasn't static, or non-moving, after all. Einstein therefore abandoned the cosmological constant, and stated that the *ad hoc* term was greatly detrimental to the formal beauty of his theory.

But the artifice stubbornly refused to die, and has been repeatedly invoked whenever cosmologists have had trouble reconciling their theories with observations. All they had to do was revive the term, stick it in the relevant equations, and adjust its value. As an example, Georges Lemaître (1894–1966) constructed a model universe with two periods of accelerated expansion, one at the beginning and one later, and a more gentle coasting period in between (Fig. 15.9), proclaiming that the expansion thus took place in three phases: a first period of rapid expansion a second period of slowing-up, followed by a third period of accelerated expansion (Lemaître 1931a, b). It is doubtless in the third period we find ourselves today. This interpretation is somewhat similar to some modern explanations of dark energy, beginning with rapid inflation and with a currently accelerated expansion that might invoke the cosmological constant.

Peebles and Ratra (2003) have provided an extensive review of the cosmological constant and dark energy, whereas Carroll et al. (1992) provided a review of the cosmological constant in another context.

The equations that describe a homogeneous, isotropic expanding universe, given in Chap. 14, Sect. 14.6 for zero cosmological constant $\Lambda = 0$, have to be rewritten to allow for a non-zero value of this constant. The Hubble expansion parameter $H(t)$ is then given by:

$$H^2(\mathrm{t}) = \left(\frac{\dot{R}(t)}{R(t)}\right)^2 = \left(\frac{dR(t)/dt}{R(t)}\right)^2 = \frac{8\pi}{3}G\rho(t) - \frac{kc^2}{R^2(t)} + \frac{\Lambda}{3}, \tag{15.26}$$

where $R(t)$ is the scale factor of the universe, $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} is the gravitational constant, $\rho(t)$ is the mass-energy density, which is equal to the radiation energy density $\rho_r(t)$ in the early radiation-dominated era and the matter density $\rho_m(t)$ in the current matter-dominated era, the space curvature constant is $k = -1$, 0 or +1, and the cosmological constant is denoted by Λ.

Contemporary observations are consistent with $k = 0$, and in this situation, we have:

$$\frac{H^2(t)}{H^2(t_0)} = \frac{8\pi G\rho(t)}{3H_0^2} + \frac{\Lambda}{3H_0^2} = \frac{\rho(t)}{\rho_c} + \frac{\Lambda}{3H_0^2}, \tag{15.27}$$

With

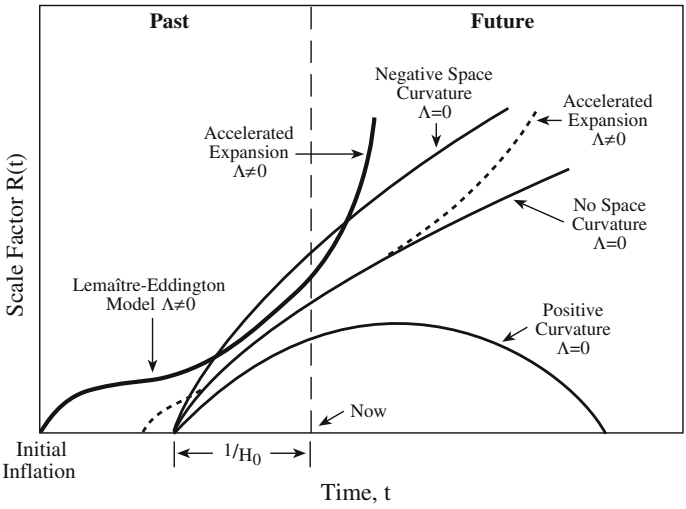


Fig. 15.9 Models of the expanding universe Schematic representation of various cosmological models showing the size, or scale factor R(t), of the expanding universe as a function of time, t. The approximate age since the expansion began is given by $1/H_0$ where H_0 is the Hubble constant. Since the expansion age was once thought to be smaller than the age of the oldest rocks on Earth, Georges Lemaître (1894–1966) and Arthur Eddington (1822–1944) independently used a cosmological repulsion term with Einstein's *General Theory of Relativity* to permit an adjustable age for the universe. The adjustable term is symbolized by a non-zero value for the cosmological constant Λ. In this interpretation (*thick solid line*), the universe began with expansion against gravity, followed by an essentially non-moving stagnation in which gravitation and cosmological repulsion were nearly in balance. This coasting period was then followed by an accelerated expansion driven by the cosmological repulsion. Three models with no cosmological constant, or with $\Lambda = 0$, describe three future possibilities for the universe with no dark energy (*thin solid lines*). It can become closed with positive space curvature, forever open with negative space curvature, and always open and no curvature of space. In recent times, this last option, of never ending expansion in space without curvature, has been modified by a non-zero cosmological constant to give a boost to some age estimates and permit an accelerated expansion by a mysterious dark energy (*dashed lines*)

$$\rho_C = \frac{3H_0^2}{8\pi G} = 1.879 \times 10^{-26}\, h^2\, \mathrm{kg\,m^{-3}} \approx 1.0 \times 10^{-26}\, \mathrm{kg\,m^{-3}}, \tag{15.28}$$

with the Hubble constant $H_0 = 100\, h$ km s^{-1} Mpc^{-1}, $h \approx 0.75$, and 1 Mpc $= 3.0857 \times 10^{19}$ km.

Once the cosmological term has grown large enough, it dominates the right side of the expansion-parameter equation and we have

$$\dot{R}(t) \approx \left(\frac{\Lambda}{3}\right)^{\frac{1}{2}} R(t), \tag{15.29}$$

which has the solution

$$R(t) \propto \exp\left[\left(\frac{\Lambda}{3}\right)^{\frac{1}{2}} t\right] = \exp(H_C t), \tag{15.30}$$

where the Hubble expansion parameter $H(t)$ has become a constant H_C, and the universe enters an exponential expansion.

The other Friedmann equation that involves the pressure P becomes:

$$\frac{\ddot{R}(t)}{R(t)} = \frac{d^2R(t)/dt}{R(t)} = -\frac{4\pi G}{3c^2}\left(\rho\,(t)c^2 + 3P\right) + \frac{\Lambda}{3}. \tag{15.31}$$

This equation implies that with large positive cosmological constant Λ the term $d^2R(t)/dt$ becomes positive and the expansion of the universe accelerates, as opposed to deceleration that could occur without such a term.

The present value of the deceleration parameter q_0 for any value of Λ is given by

$$q_0 = \frac{\ddot{R}(t_0)}{R(t_0)H^2_{0''/>}} = \frac{\Omega_m(t_0)}{2} - \frac{\Lambda c^2}{3H_0^2}, \tag{15.32}$$

where the pressure is now negligible and $H_0 = H(t_0) = 100\ h$ km s^{-1} Mpc^{-1} is the Hubble constant and $h \approx 0.75$.

The power of the cosmological constant is measured from its density parameter Ω_Λ given by

$$\Omega_\Lambda = \Omega_\Lambda(t_0) = \frac{\Lambda c^2}{3H_0^2}, \tag{15.33}$$

while the matter density parameter is given in its usual way

$$\Omega_m = \Omega_m(t_0) = \frac{\rho_m(t_0)}{\rho_c}. \tag{15.34}$$

More than a decade of observations of type SNe Ia has confirmed that their redshift – magnitude relation goes non-linear at large redshifts, permitting astronomers to measure parameters that describe the unexpected effect. The initial results, obtained in 1998, were interpreted by a total matter density parameter $\Omega_m = 0.28 \pm 0.085$, for invisible and visible matter assuming a cosmological constant of $\Lambda = 0$ (Perlmutter et al. 1998), but it was soon realized that the universe was expanding, propelled by a dark energy that could be interpreted in terms of a non-zero cosmological constant (Riess et al. 1998). Ten years later, data for hundreds of SNe Ia were used to suggest a cosmological-constant density factor of $\Omega_\Lambda = 0.713 \pm 0.027$ for a flat universe with a space curvature constant of $k = 0$ (Kowalski et al. 2008). When combined with other data, tight limits of $\Omega_m = 0.237 \pm 0.010$ were also realized (Fig. 15.10) and the previous Table 15.2 of *WMAP* results).

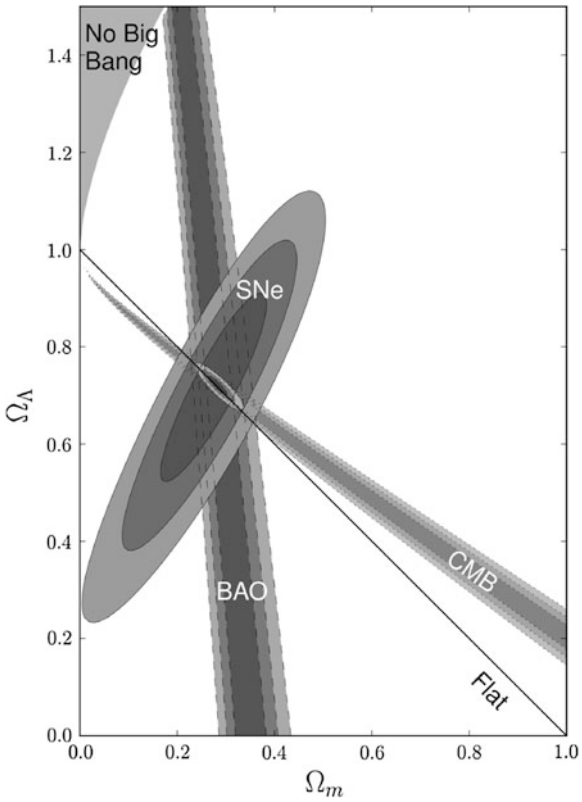


Fig. 15.10 Dark mass and dark energy constraints Three independent sets of observations provide constraints to the mass and energy content of the universe. Studies of high-redshift Type Ia supernovae, designated SNe, constrain the difference between the density of matter and the density of dark energy in the universe. This diagram illustrates the results from the Union 2.1 compilation of 833 SNe drawn from 19 datasets from 2008 to 2011. Anisotropies in the cosmic microwave background radiation, denoted CMB, constrain their sum. Investigations of Baryon Acoustic Oscillations, abbreviated BAO, detected in data from the Sloan Digital Sky Survey provide other constraints. Here the densities are given in terms of the omega parameter, denoted by Ω. It is the ratio of the inferred density to the critical mass density needed to stop the expansion in the future. The subscript Λ denotes the cosmological constant, a possible form of dark energy, while the subscript m denotes matter. Galaxy observations indicate a mass density of at most $\Omega_m = 0.3$. The theoretical expectation of an inflationary universe without spatial curvature requires $\Omega_m + \Omega_\Lambda = 1.0$, and significant dark energy with $\Omega_\Lambda = 0.7$. The observations are consistent with such a flat universe, described by ordinary Euclidean geometry. [Courtesy of the Supernova Cosmology Project, whose compilation of 580 SNe is available at http://supernova/lbl.gov/ - also see N. Suzuki et al, "The *Hubble Space Telescope* cluster supernova survey: V. Improving the dark energy constraints above z > 1 and building an early-type-hosted supernova sample", *Astrophysical Journal* **746**, 85 (2012).]

As also previously mentioned, recent *Planck* mission results indicate $\Omega_m = 0.317$ and $\Omega_\Lambda = 0.683$ also with a flat universe of $\Omega_m + \Omega_\Lambda = 1.000$.

The SNe Ia redshift–magnitude diagram, the *WMAP* fluctuations in the cosmic microwave background radiation, and other astronomical observations, have been combined to show that $\Omega_\Lambda = 0.728 \pm 0.015$, and $\Omega_m h^2 = 0.1334 \pm 0.0156$ with $h \approx 0.75$, so $\Omega_m + \Omega_\Lambda \approx 1$ with a space curvature constant $k = 0$ and non-curved Euclidean space. In this condition, at any time, t,

$$\Omega_m + \Omega_\Lambda = \Omega_m(t) + \Omega_\Lambda(t) = 1. \tag{15.35}$$

Allen et al. (2011) have reviewed the determination of cosmological parameters from observations of galaxy clusters. Leibundgut (2001) has reviewed the cosmological implications of observations of Type Ia supernovae; and Branch (1998) has reviewed Type Ia supernovae and the Hubble constant .

15.6.3 When Stars Cease to Shine

Ever since the discovery of the expansion of the universe, we have known that the universe is slowly and inexorably approaching an end. There has never been any known force that can prevent the observable universe from steadily moving into darkness. As Georges Lemaître (1894–1966) so eloquently stated "The evolution of the world can be compared to a display of fireworks that has just ended: some few red wisps, ashes, and smoke. Standing on a well-chilled cinder, we see the slow fading of the suns, and we try to recall the vanished brilliance of the origin of the worlds (Lemaître 1931a, 1931, 1950)."

Dark energy has been found, but that discovery may not help matters. It gives an extra outward push to the expansion and further reduces the power of mass to stop the expanding universe in the future. If the acceleration caused by dark energy continues unabated at the current rate, all the galaxies will be moving apart so quickly that they cannot communicate with one another in about 150 billion years, disappearing over the cosmic horizon.

Eventually, in about 100 trillion years, all of the interstellar gas and dust from which new stars condense finally will be used up, and new stars will cease to form in any galaxy. But if dark energy weakens as time goes on, expending its strength, then gravity and mass may take over and eventually pull back on the outward moving galaxies, ultimately reversing the expansion and dragging the universe back, melting it down and remaking the big bang. So no one knows for sure just how it will all end.

Chapter 16
References

A'Hearn, M. F. (2011): Comets as building blocks. *Annual Review of Astronomy and Astrophysics* **49**, 281-299.

Aannestad, P. A., Purcell, E. M.: Interstellar grains. *Annual Review of Astrononomy and Astrophysics* **11**, 309 (1973).

Abell, G. O. (1965): Clustering of galaxies. *Annual Review of Astronomy and Astrophysics* **3**, 1.

Abney, W. de W. (1877): Effect of a star's rotation on its spectrum. *Monthly Notices of the Royal Astronomical Society* **37**, 278.

Adams, F. C. (2010): The birth environment of the solar system. *Annual Review of Astronomy and Astrophysics* **48**, 47-85.

Adams, W. S. (1914): An A-type star of very low luminosity. *Publications of the Astronomical Society of the Pacific* **26**, 198. Reproduced in Lang and Gingerich (1979).

Adams, W. S. (1915): The spectrum of the companion of Sirius. *Publications of the Astronomical Society of the Pacific* **27**, 236-237. Reproduced in Lang and Gingerich (1979).

Adams, W. S. (1925): The relativity displacement of the spectral lines in the companion of Sirius. *Proceedings of the National Academy of Sciences* **11**, 382.

Adams, W. S., Kohlschütter, A. (1914): Some spectral criteria for the determination of absolute stellar magnitudes. *The Astrophysical Journal* **40,** 385-398. Reproduced in Lang and Gingerich (1979).

Aggarwal, H. R., Oberbeck, V. R. (1974): Roche limit of a solid body. *The Astrophysical Journal* **191**, 577.

Ahmed, S. N., et al. (2004): Measurement of the total active ^{8}B solar neutrino flux at the Sudbury Neutrino Observatory with enhanced neutral current sensitivity. *Physical Review Letters* **92**, 181301.

Alder, K. (2002): *The Measure of All Things: The Seven-Year Odyssey and Hidden Error that Transformed the World.* New York: Free Press 2002.

Alfvén, H. (1942a): On the existence of electromagnetic-hydrodynamic waves. *Arkiv. f. Mat., Astron., Physik* **29**, 1.

Alfvén, H. (1942b): The existence of electromagnetic-hydrodynamic waves. *Nature* **150**, 405.

Alfvén, H. (1947): Granulation, magneto-hydrodynamic waves, and the heating of the solar corona. *Monthly Notices of the Royal Astronomical Society* **107**, 211-219.

Allen, S. W., Evrard, A. E., Mantz, A. B. (2011): Cosmological parameters from observations of galaxy clusters. *Annual Review of Astronomy and Astrophysics* **49**, 409-470.

Alpher, R. A., Bethe, H., Gamow, G. (1948): The origin of chemical elements. *Physical Review* **73**, 803-804. Reproduced in Lang and Gingerich (1979).

Alpher, R. A., Follin, J. W., Herman, R. C. (1953): Physical conditions in the initial stages of the expanding universe. *Physical Review* **92**, 1347.

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7_16, © Springer-Verlag Berlin Heidelberg 2013

Alpher, R. A., Herman, R. C. (1948): Evolution of the universe. *Nature* **162**, 774. Reproduced in Lang and Gingerich (1979).
Alpher, R. A., Herman, R. C. (1950): Theory of the origin and relative abundance distribution of the elements. *Review of Modern Physics* **22**, 153-212.
Ambartsumian, V. A. (1949): Expanding stellar associations. *Astronomicheskii zhurnal* **26**, 1-9. English translation in Lang and Gingerich (1979).
Ambartsumian, V. A. (1955): Stellar systems of positive energy. *Observatory* **75**, 72-78.
Ambartsumian, V. A. (1958): On the evolution of galaxies. In: *La structure et l'evolution de l'univers. Institut international de physique solvay* (ed. R. Stoops). Brussels: Coudenberg 1958. Reproduced in Lang and Gingerich (1979).
Anderson, C. D. (1932a): The apparent existence of easily deflectable positives. *Science* **76**, 238-239.
Anderson, C. D. (1932b): Energies of cosmic-ray particles. *Physical Review* **41**, 405-421.
Anderson, C. D. (1933): The positive electron. *Physical Review* **43**, 491-494.
Anderson, C. D., Neddermeyer, S. H. (1937): Cosmic ray particles of intermediate mass. *Physical Review* **54**, 88.
Anderson, L. S., Athay, R. G. (1989): Chromospheric and coronal heating. *Astrophysical Journal* **336**, 1089-1091.
Anderson, W. (1929): Über die Grenzdichte der Materie und der Energie (About the interface of matter and energy). *Zeitshrift fur Physik* **56**, 851-856.
Andrade, E. N. da C. (1959): Doppler and the Doppler effect. *Endeavor* ***18***, No. 69, January 1959.
Angel, J. R. P. (1978): Magnetic white dwarfs. *Annual Review of Astronomy and Astrophysics* **16**, 487.
Ångström, A. J. (1868): *Recherches sur le Spectre Solaire*. Uppsala: W. Schultz 1868.
Appleton, E. V. (1932): Wireless studies of the ionosphere. *Proceedings of the Institute of Electrical Engineers (London)* **71**, 642-650.
Appleton, E. V., Barnett, M. A. F. (1925): Local reflection of wireless waves from the upper atmosphere. *Nature* **115**, 333-334.
Appleton, E. V., Barnett, M. A. F. (1925): On some direct evidence for downward atmospheric reflection of electric rays. *Proceedings of the Royal Society (London)* **A109**, 621-641.
Appleton, E. V., Hey, J. S. (1946): Solar radio noise. *Philosophical Magazine* **37**, 73-84.
Armitage, P. J. (2011): Dynamics of protoplanetary disks. *Annual Review of Astronomy and Astrophysics* **49**, 195-236.
Arnett, W. D. (1973): Explosive nucleosynthesis in stars. *Annual Review of Astronomy and Astrophysics* **11**, 73.
Arnett, W. D. (1995): Explosive nucleosynthesis revisited: Yields. *Annual Review of Astronomy and Astrophysics* **33**, 115.
Arnett, W. D., Bahcall, J. N., Kirshner, R. P., Woosley, S. E. (1989): Supernova 1987A. *Annual Review of Astronomy and Astrophysics* **27**, 629.
Arp, H. (1962): The globular cluster M 5. *The Astrophysical Journal* **135**, 311.
Arrhenius, S. A. (1896): On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine and Journal of Science* **41**, 237-268.
Aschwanden, M. J. (2006): *Physics of the Solar Corona: An Introduction*. New York: Springer-Verlag, Second Edition 2006.
Aschwanden, M. J., Poland, A., Rabin, D. M. (2001): The new solar corona. *Annual Review of Astronomy and Astrophysics* **39**, 175-210.
Ash, M. E., Shapiro, I. I., Smith, W. B. (1967): Astronomical constants and planetary ephemerides deduced from radar and optical observations. *The Astronomical Journal* **72**, 338.
Asplund, M. et al. (2009): The chemical composition of the Sun. *Annual Review of Astronomy and Astrophysics* **47**, 481-522
Aston, F. W. (1919): The constitution of the elements. *Nature* **104**, 393.
Aston, F. W. (1920): The mass-spectra of chemical elements. *Philosophical Magazine and Journal of Science* **39**, 611-625.
Aström, E. (1950): On waves in an ionized gas. *Arkiv. Physik* **2**, 443.

Athay, R. G., White, O. R. (1978): Chromospheric and coronal heating by sound waves. *The Astrophysical Journal* **226**, 1135-1139.
Athay, R. G., White, O. R. (1979): Chromospheric oscillations observed with *OSO 8* IV. Power and phase spectra for CIV. *The Astrophysical Journal* **229**, 1147-1162.
Atkinson, R. d'E. (1931): Atomic synthesis and stellar energy I, II. *The Astrophysical Journal* **73**, 250-295, 308-347. Reproduced in Lang and Gingerich (1979).
Atkinson, R. d'E. (1936): Atomic synthesis and stellar energy III. *The Astrophysical Journal* **84**, 73.
Atkinson, R. d'E., Houtermans, F. G. (1929): Zur Frage de Aufbaumöglichkeit der Elemente in Sternen. *Zeitshrift fur Physik* **54**, 656-665.
Audouze, J., Tinsley, B. M. (1976): Chemical evolution of galaxies. *Annual Review of Astronomy and Astrophysics* **14**, 43.
Aumann, H. H. (1985): *IRAS* observations of matter around nearby stars. *Publications of the Astronomical Society of the Pacific* **97**, 885-891.
Axford, W. I. (1985): The solar wind. *Solar Physics* **100**, 575-586.
Baade, W. (1942): The Crab Nebula. *The Astrophysical Journal* **96**, 188.
Baade, W. (1944): The resolution of Messier 32, NGC 205, and the central region of the Andromeda nebula. *The Astrophysical Journal* **100**, 137. Reproduced in Lang and Gingerich (1979).
Baade, W. (1945): B Cassiopeiae as a supernova of Type I. *Astrophysical Journal* **102**, 309-319.
Baade, W. (1952): A revision of the extra-galactic distance scale. *Transactions of the International Astronomical Union* **8**, 397. Reproduced in Lang and Gingerich (1979).
Baade, W. (1957): The polarization and distribution in the Crab nebula derived from plates taken with the 200-inch telescope. *Bulletin of the Astronomical Institutes of the Netherlands* **13**, 301-311.
Baade, W., Minkowski, R. (1954): Identification of the radio sources in Cassiopeia, Cygnus A, and Puppis A. *The Astrophysical Journal* **119**, 206-214. Reproduced in Lang and Gingerich (1979).
Baade, W., Swope, H. H. (1955): The Palomar survey of variables in M 31. *Astonomical Journal* **60**, 151-157.
Baade, W., Zwicky, F. (1934a): On super-novae. *Proceedings of the National Academy of Science* **20**, 254. Reproduced in Lang and Gingerich (1979).
Baade, W., Zwicky, F. (1934b): Cosmic rays from super-novae. *Proceedings of the National Academy of Sciences* **20** (5), 259-263.
Baade, W., Zwicky, F. (1934c): Remarks on super-novae and cosmic rays. *Physical Review* **46**, 76.
Baade, W., Zwicky, F. (1934d): Supernovae and cosmic rays. *Physical Review* **45**, 138.
Babcock, H. W. (1961): The topology of the sun's magnetic field and the 22-year cycle. *The Astrophysical Journal* **133**, 572-587.
Babcock, H. W., Babcock, H. D. (1955): The sun's magnetic field, 1952-1954. *The Astrophysical Journal **121**, 349-366.*
Backer, D. C., Hellings, R. W. (1986): Pulsar timing and general relativity. *Annual Review of Astronomy and Astrophysics* **24**, 537.
Bahcall, J. N. (1964): Solar neutrinos I. Theoretical. *Physical Review Letters* **12**, 300-302. Reproduced in Lang and Gingerich (1979).
Bahcall, J. N. (1965): Observational neutrino astronomy. *Science* **147**, 115.
Bahcall, J. N. (1978): Solar neutrino experiments. *Review of Modern Physics* **50**, 881-903.
Bahcall, J. N., Bethe, H. A. (1990): A solution to the solar neutrino problem. *Physical Review Letters* **65**, 2233-2235.
Bahcall, J. N., Gonzalez-Garcia, M. C., Pena-Garay, C. (2003): Does the Sun shine by pp or CNO fusion reactions? *Physical Review Letters* **90**, 131301-131305.
Bahcall, J. N., Pinsonneault, M. H. (2004): What do we (not) know theoretically about solar neutrino fluxes? *Physical Review Letters* **92**, 121301.

Bahcall, N. A. (1977): Clusters of galaxies. *Annual Review of Astronomy and Astrophysics* **15**, 505.

Bahcall, N. A. (1988): Large-scale structure in the universe indicated by galaxy clusters. *Annual Review of Astronomy and Astrophysics* **26**, 631.

Bahcall, N. A. (1993): Clusters, superclusters, and large-scale structure – a consistent picture. *Proceedings of the National Academy of Science* **90**, 4848.

Baker, J. G., Menzel, D. H. (1938): Physical processes in gaseous nebulae III. The Balmer decrement. *The Astrophysical Journal* **88**, 52. Reproduced in Menzel (1962).

Baker, N., Kippenhahn, R. (1962): The pulsation of models of δ Cephei stars. *Zeitschrift fur Astrophysik* **54**, 114.

Balick, B., Frank, A. (2002): Shapes and shaping of planetary nebulae. *Annual Review of Astronomy and Astrophysics* **40**, 439-486.

Balmer, J. J. (1885): Notiz über die Spectrallinien des Wasserstoffs (A note on the spectral lines of hydrogen). *Annalen der Physik und Chemie* **25**, 80-85.

Barnard, E. E. (1916): A small star with large proper motion. *Astronomical Journal* **29** (695), 181.

Barnard, E. E. (1919): On the dark markings of the sky with a catalogue of 182 such objects. *The Astrophysical Journal* **49**, 1.

Barnard, E. E. (1927): *A Photographic Atlas of Regions of the Milky Way*. Washington, D.C.: Carnegie Institution.

Barnes, J. E., Hernquist, L. (1992): Dynamics of interacting galaxies. *Annual Review of Astronomy and Astrophysics* **30**, 705.

Barstow, M. A, et al. (2005): *Hubble Space Telescope* spectroscopy of the Balmer lines in Sirius B. *Monthly Notices of the Royal Astronomical Society* **362**, 1134-1142.

Bastian, N., Covey, K. R., Meyer, M. R. (2010): A universal stellar initial mass function? A critical look at variations. *Annual Review of Astronomy and Astrophysics* **48**, 339-430.

Bastian, T. S., Benz, A. O., Gary, D. E. (1998): Radio emission from solar flares. *Annual Review of Astronomy and Astrophysics* **36**, 131-188.

Baym, G., Pethick, C. (1979): Physics of neutron stars. *Annual Review of Astronomy and Astrophysics* **17**, 415.

Beck, R., Brandenburg, A., Moss, D., Shukurov, A., Sokoloff, D. (1996): Galactic magnetism: recent developments and perspectives. *Annual Review of Astronomy and Astrophysics* **34**, 155-206.

Becquerel, A. H. (1896): Sur les radiations émises par phosphoescence (On the invisible rays emitted by phosphorescent bodies). *Comptes Rendus de l'Académie des Sciences* **122**, 420-421, 501-503.

Becquerel, A. H. (1900): Déviation du rayonnement du radium dans un champ électrique. *Comptes Rendus de l'Académie des Sciences*, **130**, 809-815.

Beers, T. C., Christlieb, J. (2005): The discovery and analysis of very metal-poor stars in the galaxy. *Annual Review of Astronomy and Astrophysics* **43**, 531-580.

Begelman, M. C., Blandford, R. D., Rees, M. J. (1984): Theory of extragalactic radio sources. *Reviews of Modern Physics* **56**, 255-351.

Belcher, J. W., Davis, L. Jr., Smith, E. J. (1969): Large-amplitude Alfvén waves in the interplanetary medium: *Mariner 5*. *Journal of Geophysical Research* **74**, 2303-2308.

Benacquista, M. (2013): *An Introduction to the Evolution of Single and Binary Stars*. New York: Springer 2013.

Benedict, G., et al. (2002): Astrometry with the *Hubble Space Telescope*: A parallax of the fundamental distance calibrator Delta Cephei. *The Astronomical Journal* **124**, 1695-1705.

Benedict, G., et al. (2007): *Hubble Space Telescope* fine guidance sensor parallaxes of galactic Cepheid variable stars: Period-luminosity relations. *Astronomical Journal* **133**, 1810-1827.

Bennett, C. L. et al. (1993): Scientific results from the *Cosmic Background Explorer* (*COBE*). *Proceedings of the National Academy of Science* **90**, 4766.

Bennett, C. L. et al. (1996): Four-year COBE DMR cosmic microwave background observations - maps and basic results. *Astrophysical Journal Letters* **464**, L1.

Bennett, C. L. et al. (2003): First-year *Microwave Anisotropy Probe* (*WMAP*) observations: preliminary maps and basic results. *Astrophysical Journal Supplement* **148**, Issue 1, 1-27.
Benz, A. O., Güdel, M. (2010): Physical processes in magnetically driven flares on the Sun, stars and young stellar objects. *Annual Review of Astronomy and Astrophysics* **48**, 241-287.
Bergeron, P., Liebert, J., Fulbright, M. S. (1995): Masses of DA white dwarfs with gravitational redshift determinations. *The Astrophysical Journal* **444**, 810.
Bergin, E. A., Tafalla, M. (2007): Cold dark clouds: the initial conditions for star formation. *Annual Review of Astronomy and Astrophysics* **45**, 339-396.
Bertotti, B., Iess, L., Tortora, P. (2003): A test of general relativity using radio links with the Cassini spacecraft. *Nature* **425**, 374-376.
Bessel, F. W. (1839): Bestimmung der Entfemung des 61 Sterns des Schwans (A new determination of the distance of 61 Cygni) – von Hernn Geheimen – Rath und Ritter Bessel. *Astronomische Nachrichten* **16**, No. 356 (5-6), 65.
Bessel, F. W. (1844): On the variations of the proper motions of Procyon and Sirius. *Monthly Notices of the Royal Astronomical Society* **6**, 136.
Bethe, H. A. (1932): Bremsformel für Elektronen relativistischer Geschwindigkeit (A braking formula for relativistic electrons*). Zeitschrift fur Physik* **76**, 293.
Bethe, H. A. (1939): Energy production in stars. *Physical Review* **55**, 434-456. Reproduced in Lang and Gingerich (1979).
Bethe, H. A. (1967): Energy production in stars. Nobel Lecture.
Bethe, H. A. (1990): Supernova mechanisms. *Reviews of Modern Physics* **62**, 801-866.
Bethe, H. A., Critchfield, C. L. (1938): The formation of deutrons by proton combination. *Physical Review* **54**, 248.
Bethe, H. A., Heitler, W. (1934): On the stopping of fast particles and on the creation of positive electrons. *Proceedings of the Royal Society of London* **A146**, 83.
Bhattacharjee, A. (2004): Impulsive magnetic reconnection in the earth's magnetotail and the solar corona. *Annual Review of Astronomy and Astrophysics* **42**, 365-384.
Biermann, L. F. (1946): Zur Deutung der chromosphärischen Turbulenz und des Exzesses der UV-Strahlung der Sonne (An explanation of chromospheric turbulence and the UV excess of solar radiation). *Naturwissenschaften* **33**, 118.
Biermann, L. F. (1947): Über die Ursache der chromosphärischen Turbulenz und des UV-Exzesses der Sonnenstrahlung (About the cause of the chromospheric turbulence and the UV excess of the solar radiation). *Zeitschrift für Astrophysick* **25**, 161.
Biermann, L. F. (1948): Über die Ursache der chromosphärischen Turbulenz und des UV-Exzesses der Sonnenstrahlung. *Zeitschrift für Astrophysik* **25**, 161-177.
Biermann, L. F. (1951): Kometenschweife und solare Korpuskularstrahlung (The tails of comets and the solar corpuscular radiation). *Zeitschrift für Astrophysik* **29**, 274-286.
Biermann, L. F. (1957): Solar corpuscular radiation and the interplanetary gas. *Observatory* **77**, 109. Reproduced in Lang and Gingerich (1979).
Binggeli, B., Sandage, A., Tammann, G. A. (1988): The luminosity function of galaxies. *Annual Review of Astronomy and Astrophysics* **26**, 509.
Binney, J. (1982): Dynamics of elliptical galaxies and other spheroidal components. *Annual Review of Astronomy and Astrophysics* **20**, 399.
Bionta, R. M., *et al*. (1987): Observation of a neutrino burst in coincidence with supernova 1987A in the Large Magellanic Cloud. *Physical Review Letters* **58**, 1494.
Blaauw, A. (1961): On the origin of the O- an B- type stars with high velocities (the "run-away stars") and some related problems. *Bulletin of the Astronomical Institutes of the Netherlands* **15**, 265-290. Reproduced in Lang and Gingerich (1979).
Blaauw, A. (1964): The O associations in the solar neighborhood. *Annual Review of Astronomy and Astrophysics* **2**, 213.
Blaauw, A., Gum, C. S., Pawsey, J. L., Westerhout, G. (1960): The new I.A.U. system of galactic coordinates. *Monthly Notices of the Royal Astronomical Society* **121**, 123-131.
Blaauw, A., Moran, W. W. (1954): Space motions of AE Aurigae and mu Columbae with respect to the Orion Nebula. *The Astrophysical Journal* **119**, 625.

Blackett, P. M. S. (1925): The ejection of protons from nitrogen nuclei, photographed by the Wilson method. *Proceedings of the Royal Society of London*. Series A **107** (742), 349-360.

Blaise, C. (2000): *Time Lord: Sir Sandford Fleming and the Ceation of Standard Time*. New York: Pantheon Books 2000.

Blamont, J. E., Roddier, F. (1961): Precise observation of the profile of the Fraunhofer strontium resonance line. Evidence for the gravitational redshift of the Sun. *Physical Review Letters* **7**, 437.

Blandford, R. D., Narayan, R. (1992): Cosmological applications of gravitational lensing. *Annual Review of Astronomy and Astrophysics* **30**, 311.

Blandford, R. D., Ostriker, J. P. (1980): Supernova shock acceleration of cosmic rays in the Galaxy. *The Astrophysical Journal* **237**, 793.

Blanton, M. R., Moustakas, J. (2009): Physical properties and environments of nearby galaxies. *Annual Review of Astronomy and Astrophysics* **47**, 159-210.

Blum, J., Wurm, G. (2008): The growth mechanisms of macroscopic bodies in protoplanetary disks. *Annual Review of Astronomy and Astrophysics* **46**, 21-56.

Boesgaard, A. M., Steigman, G. (1985): Big bang nucleosynthesis – theories and observations. *Annual Review of Astronomy and Astrophysics* **23**, 319.

Böhm-Vitense, E. (1953): Die Wasserstoffkonvektionszone der Sonne (The hydrogen convection zone of the Sun). *Zeitschrift fur Astrophysik* **32**, 135.

Böhm-Vitense, E. (1958): Über die Wasserstoffkonvektionszone in Sternen verschiedener Effektivtemperaturen und Leuchtkräfte (About the hydrogen convection zone in stars of various effective temperatures and luminosities). *Zeitschrift fur Astrophysik* **46**, 108.

Böhm-Vitense, E. (1981): The effective temperature scale. *Annual Review of Astronomy and Astrophysics* **19**, 295.

Böhm-Vitense, E. (1989, 1992): *Introduction to Stellar Astrophysics. Volume 1. Basic Stellar Observations and Data. Volume 2. Stellar Atmospheres. Volume 3. Stellar Structure and Evolution*. Cambridge, England: Cambridge University Press, 1989, 1992.

Bohr, N. (1913): On the constitution of atoms and molecules. *Philosophical Magazine* **26**, 1, 10, 476, 857.

Bohr, N. (1914): On the spectrum of hydrogen. *Fysisk Tidsskrift* **12**, 97.

Bohr, N. (1930): Faraday lecture: Chemistry and the Quantum Theory of Atomic Constitution. *Journal of the Chemical Society of London*, 349-389 (1930) – delivered 8 May 1930.

Bolton, C. T. (1972): Identification of Cygnus X-1 with HDE 226868. *Nature* **235** (2), 271-273. Reproduced in Lang and Gingerich (1979).

Bolton, C. T. (1975): Optical observations and model for Cygnus X-1. *The Astrophysical Journal* **200**, 269-277.

Bolton, J. G., Stanley, G. J., Slee, O. B. (1949): Positions of three discrete sources of galactic radio-frequency radiation. *Nature* **164**, 101. Reproduced in Lang and Gingerich (1979).

Bolton, J. G., Wild, J. P. (1957): On the possibility of measuring interstellar magnetic fields by 21-cm Zeeman splitting. *Astrophysical Journal* **125**, 296.

Boltwood, B. B. (1907): On the ultimate disintegration products of the radioactive elements II. *American Journal of Science* **23**, 77-88.

Boltwood, B. B. (1907): The origin of radium. *Nature* **76**, 589.

Boltzmann, L. (1868): Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten (Studies of the equilibrium and the life force between material points). *Wien. Ber.* **58**, 517. English translation in Brush (1948).

Boltzmann, L. (1872): Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen (Further studies on the thermal equilibrium of gas molecules). *Sitz. Acad. Wiss.* **66**, 275 (1872).

Bond, G. P. (1863): Light of the moon and of Jupiter. *Memoirs of the American Academy* **8**, 221.

Bondi, H., Gold, T. (1948): The steady-state theory of the expanding universe. *Monthly Notices of the Royal Astronomical Society* **108**, 252-270. Reproduced in Lang and Gingerich (1979).

Bonnefoy, M. et al. (2011): High angular resolution detection of Beta Pictoris b at 2.18 micrometers. *Astronomy and Astrophysics* **528**, L15.

Bowen, I. S. (1927): The origin of the nebulium spectrum. *Nature* **120**, 473.
Bowen, I. S. (1928): The origin of the nebular lines and the structure of the planetary nebulae. *The Astrophysical Journal* **67**, 1. Reproduced in Lang and Gingerich (1979).
Bowyer, S. et al. (1964): Lunar occultation of x-ray emission from the Crab Nebula. *Science* **146** (3646), 912-917.
Bowyer, S. et al. (1965): Cosmic x-ray sources. *Science* **147** (3656), 394-398.
Boyle, R. (1660): The spring of the air. In *New experiments physico-mechanical, touching the spring of air, and its effects; made for the most part in a new pneumatical engine*. Oxford, 1660. Reproduced in Brush (1948).
Boyle, R. (1662): A defense of the doctrine touching the spring and weight of the air,…. Oxford 1662.
Brackett, F. S. (1922): Visible and infra-red radiation of hydrogen. *The Astrophysical Journal* **56**, 154.
Bradley, J. (1728): A new apparent motion in the fixed stars discovered, its cause assigned, the velocity and equable motion of light deduced. *Philosophical Transactions of the Royal Society of London* **6**, 149.
Bradley, J. (1748): An apparent motion in some of the fixed stars. *Philosophical Transactions of the Royal Society of London* **10**, 32.
Bradt, H. L., Peters, B. (1950): The heavy nuclei of the primary cosmic radiation. *Physical Review* **77**, 54-70.
Branch, D. (1998): Type Ia supernovae and the Hubble constant. *Annual Review of Astronomy and Astrophysics* **36**, 17-55.
Branch, D., Tammann, G. A. (1992): Type Ia supernovae as standard candles. *Annual Review of Astronomy and Astrophysics* **30**, 359.
Breton, R. P., et al. (2008): Relativistic spin precession in the double pulsar. *Science* **321** (5885), 104-107.
Bridle, A. H., Perley, R. A. (1984): Extragalactic radio jets. *Annual Review of Astronomy and Astrophysics* **22**, 319.
Brodie, J. P., Stader J. (2006): Extragalactic globular clusters and galaxy formation. *Annual Review of Astronomy and Astrophysics* **44**, 193-267.
Bromm, V., Larson, R. B. (2004): The first stars. *Annual Review of Astronomy and Astrophysics* **42**, 79-118.
Bromm, V., Yoshida, N. (2011): The first galaxies. *Annual Review of Astronomy and Astrophysics* **49**, 373-407.
Bruner, E. C., Jr. (1981): *OSO 8* observational limits to the acoustic coronal heating mechanism. *The Astrophysical Journal* **247**, 317-324.
Bruno, G. (1584*): De l'infinito universe et mondi* (On the Infinite Universe and Worlds). Venice, Stampato in Venetia 1584. Third dialogue between Burchio, Elpino, Faracastoro and Philotheo. English translation by Dorothea Waley Singer, New York: Henry Schuman 1950.
Brush, S. G. (1948): *Kinetic theory, vol. I. The nature of gases and of heat.* Oxford: Pergamon Press 1948.
Brush, S. G., Holton, G. J. (2001): *Physics, the human adventure: from Copernicus to Einstein and beyond*. New Brunswick, New Jersey: Rutgers University Press pp. 137.
Buffon, Comte de (Georges Louis Leclerc) (1774): *Epoques de la nature* (Paris), Partial English translation in *Source book in Geology 1400-1900*, ed. K. F. Mather and S. L. Mason (Cambridge, Mass: Harvard University Press, 1970) 65-73.
Burbidge, E. M., Burbidge, G. R., Fowler, W. A., Hoyle, F. (1957): Synthesis of the elements in stars. *Reviews of Modern Physics* **29**, 547-650. Reproduced in Lang and Gingerich (1979).
Burbidge, G. R., Burbidge, E. M., Sandage, A. R. (1963): Evidence for the occurrence of violent events in the nuclei of galaxies. *Reviews of Modern Physics* **35**, 947-972.
Burchfield, J. D. (1975): *Lord Kelvin and the Age of the Earth*. New York: Science History Publications 1975. Chicago: University of Chicago Press 1990.
Burstein, D. (1990): Large-scale motions in the universe: a review. *Reports Progress Physics* **53**, 421-481.

Busso, M., Gallino, R., Wasserburg, G. J. (1999): Nucleosynthesis in asymptotic giant branch stars: Relevance for galactic enrichment and solar system formation. *Annual Review of Astronomy and Astrophysics* **37**, 239-309.
Byram, E. T., Chubb, T. A., Friedman, H. (1953): The contribution of solar X-rays to E-layer ionization. *Physical Review* **92**, 1066-1067.
Byram, E. T., Chubb, T. A., Friedman, H. (1954): Solar X-ray emission. *Physical Review* **96**, 860.
Byram, E. T., Chubb, T. A., Friedman, H. (1956): The solar X-ray spectrum and the density of the upper atmosphere. *Journal of Geophysical Research* **61**, 251-263.
Cameron, A. G. W. (1957): Nuclear reactions in stars and nucleogenesis. *Publications of the Astronomical Society of the Pacific* **69**, 201.
Cameron, A. G. W. (1957): Stellar evolution, nuclear astrophysics and nucleogenesis. *Atomic Energy of Canada, Chalk River Project AECL 454, CRL-41*.
Canal, R., Isern, J., Labay, J. (1990): The origin of neutron stars in binary systems. *Annual Review of Astronomy and Astrophysics* **28**, 183.
Cannon, A. J., Pickering, E. C. (1918-1924): The Henry Draper Catalogue. *Annals of the Harvard College Observatory* **91-99**.
Carilli, C. L., Taylor, G. B. (2002): Cluster magnetic fields. *Annual Review of Astronomy and Astrophysics* **40**, 319-348.
Carnot, S. (1824): *Reflexions sur la Puissance Motrice du Feu et sur les Machines* (*Reflections on the Motive Power of Fire and the Machines*). Bachelier, Paris 1824, Reproduced New York: Dover 1960.
Carr, B. (1994): Baryonic dark matter. *Annual Review of Astronomy and Astrophysics* **32**, 531.
Carrington, R. C. (1858): On the distribution of the solar spots in latitude since the beginning of the year 1854. *Monthly Notices of the Royal Astronomical Society* **19**, 1-3. Reproduced in Meadows (1970).
Carrington, R. C. (1860): Description of a singular appearance seen in the Sun on September 1, 1859. *Monthly Notices of the Royal Astronomical Society* **20**, 13-15. Reproduced in Meadows (1970).
Carrington, R. C. (1863): *Observations of the Spots on the Sun*. London: Williams and Norgate 1863.
Carroll, S. M., Press, W. H., Turner, E. L. (1992): The cosmological constant. *Annual Review of Astronomy and Astrophysics* **30**, 499.
Carson, T. R. (1976): Stellar opacity. *Annual Review of Astronomy and Astrophysics* **14**, 95.
Carter, B., Carter, M. S. (2002): *Latitude: How American Astronomiers Solved the Mystery of Variation*. Annapolis, Maryland: Naval Institute Press 2002.
Cassinelli, J. P. (1979): Stellar winds. *Annual Review of Astronomy and Astrophysics* **17**, 275.
Caughlan, G. R., Fowler, W. A. (1962): The mean lifetimes of carbon, nitrogen, and oxygen nuclei in the CNO bi-cycle. *The Astrophysical Journal* **136**, 453.
Cavendish, H. (1798): Experiments to determine the density of Earth. *Philosophical Transactions of the Royal Society of London* **88**, 469-526.
Chaboyer, B. (1995): Absolute ages of globular clusters and the age of the universe. *Astrophysical Journal Letters* **444**, L9.
Chadwick, J. (1932a): Possible existence of a neutron. *Nature* **192**, 312.
Chadwick, J. (1932b): The existence of a neutron. *Proceedings of the Royal Society of London* Series A **136**, 692-708.
Chandrasekhar, S. (1931): The maximum mass of ideal white dwarfs. *The Astrophysical Journal* **74**, 81.
Chandrasekhar, S. (1938, 1939): *An Introduction to the Study of Stellar Structure*. Chicago, Illinois: University of Chicago Press 1938, 1939.
Chandrasekhar, S. (1961): *Hydrodynamics and hydromagnetic stability*. Oxford: Oxford at the Clarendon Press 1961.
Chandrasekhar, S. (1969): *Ellipsoidal Figures of Equilibrium*. New Haven, Conn.: Yale University Press 1969.

Charlier, C. V. L. (1922): How an infinite world may be built up. *Arkiv för Matematik, Astronomi och Fysik* **16**, No. 22, 1.

Cherenkov, P. A. (1937): Visible radiation produced by electrons moving in a medium with velocities exceeding that of light. *Physical Review* **52**, 378.

Cheung, A. C., Rank, D. M., Townes, C. H., Thornton, D. D., Welch, W. J. (1968): Detection of NH_3 molecules in the interstellar medium by their microwave emission. *Physical Review Letters* **21**, 1701.

Chiosi, C., Bertelli, G. Bressan, A. (1992): New developments in understanding the H−R diagram. *Annual Review of Astronomy and Astrophysics* **30**, 235.

Chiosi, C., Maeder, A. (1986): The evolution of massive stars with mass loss. *Annual Review of Astronomy and Astrophysics* **24**, 329.

Christy, R. F. (1966): Pulsation theory. *Annual Review of Astronomy and Astrophysics* **4**, 353.

Chupp, E. L. (1984): High energy neutral radiations from the Sun. *Annual Review of Astronomy and Astrophysics* **22**, 359.

Churchwell, E. (2002): Ultra-compact H II regions and massive star formation. *Annual Review of Astronomy and Astrophysics* **40**, 27-62.

Chwolson, O. (1924): Über eine mögliche Form fiktiver Doppelsterne. *Astronomische Nachrichten* **221** (20), 329.

Clairault, A. C. (1743): *Theory of the Figure of the Earth, and the Principles of Hydrostatics*. Paris: Du Mond 1743.

Clapeyron, E. (1834): Mémoire sur la puissance motrice de la chaleur. *Journal de l'École Polytechnique* **14**, 153-190.

Clausius, R. (1850): Über die bewegende Kraft der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen (On the moving force of heat and the laws of thermodynamics that can be deduced from it). *Annalen der Physik und Chemie* **79**, 368-397, 500-524. English translation in *Philosophical Magazine* **2**, 1 (1851).

Clausius, R. (1857): Über die Art der Bewegung, welche wir Wärme nennen (The nature of the motion which we call heat). *Annalen der Physik und Chemie* **100**, 353-380. English translation in *Philosophical Magazine* **14**, 108 (1857) and Reproduced in Brush (1948).

Clausius, R. (1858): Über die mittlere Länge der Wege, welche bei Molecularbewegung gasförmigen Körper von den einzelnen Molecülen zurückgelegt werden, nebst einigen anderen Bemerkungen über die mechanischen Wärmetheorie (On the mean lengths of the paths described by the separate molecules of gaseous bodies). *Annalen der Physik* **105,** 239-258. English translation by F. Guthrie in *Philosophical Magazine* **17**, 81-91 (1859) and reproduced in Brush (1948).

Clausius, R. (1870): Ueber einen auf die Wärme anwendbaren mechanischen Satz (On a mechanical theorem applicable to heat). *Sitzungsberichte der Niedderrheinischen Gesellschaft*, Bonn.114-119. English translation reproduced in Brush (1948).

Clay, J. (1932): The Earth-magnetic effect and the corpuscular nature of (cosmic) ultra-radiation. IV. *Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam, Proceedings of the Section of Sciences* **35**, 1282-1290.

Clayton, D. D. (1968, 1984): *Principles of Stellar Evolution and Nucleosynthesis*. Chicago, Illinois: University of Chicago Press 1968, 1984.

Cleveland, B. T., et al. (1998): Measurement of the solar electron neutrino flux with the Homestake chlorine detector, *The Astrophysical Journal* **496**, 505-526.

Cockcroft, J. D., Walton, E. T. S. (1932a): Disintegration of lithium by swift protons. *Nature* **129**, 242.

Cockcroft, J. D., Walton, E. T. S. (1932b): Experiments with high velocity positive ions. *Proceedings of the Royal Society of London* **136**, 619-630.

Colombo, G., Shapiro, I. I. (1966): The rotation of the planet Mercury. *The Astrophysical Journal* **145**, 296.

Combes, F. (1991): Distribution of CO in the Milky Way. *Annual Review of Astronomy and Astrophysics* **29**, 195.

Comella, J. M., Craft, H. D., Lovelace, R. V. E., Sutton, J. M., Tyler, G. L. (1969): Crab Nebula pulsar NP 0532. *Nature* **221**, 453.

Compton, A. H. (1923a): A quantum theory of the scattering of X-rays by light elements. *Physical Review* **21**, 207, 483.

Compton, A. H. (1923b): The spectrum of scattered X-rays. *Physical Review* **22** (5), 409.

Compton, A. H. (1932): Variation of the cosmic rays with latitude. *Physical Review* **41**, 111-113.

Copi, C. J., Schramm, D. N., Turner, M. S. (1995): Big-bang nucleosynthesis and the baryon density of the universe. *Science* **267**, 192.

Coulomb, C.-A. (1785): Priemier, Second, Troisieme mémoire sur l'électricité et le magnétisme. *Histoire de l'Academie Royale des Sciences*, 569-577, 578-611, 612-638.

Coulomb, C.-A. (1787): Quatrième, Cinquième, Sixième mémoire sur l'electricite. *Histoire de l'Academie Royale des Sciences*, 67-77, 421-467, 617-705.

Coulomb, C.-A. (1789): Septieme mémoire sur l'électricité et le magnétisme. *Histoire del'Academie Royale des Sciences*, 455-505.

Couvidat, S., Turck-Chieze, S., Kosovichev, A. G. (2003): Solar seismic models and the neutrino predictions. *The Astrophysical Journal* **599**, 1434-1448.

Cowan, C. L., Jr., et al. (1956): Detection of the free neutrino: a confirmation. *Science **124,** 103.*

Cowling, T. G. (1945a): On the Sun's general magnetic field. *Monthly Notices of the Astronomical Society* **105**, 166.

Cowling, T. G. (1945b): The electrical conductivity of an ionized gas in a magnetic field, with applications to the solar atmosphere and the ionosphere. *Proceedings of the Royal Society of London* **A183**, 453.

Cox, A. N. (1980): The masses of Cepheids. *Annual Review of Astronomy and Astrophysics* **18**, 15.

Cox, D. P., Reynolds, R. J. (1987): The local interstellar medium. *Annual Review of Astronomy and Astrophysics* **25**, 303.

Cox, J. P. (1980): *Theory of Stellar Pulsation*. Princeton, New Jersey: Princeton University Press 1980.

Cranmer, S. R. (2002): Coronal holes and the high-speed solar wind. *Space Science Reviews* **101**, 229-294.

Cranmer, S. R., Van Ballegooijen, A. A. (2005): On the generation, propagation and reflection of Alfvén waves from the photosphere to the distant heliosphere. *Astrophysical Journal Supplement Series* **156**, 265-293.

Crenshaw, D. M., Kraemer, S. B., George, I. M. (2003): Mass loss from the nuclei of active galaxies. *Annual Review of Astronomy and Astrophysics* **41**, 117-167.

Crutcher, R. M. (2012): Magnetic fields in molecular clouds. *Annual Review of Astronomy and Astrophysics* **50**, 29-63.

Curie, M. S. (1898): Rays emitted by compounds of uranium and of thorium. *Comptes Rendus* **126**, 1101-1103.

Curie, P., Curie, M. S. (1898): Sur une nouvelle substance fortement radio-active, contenue dans la pechblende. (On a strongly radioactive substance contined in pitchblende. *Comptes Rendus de l'Académie des Sciences* **127**, 1215-1217.

Curtis, H. D. (1919): Modern theories of the spiral nebulae. *Journal of the Washington Academy of Sciences* **9**, 217-227. Reproduced in Lang and Gingerich (1979).

D'Antona, F., Mazzitelli, I. (1990): Cooling of white dwarfs. *Annual Review of Astronomy and Astrophysics* **28**, 139.

Davis, L. Jr. (1955): Interplanetary magnetic fields and cosmic rays. *Physical Review* **100**, 1440-1444.

Davis, L. Jr., Greenstein, J. L. (1951): The polarization of starlight by aligned dust grains. *The Astrophysical Journal* **114**, 506.

Davis, M., Efstathiou, G., Frenk, C. S., White, S. D. M. (1985): The evolution of large-scale structure in a universe dominated by cold dark matter. *The Astrophysical Journal* **292**, 371.

Davis, M., Peebles, P. J. E. (1983): Evidence for local anisotropy of the Hubble flow. *Annual Review of Astronomy and Astrophysics* **21**, 109.

Davis, R., Jr. (1964): Solar neutrinos II. Experimental. *Physical Review Letters* **12**, 303-305.
Davis, R., Jr., Harmer, D. S., Hoffman, K. C. (1968): Search for neutrinos from the Sun. *Physical Review Letters* **20**, 1205. Reproduced in Lang and Gingerich (1979).
De Bort, T. P. T. (1902): Variations de la température de l'air libre, dans la zone comprise entre 9 et 5 kilomètres d'altitude. *Comptes rendus de l'Académie des sciences* **134**, 987-989.
De Broglie, L. (1923): Waves and quanta. *Nature* **112**, 540.
De Pontieu, B., et al. (2007): Chromospheric Alfvenic waves strong enough to power the solar wind. *Science* **318**, 1574-1576.
De Vaucouleurs, G. (1953): Evidence for a local supergalaxy. *Astronomical Journal* **58**, 29.
Debye, P. (1909): Der Lichtdruck auf Kugeln von beliebigem Material (The light pressure on spheres of arbitrary material). *Annalen der Physik.* **30**, 57.
Decker, R. B., et al. (2005): *Voyager 1* in the foreshock, termination shock, and heliosheath. *Science* **309**, 2020-2024.
Defrancesco, S. (1988): Schiaparelli's determination of the rotation period of Mercury: a re-examination. *Journal of the British Astronomical Association* **98**, no. 3, 146-150.
Demorest, P. B. et al. (2010): A two-solar-mass neutron star measured using Shapiro delay. *Nature* **467** (7319), 1081-1083.
Deubner, F.-L. (1975): Observations of low wave number nonradial eigenmodes of the Sun. *Astronomy and Astrophysics* **44**, 371-375.
Deubner, F.-L., Gough, D. (1984): Helioseismology – oscillations as a diagnostic of the solar interior. *Annual Review of Astronomy and Astrophysics* **22**, 593-619.
Dicke, R. H., Peebles, P. J. E., Roll, P. G., Wilkinson, D. T. (1965): Cosmic black-body radiation. *The Astrophysical Journal* **142**, 414-419.
Dickey, J. O., *et al*. (1994): Lunar laser ranging - a continuing legacy of the Apollo program. *Science* **265**, 482.
Dirac, P. A. M. (1928): The quantum theory of the electron. *Proceedings of the Royal Society of London, Series A* **117** (778), 610-624.
Doeleman, S. S. et al. (2008): Event-horizon-scale structure in the supermassive black hole candidate at the galactic centre. *Nature* **455** (7209), 78-80.
Donati, J.-F., Landstreet, J. D. (2009): Magnetic fields of nondegenerate stars. *Annual Review of Astronomy and Astrophysics* **47**, 333-370.
Doppler, C. (1842): Über das farbige Licht der Doppelsterne und einiger anderer Gestirne des Himmels (On the colored light of double stars, etc.). *Abhandlungen der königlichen. Böhmischen Gesellschaft der Wissenschaften*. **2**, 465.
Doppler, C. (1846): Bemerkungen zu meiner Theorie des farbigen Lichtes der Doppelserne. *Annalen der Physik und Chemie* **68**, 1-35..
Draine, B. T. (2003): Interstellar dust grains. *Annual Review of Astronomy and Astrophysics* **41**, 241-289,
Dulk, G. A. (1985): Radio emission from the Sun and other *stars. Annual Review of Astronomy and Astrophysics* **23**, 169-224.
Dullemand, C. P., Monnier, J. D. (2010): The inner regions of protoplanetary disks. *Annual Review of Astronomy and Astrophysics* **48**, 205-239.
Duncan, J. C. (1921): Changes observed in the Crab Nebula in Taurus. *Proceedings of the National Academy of Science* **7** (6), 179-180.
Duncan, J. C. (1939): Second report on the expansion of the Crab Nebula. *The Astrophysical Journal* **89**, 482.
Dungey, J. W. (1961): Interplanetary magnetic field and the auroral zones. *Physical Review Letters* **6**, 47-48.
Dupree, A. K. (1986): Mass loss from cool stars. *Annual Review of Astronomy and Astrophysics* **24**, 377.
Dyce, R. B., Pettengill, G. H., Shapiro, I. I. (1967): Radar determinations of the rotations of Venus and Mercury. *Astronomical Journal* **72**, 351-359. Reproduced in Lang and Gingerich (1979).

Dyson, F. W., Eddington, A. S., Davidson, C. (1920): A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London, Series A* **220**, 291-333. Reproduced in Lang and Gingerich (1979).

Eardley, D. M., Press, W. H. (1975): Astrophysical processes near black holes. *Annual Review of Astronomy and Astrophysics* **13**, 381.

Earl, J. A. (1961): Cloud-chamber observations of primary cosmic-ray electrons. *Physical Review Letters* **6**, 125-128.

Eddington, A. S. (1916): The kinetic energy of a star cluster. *Monthly Notices of the Royal Astronomical Society* **76**, 525-528.. Reproduced in Lang and Gingerich (1979).

Eddington, A. S. (1916/17): On the radiative equilibrium of stars. *Monthly Notices of the Royal Astronomical Society* **77**, 16-35, 596-612. Reproduced in Lang and Gingerich (1979).

Eddington, A. S. (1917): The pulsation theory of Cepheid variables. *The Observatory* **40**, 290.

Eddington, A. S. (1918): *Report on the Relativity Theory of Gravitation*. London: Physical Society of London 1918.

Eddington, A. S. (1918/19): On the pulsations of a gaseous star and the problem of the Cepheid variables. *Monthly Notices of the Royal Astronomical Society* **79**, 2-22, 177-188. Reproduced in Lang and Gingerich (1979).

Eddington, A. S. (1919): The total eclipse of 1919 May 29 and the influence of gravitation on light. *Observatory* **42**, 119.

Eddington, A. S. (1920): The internal constitution of the stars. *Nature* **106**, 14-20, *Observatory* **43**, 341 (1920). Reproduced in Lang and Gingerich (1979).

Eddington, A. S. (1923): *The Mathematical Theory of Relativity*. Cambridge, England: Cambridge University Press 1923.

Eddington, A. S. (1924): On the relation between the masses and luminosities of the stars. *Monthly Notices of the Royal Astronomical Society* **84**, 308-332. Reproduced in Lang and Gingerich (1979).

Eddington, A. S. (1926a): Diffuse matter in interstellar space. *Proceedings of the Royal Society of London* **A111**, 424.

Eddington, A. S. (1926b): *The Internal Constitution of the Stars*. Cambridge, England: Cambridge University Press 1926.

Eddington, A. S. (1933): *The Expanding Universe*. Cambridge, England: Cambridge University Press 1933.

Eddington, A. S. (1941): On the cause of Cepheid pulsation. *Monthly Notices of the Royal Astronomical Society* **101**, 182.

Eddington, A. S. (1942): Conditions in the hydrogen convection zone. *Monthly Notices of the Royal Astronomical Society* **102**, 154.

Edlén, B. (1941): An attempt to identify the emission lines in the spectrum of the solar corona. *Arkiv för Matematik, Astronomi och Fysik* **28B**, 1-4. Reproduced in Lang and Gingerich (1979).

Edlén, B. (1945): The identification of the coronal lines. *Monthly Notices of the Royal Astronomical Society* **105**, 323-333.

Eggen, O. J., Lynden-Bell, D., Sandage, A. R. (1962): Evidence from the motions of old stars that the Galaxy collapsed. *The Astrophysical Journal* **136**, 748.

Eguci, K., et al. (2003): First results from Kamland: evidence for reactor anti-neutrino disappearance. *Physical Review Letters* **90**, 021802.

Ehrenfreund, P., Chamley, S. B. (2000): Organic molecules in the interstellar medium, comets and meteorites: a voyage from dark clouds to the early Earth. *Annual Review of Astronomy and Astrophysics* **38**, 427-483.

Einasto, M. et al. (2011): The Sloan great wall. Morphology and galaxy content. *The Astrophysical Journal* **736** , 51.

Einstein, A. (1905a): Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt (On a heuristic point of view about the creation and converson of light). *Annalen der Physik* **17**, 132-148.

Einstein, A. (1905b): Ist die Trägheit eines Körpers von seinem Energeieinhalt abhängig? (Does the inertia of a body depend upon its energy content?). *Annalen der Physik* **18**, 639-641. English translation in Lang and Gingerich (1979).

Einstein, A. (1906): Das Prinzip von der Erhaltung der Schwerpunktsbewegung und die Trägheit der Energie. (The principle of the conservation of the motion of the center of gravity and the inertia of energy). *Annalen der Physik* **20**, 627-633. English translation in Lang and Gingerich (1979).

Einstein, A. (1907): Über das Relativitätsprinzip und die aus demselben gezogenen Fogerungen (About the principle of relativity and the conclusions drawn therefrom). *Jahrb. Radioact. Elekt.* **4**, 411. English translation in: *The Principle of Relativity* (ed. A. Sommerfeld). New York: Dover 1952.

Einstein, A. (1907): Über die vom Relativitätsprinzip geforderte Trägheit der Energie (Variation of intertia with energy on the principle of relativity). *Annalen der Phyik* **23**, 371.

Einstein, A. (1911): Über den Einfluβ der Schwerkraft auf die Ausbreitung des Lichtes (On the influence of gravity on the propagation of light). *Ann. Phys.* **35**, 898. English translation in *The Principle of Relativity*. New York, Dover 1952.

Einstein, A. (1915): Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie (Explanation of the perihelion motion of Mercury by means of the general theory of relativity). *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin* **11**, 831-839. English translation in Lang and Gingerich (1979).

Einstein, A. (1916): Die Grundlage der allgemeinen Relativitätstheorie (The foundation of the general theory of relativity). *Annalen der Physik* **49** (7), 769-822. English translation in *The Principle of Relativity* (ed. A. Sommerfeld). New York: Dover 1952.

Einstein, A. (1917a): Zur Quantentheorie der Strahlung. (On the quantum mechanics of radiation. *Physikalische Zeitschrift* **18**, 121-128.

Einstein, A. (1917b): Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie (Cosmological considerations of the general theory of relativity). *Acad. Wiss.* **1**, 142 (1917). English translation in: *The Principle of Relativity* (ed. A. Sommerfeld). New York: Dover 1952.

Einstein, A. (1936): Lens-like action of a star by the deviation of light in the gravitational field. *Science* **84** (2188), 506-507.

Einstein, A., de Sitter, W. (1932): On the relation between the expansion and the mean density of the universe. *Proceedings of the National Academy of Sciences* **18**, 213-214. Reproduced in Lang and Gingerich (1979).

Eisenstein, D. J. et al. (2005): Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *The Astrophysical Journal* **633**, 560-574.

Elder, R. R. et al. (1947): Radiation from electrons in a synchrotron. *Physical Review* **71**, 829-830.

Elmegreen, B. G., Lada, C. J. (1977): Sequential formation of subgroups in OB associations. *The Astrophysical Journal* **214**, 725.

Encrenaz, T. (2008): Water in the solar system. *Annual Review of Astronomy and Astrophysics* **46**, 57-87.

Essen, L. (1969): The measurement of time. *Vistas in Astronomy* **2**, 45.

Evans, N. J. II (1999): Physical conditions in regions of star formation. *Annual Review of Astronomy and Astrophysics* **37**, 311-362.

Ewen, H. I., Purcell, E. M. (1951): Radiation from galactic hydrogen at 1,420 MHz. *Nature* **168**, 356. Reproduced in Lang and Gingerich (1979).

Faber, S. M., Gallagher, J. S. (1979): Masses and mass-to-light ratios of galaxies. *Annual Review of Astronomy and Astrophysics* **17**, 135.

Fabian, A. C. (2012): Observational evidence of active galactic nuclei feedback. *Annual Review of Astronomy and Astrophysics* **50**, 455-489.

Faraday, M. (1839): *Experimental Researches in Electricity, Volume* 1. London: Richard and John Edward Taylor 1839.

Faraday, M. (1843): On static electrical inductive action. *Philosophical Magazine* **22**, 200.

Faraday, M. (1844): *Experimental Researches in Electricity, Volume 2*. London: Richard and John Edward Taylor 1844.

Feast, M. W., Walker, A. R. (1987): Cepheids as distance indicators. *Annual Review of Astronomy and Astrophysics* **25**, 345.

Feast, M. W., Whitelock, P. (1997): Galactic kinematics of Cepheids from *HIPPARCOS* proper motions. *Monthly Notices of the Royal Astronomical Society* **291**, 683.

Fender, R., Belloni, T. (2004): GRS 1915+105 and the disc-jet coupling in accreting black hole systems. *Annual Review of Astronomy and Astrophysics* **42**, 317-364.

Fermi, E. (1926): Zur Quantelung des idealen einatomigen Gases (On quantisation of the ideal monatomic gas). *Zeitschrift fur Physik* **36**, 902.

Fermi, E. (1928): Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente (A statistical method for determining the eigenstate of atoms and its application to the theory of the periodical system of the elements). *Zeitschrift fur Physik* **48**, 73.

Fermi, E. (1934): Versuch einer Theorie der β-strahlen I. (An attempt at the theory of β-rays). *Zeitschrift fur Physik* **88**, 161. English translation in *American Journal of Physics* **36**, 1150 (1968).

Fernie, J. D. (1969): The period-luminosity relation – a historical review. *Publications of the Astronomical Society of the Pacific* **81**, 707.

Fiegelson, E. D., Montmerle, T. (1999): High-energy processes in young stellar objects. *Annual Review of Astronomy and Astrophysics* **37**, 363-408.

Field, G. B., Somerville, W. B., Dressler, K. (1966): Hydrogen molecules in astronomy. *Annual Review of Astronomy and Astrophysics* **4**, 207.

Filippenko, A. (1997): Optical spectra of supernovae. *Annual Review of Astronomy and Astrophysics* **35**, 309-355.

Fish, M., Tremaine, S. (1991): The mass of the Galaxy. *Annual Review of Astronomy and Astrophysics* **29**, 409.

Fishman, G. J., Meegan, C. A. (1995): Gamma ray bursts. *Annual Review of Astronomy and Astrophysics* **33**, 415.

Fixler, J. B. et al. (2007): Atom interferometer measurement of the Newtonian constant of gravity *Science* **315** (5808), 74-77.

Fomalont, E. B. et al. (2009): Progress in measurements of the gravitational bending of radio waves using the VLBA. *The Astrophysical Journal* **699** (2), 1395-1402.

Fomalont, E. B., Sramek, R. A. (1976): Measurements of the solar gravitational deflection of radio waves in agreement with general relativity. *Physical Review Letters* **36**, 1475.

Forbes, E. G. (1970): The Einstein effect and the observed solar red shifts. *Observatory* **90**, 149.

Forbush, S. E. (1950): Cosmic-ray intensity variations during two solar cycles. *Journal of Geophysical Research* **63**, 651-669.

Fourier, J. (1824): Remarques générales sur les températures du globe terrestre et des espaces planétaires (General remarks on the temperature of the Earth and outer space). *Annales de Chimie et de Physique* **27**, 136-167. English translation by Ebeneser Burgess, *American Journal of Science* **32**, 1-20 (1837).

Fourier, J. (1827): Mémoire sur les températures du globe terrestre et des espaces planétaires. *Mémoires de l'Académie Royale des Sciences* **7,** 569-604.

Fowler, R. H. (1926): On dense matter. *Monthly Notices of the Royal Astronomical Society* **87**, 114-122. Reproduced in Lang and Gingerich (1979).

Fowler, R. H., Milne, E. A. (1924): The maxima of absorption lines in stellar spectra (second paper). *Monthly Notices of the Royal Astronomical Society* **94**, 499.

Fowler, W. A. (1984): Experimental and theoretical astrophysics: the quest for the origin of the elements. *Reviews of Modern Physics* **56**, 149.

Fowler, W. A., Caughlan, G. R., Zimmerman, B. A. (1975): Thermonuclear reaction rates II. *Annual Review of Astronomy and Astrophysics* **13**, 69-112.

Fowler, W. A., Hoyle, F. (1960): Nuclear cosmochronology. *Annals of Physics* **10**, 280.

Fowler, W. A., Hoyle, F. (1964): Neutrino processes and pair formation in massive stars and supernovae. *Astrophysical Journal Supplement* **9**, 201.

Frail, D. A., Kulkarni, S. R., Nicastro, L., Feroci, M., Taylor, G. B. (1997): The radio afterglow from the γ-ray burst of 8 May 1997. *Nature* **389**, 261.

Fraunhofer, J. (1817): Bestimmung des Brechungs- und des Farbenzerstreungs- Vermögens verschiedener Glasarten, in Bezug auf die Vervollkommnung achromatischer Fernröhre. *Annalen der Physik* **56**, 264-313. Also see Shapley and Howarth (1929).

Freedman, W. L. et al. (2001): Final results from the *Hubble Space Telescope* key project to measure the Hubble constant. *The Astrophysical Journal* **553**, 47-72.

Freedman, W. L. et al. (2012): Carnegie Hubble program: A mid-infrared calibration of the Hubble constant. *The Astrophysical Journal* **758**, 24-34.

Freedman, W. L., Madore, B. F. (2010): The Hubble constant. *Annual Review of Astronomy and Astrophysics* **48**, 673-710.

Freeman, K., Bland-Hawthom, J. (2002): The new Galaxy: signatures of its formation. *Annual Review of Astronomy and Astrophysics* **40**, 487-537.

Frenk, C. S., et al. (1985): Cold dark matter, the structure of galactic haloes and the origin of the Hubble sequence. *Nature* **317**, 595.

Friedmann, A. (1922): Über die Krümmung des Raumes (On the curvature of space). *Zeitschrift für Physik* **10**, 377. English translation in Lang and Gingerich (1979).

Friedmann, A. (1924): Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes (On the possibility of a universe with a constant negative curvature of space). *Zeitschrift für Physik* **21**, 326.

Frieman, J. A., Turner, M. S., Huterer, D. (2008): Dark energy and the accelerating universe. *Annual Review of Astronomy and Astrophysics* **46**, 385-432.

Frisch, P. C., Redfield, S., Slavin, J. D. (2011): The interstellar medium surrounding the Sun. *Annual Review of Astronomy and Astrophysics* **49**, 237-279.

Fukuda, Y. et al. (1998a): Evidence for oscillation of atmospheric neutrinos. *Physical Review Letters* **81**, 1562-1567.

Fukada, Y. et al. (1998b): Measurements of the solar neutrino flux from super-kamiokande's first 300 days. *Physical Review Letters* **81** (6), 1158-1162.

Fukui, Y., Kawamura, A. (2010): Molecular clouds in nearby galaxies. *Annual Review of Astronomy and Astrophysics* **48**, 547-580.

Fynbo, H. O. U. et al. (2004): Revised rates for the stellar triple-alpha process from measurement of 12C nuclear resonance. *Nature* **433** (7022) 136-139.

Gaensler, B. M., Slane, P. O. (2006): The evolution and structure of pulsar wind nebulae. *Annual Review of Astronomy and Astrophysics* **44**, 17-47.

Galilei, G. (1610): *Sidereus Nuncius* (*The Siderel Messenger*). Ventis, Apud Thoman Baglionum, 1610. English translation by Albert van Helden: Chicago, Illinois: University of Chicago Press 1989.

Galilei, G. (1632): *Dialogo Massimi Sistemi del Mondo Tolemaico, e Copernicano* (*Dialogue Concerning the Two Chief World Systems – Ptolemaic and Copernican).* Fiorenza, Batifta Landini 1632. English translation by Stillman Drake, Berkeley, California: University of California Press 1953, 1957.

Galilei, G. (1638): *Discorsi e Dimostrazioni Matematiche Intorno à Due Nuoue Scienze Mecanica & Movimenti Locali* (*Discourses and Mathematical Demonstrations Concerning Two New Sciences Pertaining to Mechanics and Local Motions)*. Leyden: At the Elzevirs 1638. English translation by Stillman Drake, Madison, Wisconsin: The University of Wisconsin Press 1974.

Gallagher, J. S., Starrfield, S. (1978): Theory and observations of classical novae. *Annual Review of Astronomy and Astrophysics* **16**, 171.

Gallart, C., Zoccali, M., Aparicio, A. (2005): The adequacy of stellar evolution models for the interpretation of color-magnitude diagrams of resolved stellar populations. *Annual Review of Astronomy and Astrophysics* **43**, 387-434.

Galloway, D. J., Weiss, N. O. (1981): Convection and magnetic fields in stars. *The Astrophysical Journal* **243**: 945-953.
Gamow, G. (1928): Zur quanten theorie der atomzertrümmerung (On the quantum theory of the atomic nucleus). *Zeitschrift fur Physik* **52**, 510.
Gamow, G. (1939): Physical possibilities of stellar evolution. *Physical Review* **55**, 718-725.
Gamow, G. (1946): Expanding universe and the origin of elements. *Physical Review* **70**, 572.
Gamow, G. (1948): The evolution of the universe. *Nature* **162**, 680.
Gamow, G. (1956): The physics of the expanding universe. *Vistas in Astronomy* **2**, 1726-1732. Reproduced in Lang and Gingerich (1979).
Gamow, G., Schönberg, M. (1941): Neutrino theory of stellar collapse. *Physical Review* **59**, 539-547. Reproduced in Lang and Gingerich (1979).
Gauss, C. F. (1809): *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium* (*Theory of Motion of Celestial Bodies Moving in Conic Sections around the Sun*.) English translation by C. H. Davis , reprinted New York: Dover 1963.
Gautschy, A., Saio, H. (1995): Stellar pulsations across the HR diagram. Part 1. *Annual Review of Astronomy and Astrophysics* **33**, 75-114.
Gautschy, A., Saio, H. (1996): Stellar pulsations across the HR diagram: Part 2. *Annual Review of Astronomy and Astrophysics* **34**, 551-606.
Gebhardt, K., Thomas, J. (2009): The black hole mass, stellar mass-to-light ratio, and dark halo in M87. *The Astrophysical Journal* **700** (2), 1690-1701.
Gehrels, N., Ramirez-Ruiz, E., Fox, D. B. (2009): Gamma-ray bursts in the *Swift* era. *Annual Review of Astronomy and Astrophysics* **47**, 567-617.
Geller, M. J., Huchra, J. P. (1989): Mapping the universe. *Science* **246** (4932), 897-903,
Genzel, R., Townes, C. H. (1987): Physical conditions, dynamics and mass distribution in the center of the Galaxy. *Annual Review of Astronomy and Astrophysics* **25**, 377.
Georgelin, Y. M., Georgelin, Y. P. (1976): The spiral structure of our Galaxy determined from H II regions. *Astronomy and Astrophysics* **49**, 57.
Ghez, A. M. et al. (2005): Stellar orbits around the galactic center black hole. *The Astrophysical Journal* **620** (2), 744-757.
Ghez, A. M. et al. (2008): Measuring distance and properties of the Milky Way's central supermassive black hole with stellar orbits. *The Astrophysical Journal* **689** (2), 1044-1062.
Giacconi, R. et al. (1962): Evidence for x-rays from sources outside the solar system. *Physical Review Letters* **9**, 439-443. Reproduced in Lang and Gingerich (1979).
Giacconi, R. et al. (1971): Discovery of periodic x-ray pulsations in Centaurus X-3 from *UHURU*. *The Astrophysical Journal* **167**, L67.
Giavalisco, M. (2002): Lyman-break galaxies. *Annual Review of Astronomy and Astrophysics* **40**, 579-641.
Gilbert, W. (1600): *De Magnete, Magneticisque Corporibus, et de Magno Magnete Tellure: Physiologia Nova, Plurimis and Argumentis, and Experimentis Demonstrata* London 1600. English translation by P. Fleury Mottelay, *William Gilbert of Colchester… on the great magnet of the earth*. Ann Arbor 1893 and Silvanus P. Thompson, reprinted from the 1900 edition by Basic Books, New York 1958.
Gilles, C., Baraffe, I. (2000): Theory of low-mass stars and substellar objects. *Annual Review of Astronomy and Astrophysics* **38**, 337-377.
Gillies, G. T. (1997): The Newtonian gravitational constant: recent measurements and related studies. *Reports on Progress in Physics* **60,** 151-225.
Gillessen, S. et al. (2009): Monitoring stellar orbits around the massive black hole in the galactic center. *The Astrophysical Journal* **692**, 1075-1109.
Gilman, P. A. (1974): Solar rotation. *Annual Review of Astronomy and Astrophysics* **12**, 47-70.
Ginzburg, V. L. (1956): The nature of cosmic radio emission and the origin of cosmic rays. *Nuovo Cimento Supplement* **3**, 38-48. Reproduced in Lang and Gingerich (1979).
Ginzburg, V. L., Syrovatskii, S. I. (1965): Cosmic magnetobremsstrahlung (Synchrotron radiation). *Annual Review of Astronomy and Astrophysics* **3**, 297.

Giovanelli, R. G. (1947): Magnetic and electric phenomena in the Sun's atmosphere associated with sunspots. *Monthly Notices of the Royal Astronomical Society* **107**, 338-355.
Giovanelli, R. G. (1949): A note on heat transfer in the upper chromosphere and corona. *Monthly Notices of the Royal Astronomical Society* **109**, 372.
Giovanelli, R., Haynes, M. P. (1991): Redshift surveys of galaxies. *Annual Review of Astronomy and Astrophysics* **29**, 499.
Gizon, L., Birch, A. C., Spruit, H. C. (2010): Local helioseismology: three-dimensional imaging of the solar interior. *Annual Review of Astronomy and Astrophysics* **48**, l289-338.
Gliese, W. (1969): *Catalogue of Nearby Stars*. Veroff-Astron. Rechen-Institut Heidelberg: Braun, Karlsruhe 1969.
Gliese, W., Jahreiss, H. (1979): Nearby star data published 1969-1978. *Astronomy and Astrophysics Supplement* **38**, 423-338.
Gold, T. (1959a): Magnetic field in the solar system. *Nuovo Cimento Supplemento* **13**, 318-323.
Gold, T. (1959b): Plasma and magnetic fields in the solar system. *Journal of Geophysical Research* **64**, 1665-1674.
Gold, T. (1968): Rotating neutron stars as the origin of the pulsating radio sources. *Nature* **218**, 73-732. Reproduced in Lang and Gingerich (1979).
Gold, T., Hoyle, F. (1960): On the origin of solar flares. *Monthly Notices of the Royal Astronomical Society* **120**, 89-105.
Goldreich, P., Kumar, P. (1990): Wave generation by turbulent convection. *The Astrophysical Journal* **363**, 694-704.
Goldreich, P., Peale, S. J. (1966): Resonant spin states in the solar system. *Nature* **209**, 1078-1079.
Goldreich, P., Tremaine, S. (1982): The dynamics of planetary rings. *Annual Review of Astronomy and Astrophysics* **20**, 249-283.
Golimowski, D. A. et al. (2006): *Hubble Space Telescope* ACS multiband coronagraphic imaging of the debris disk around Beta Pictoris. *The Astronomical Journal* **131**, issue 6, 3109-3130.
Goodricke, J. (1783): A series of observations on, and a discovery of, the period of the variation of the light of the bright star in the head of Medufa, called Algol. *Philosophical Transactions of the Royal Society of London* **73**, 474.
Goodricke, J. (1785): Observations of a new variable star. *Philosophical Transactions of the Royal Society of London* **75**, 153-164.
Gordon, M. A., Burton, W. B. (1976): Carbon monoxide in the Galaxy. I. The radial distribution of CO, H2, and nucleons. *The Astrophysical Journal* **208**, 346.
Gott, J. R., et al. (1974): An unbound universe? *The Astrophysical Journal* **194**, 543.
Gott, J. R., et al. (2005): A map of the universe. *The Astrophysical Journal* **624** (2), 463-484.
Gratton, R. G., Sneden, C., Carretta, E. (2004): Abundance variations within globular clusters. *Annual Review of Astronomy and Astrophysics* **42**, 385-440.
Greenstein, J. L., Matthews, T. A. (1963): Red-shift of the unusual radio source: 3C 48. *Nature* **197**, 1041-1042. Reproduced in Lang and Gingerich (1979).
Greenstein, J. L., Oke, J. B., Shipman, H. L. (1971): Effective temperature, radius, and gravitational redshift of Sirius B. *The Astrophysical Journal* **169**, 563.
Greenstein, J. L., Trimble, V. L. (1972): The gravitational redshift of 40 Eridani B. *Astrophysical Journal Letters* **175**, L1.
Gribov, V. N., Pontecorvo, B. M. (1969): Neutrino astronomy and lepton charge. *Physics Letters* **B28**, 493-496.
Gringauz, K. I. (1961): Some results of experiments in interplanetary space by means of charged particle traps on Soviet space probes. *Space Research* **2**, 539-553.
Grotrian, W. (1934): Über das Fraunhofersche Spektrum der Sonnenkorona. *Zeitschrift für Astrophysik* **8**, 124-146.
Grotrian, W. (1939): On the question of the significance of the lines in the spectrum of the solar corona. *Naturwissenschaften* **27**, 214. English translation in Lang and Gingerich (1979).
Gurney, R. W., Condon, E. U. (1928): Wave mechanics and radioactive disintegration. *Nature* **122**, 493.

Guth, A. H. (1981): Inflationary universe: a possible solution to the horizon and flatness problems. *Physical Review* **D23**, 347.
Hafele, J. C., Keating, R. E. (1972): Around-the-world atomic clocks – predicted relativistic time gains. *Science* **177**, 166.
Hafele, J. C., Keating, R. E. (1972): Around-the-world atomic clocks – observed relativistic time gains. *Science* **177**, 168.
Hahn, O. (1921): Über rein neues radioaktives Zerfallsprodukt im Uran. *Die Naturwissenschaftern* **9** (5), 84.
Hahn, O., Strassmann, F. (1939a): Über den Nachweis und das Verhalten der bei der Bestrahlung des Urans mittels Neutronen entstehenden Erdalkalimetalle. Concerning the existence of alkaline earth metals resulting from neutron irradiation of uranium. *Die Naturwissenschaften* **27**, 11-15. English translation in *American Journal of Physics*, January 1964, 9-15.
Hahn, O., Strassmann, F. (1939b): Proof of the formation of active isotopes of barium from uranium and thorium irradiated with neutrons; proof of the existence of more active fragments produced by uranium fission. *Die Naturwissenschaften* **27**, 89-95. English translation in *Journal of Chemical Education*, May 1989, 363.
Hale, G. E. (1908a): On the probable existence of a magnetic field in Sun-spots. *The Astrophysical Journal* **28**, 315-343. Reproduced in Lang and Gingerich (1979).
Hale, G. E. (1908b): The Zeeman effect in the Sun. *Publications of the Astronomical Society of the Pacific* **20**, 287-288.
Hale, G. E., Ellerman, F., Nicholson, S. B., Joy, A. H. (1919): The magnetic polarity of sun-spots. *The Astrophysical Journal* **49**, 153-178.
Hall, J. S. (1949): Observations of the polarized light from stars. *Science* **109**, 166-167. Reproduced in Lang and Gingerich (1979).
Halley, E. (1717): Considerations on the change of the latitudes of some of the principal fixt stars. *Proceedings of the Royal Society of London* **30**, 736-738.
Hanaoka, Y. (1996): Flares and plasma flow caused by interacting coronal loops. *Solar Physics* **165**, 275-301.
Hansen, B. M. S., Liebert, J. (2003): Cool white dwarfs. *Annual Review of Astronomy and Astrophysics* **41**, 465-515.
Hargreaves, R. (1908): Integral forms and their connection with physical equations. *Transactions of the Cambridge Philosophical Society* **21**, 107.
Harkins, W. D. (1917): The evolution of the elements and the stability of complex atoms – A new periodic system which shows a relation between the abundance of the elements and the structure of the nuclei of atoms. *Journal of the American Chemical Society* **39**, 856.
Harkins, W. D. (1931): The periodic system of atomic nuclei and the principle of regularity and continuity of series. *Physical Review* **38**, 1270.
Harris, D. E., Krawcynski, H. (2006): X-ray emission from extragalactic jets. *Annual Review of Astronomy and Astrophysics* **44**, 463-506.
Haxton, W. C., Robertson, H., Serenelli, A. M. (2013): Solar neutrinos. *Annual Review of Astronomy and Astrophysics* **51**.
Hayashi, C. (1950): Proton-neutron concentration ratio in the expanding universe at the stages preceding the formation of the elements. *Progress in Theoretical Physics (Japan)* **5**, 224.
Hayashi, C. (1961): Stellar evolution in early phases of gravitational contraction. *Publications of the Astronomical Society of Japan* **13**, 450-452. Reproduced in Lang and Gingerich (1979).
Hayashi, C. (1966): Evolution of protostars. *Annual Review of Astronomy and Astrophysics* **4**, 171.
Haymes, R. C. et al. (1968): Observation of gamma radiation from the Crab Nebula. *Astrophysical Journal Letters* **151,** L9.
Hazard, C., Mackey, M. B., Shimmins, A. J. (1963): Investigation of the radio source 3C 273 by the method of lunar occultations. *Nature* **197**, 1037-1039. Reproduced in Lang and Gingerich (1979).
Heaviside, O. (1902): The waste of energy from a moving electron. *Nature* **67**, 6.

Heaviside, O. (1904): The radiation from an electron moving in an elliptic, or any other orbit. *Nature* **69**, 342.

Heger, A. et al. (2003): How massive single stars end their life. *The Astrophysical Journal* **591**, 288-300.

Heisenberg, W. (1927): Ueber den anschaulichen Inhalt der quantentheoretischen Kinematick und Mechanik. *Zeitschrift für Physik* **43**, 172-198. English translation in Wheeler, J. A., and Zurek, W. H. (eds.) (1983): *Quantum Theory and Measurement*, Princeton, New Jersey: Princeton University Press 1983.

Heisenberg, W. (1927): Ueber die Grundprincipien der Quantenmechanik. *Forschungen und Fortschritte* **3**, 83.

Helmholtz, H. (1847): *Ueber die Erhaltung der Kraft* (The Conservation of Force). Berlin: G. Reimer 1847. English translation by John Tyndall reproduced in Brush (1948).

Helmholtz, H. (1856): On the interaction of natural forces. *Philosophical Magazine* **11** (4), 489-578.

Helmholtz, H. (1908): Observations on the Sun's store of forces. *Popular Scientific Lectures* **2**, 312-315. English translation by E. Atkinson reproduced in Meadows (1970).

Henderson, T. (1839): *Parallax of α Centauri*, communicated to the astronomical society january 9, 1839. See Clerk, A.: *A popular history of astronomy during the nineteenth century*. Edinburgh: Adam and Charles Black 1885.

Herbst, E., Klemperer, W. (1973): The formation and depletion of molecules in dense interstellar clouds. *The Astrophysical Journal* **185**, 505.

Herbst, E., Van Dishoeck, E. F. (2009): Complex organic interstellar molecules. *Annual Review of Astronomy and Astrophysics* **47**, 427-480.

Herlofson, N. (1950): Magneto-hydrodynamic waves in a compressible fluid conductor. *Nature* **165**, 1020.

Herschel, W. (1781): Account of a comet. *Philosophical Transactions of the Royal Society of London* **71**, 492-501.

Herschel, W. (1783): On the proper motion of the sun and solar system, with an account of several changes that have happened among the fixed stars since the time of Mr. Flamsteed. *Philosophical Transactions of the Royal Society of London* **73**, 247-283

Herschel, W. (1785): On the construction of the heavens. *Philosophical Transactions of the Royal Society of London* **75**, 213-266.

Herschel, W. (1786): Catalogue of one thousand new nebulae and clusters of stars. *Philosophical Transactions of the Royal Society* (*London*) **76**, 457-499.

Herschel, W. (1800): Experiments on the refrangibility of the invisible rays of the Sun. *Philosophical Transactions of the Royal Society* (*London*) **80**, 284-292.

Herschel, W. (1801): Observations tending to investigate the nature of the Sun, in order to find the causes or symptoms of its variable emission of light and heat; with remarks on the use that may possibly be drawn from solar observations. *Philosophical Transactions of the Royal Society of London* **91**, 265-318.

Hertz, H. R. (1887): Über sehr schnelle electrische Schwingungen. *Annalen der Physik* **267**, 421-448.

Hertz, H. (1889): Die Kräfte elektrischer Schwingungen, behandelt nach der Maxwell'schen Theorie (The force of electrical oscillations treated with the Maxwell theory). *Annalen der Physik* **36**, 1.

Hertz, H. (1889): Über die Beziehungen zwischen Licht und Elektrizität (On the relations between light and electricity). *Gesammelte Werke* **1**, 340.

Hertzsprung, E. (1905): Zur Strahlung der Sterne (On the radiation of stars). *Zeitshrift fur Wissenshaften Photo.* **3**, 429. English translation in Lang and Gingerich (1979).

Hertzsprung, E. (1911): *Publikationen des Astrophysikalischen Obsservatoriums zu Potsdam* **22**, No. 63.

Hess, V. F. (1912): Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten (On observations of the penetrating radiation during seven balloon flights). *Physikalishe Zeitschrift* **13**, 1084-1901. English translation in Lang and Gingerich (1979).

Hesser, J. E., *et al*. (1987): A ccd color-magnitude study of 47 Tucanae. *Publications of the Astronomical Society of the Pacific* **99**, 739.
Hester, J. J. (2008): The Crab Nebula: an astrophysical chimera. *Annual Review of Astronomy and Astrophysics* **46**, 127-155.
Hewish, A., Bell, S. J., Pilkington, J. D, H., Scott, P. F., Collins, R. A. (1968): Observation of a rapidly pulsating radio source. *Nature* **217**, 709-713. Reproduced in Lang and Gingerich (1979).
Heyl, P. R. (1930): A redetermination of the constant of gravitation. *Journal of Research National Bureau of Standards* **5**, 1243.
Hillebrandt, W., Niemeyer, J. C. (2000): Type Ia supernova explosion models. *Annual Review of Astronomy and Astrophysics* **38**, 191-230.
Hiltner, W. A. (1949): Polarization of light from distant stars by the interstellar medium. *Science* **109**, 165. Reproduced in Lang and Gingerich (1979).
Hipparchus, 125 B.C. in Ptolemy's *Almagest*. English translation by G. J. Toomer (London: Duckworth 1984).
Hirata, K. S., *et al*. (1987): Observation of a neutrino burst from the Supernova SN1987A. *Physical Review Letters* **58**, 1490.
Hirshfeld, A. (2001): *Parallax: The Race to Measure the Cosmos*. New York, W. H. Freeman and Co. 2001.
Hjellming, R. M., Churchwell, E. (1969): An analysis of radio recombination lines emitted by the Orion nebula. *Astrophysical Letters* **4**, 165.
Ho, L. C. (2008): Nuclear activity in nearby galaxies. *Annual Review of Astronomy and Astrophysics* **46**, 475-539.
Ho, P. T. P., Townes, C. H. (1983): Interstellar ammonia. *Annual Review of Astronomy and Astrophysics* **21**, 239.
Hodgson, R. (1860): On a curious appearance seen in the Sun. *Monthly Notices of the Royal Astronomical Society* **20**, 15-16. Reproduced in Meadows (1970).
Holmes, A. (1949): Lead isotopes and the age of the Earth. *Nature* **19**, 453-456.
Hoyle, F. (1946): The synthesis of elements from hydrogen. *Monthly Notices of the Royal Astronomical Society* **106**, 343.
Hoyle, F. (1948): A new model for the expanding universe. *Monthly Notices of the Royal Astronomical Society* **108**, 372.
Hoyle, F. (1954): On nuclear reactions occurring in very hot stars: I. The synthesis of elements from carbon to nickel. *Astrophysical Journal Supplement* **1**, 121.
Hoyle, F., Fowler, W. A. (1960): Nucleosynthesis in supernovae. *The Astrophysical Journal* **132**, 565.
Hoyle, F., Schwarzschild, M. (1955): On the evolution of Type II stars. *Astrophysical Journal Supplement* **2**, 1.
Hoyle, F., Tayler, R. J. (1964): The mystery of the cosmic helium abundance. *Nature* **203**, 1108.
Hu, W., Dodelson, S. (2002): Cosmic microwave background anisotropies. *Annual Review of Astronomy and Astrophysics* **40**, 171-216.
Hubble, E. P. (1925): Cepheids in spiral nebulae. *Publications of the Astronomical Society of the Pacific* **5**, 261-264., *Observatory* **48**, 139-142 (1925). Reproduced in Lang and Gingerich (1979).
Hubble, E. P. (1926): Extra-galactic nebulae. *The Astrophysical Journal* **64**, 321-369. Reproduced in Lang and Gingerich (1979).
Hubble, E. P. (1928): Novae or temporary stars. *Astronomical Society of the Pacific Leaflet* **1**, no. 14, 55-58. Reproduced in Lang and Gingerich (1979)
Hubble, E. P. (1929): A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Science* **15**, 168-173. Reproduced in Lang and Gingerich (1979).
Hubble, E. P. (1936): *The Realm of the Nebulae*. New Haven, Connecticut: Yale University Press 1936, pp. 201, 202.

Hubble, E. P. (1953): The law of red shifts. George Darwin Lecture. *Monthly Notices of the Royal Astronomical Society* **113**, 658-666.

Hubble, E. P., Humason, M. L. (1931): The velocity-distance relation among extra-galactic nebulae. *The Astrophysical Journal* **74**, 43.

Hubble, E. P., Humason, M. L. (1934): The velocity-distance relation for isolated extra-galactic nebulae. *Proceedings of the National Academy of Science* **20,** 264-276.

Huggins, W. (1864): On the spectra of some nebulae. *Philosophical Transactions of the Royal Society of London* **154**, 437-444. Supplement to the paper: On the spectra of some of the fixed stars by Huggins, W. and Miller, W. A.

Huggins, W. (1868): Further observations on the spectra of some of the stars and nebulae, with an attempt to determine therefrom whether these bodies are moving towards or from the Earth. *Philosophical Transactions of the Royal Society of London* **158,** 529-564.

Hughes, S. A. (2009): Gravitational waves from merging compact binaries. *Annual Review of Astronomy and Astrophysics* **47**, 107-157.

Hulse, R. A. (1994): The discovery of the binary pulsar. *Reviews of Modern Physics* **66**, 699.

Hulse, R. A., Taylor, J. H. (1975): Discovery of a pulsar in a binary system. *Astrophysical Journal Letters* **195**, L5-L53..

Hulst, H. C. van de (1945): Radiogloven mit hat wereldrium: I. Ontvangst der radiogloven; II. Herkomst der radiogloven (Radio waves from space: I. Reception of radiowaves; II. Origin of radiowaves). *Ned. Tijdsch. Natuurk* **11**, 201. English translation in Lang and Gingerich (1979).

Hulst, H. C. van de (1949): The solid particles in interstellar space. *Rech. Astr. Obs. Utrecht* **11**, part 2. Reproduced in Lang and Gingerich (1979),.

Humason, M. L., Mayall, N. U., Sandage, A. R. (1956): Redshifts and magnitudes of extra-galactic nebulae. *Astronomical Journal* **61**, 97-162. Reproduced in Lang and Gingerich (1979).

Hunten, D. M. (1993): Atmospheric evolution of the terrestrial planets. *Science* **259** (5097), 915-920.

Hurford, G. J., et al. (2006): Gamma-ray imaging of the 2003 October/November solar flares. *Astrophysical Journal (Letters)* **644**, L93-L96.

Iben, I. Jr., Renzini, A. (1983): Asymptotic giant branch evolution and beyond. *Annual Review of Astronomy and Astrophysics* **21**, 271-342.

Ivezic, Z., Beers, T. C., and Evans, N. J. (2012): Galactic stellar populations in the era of SDSS and other large surveys. *Annual Review of Astronomy and Astrophysics* **50**, 251-304.

Jackson, J. (1956): The distances of the stars: A historical review. *Vistas in Astronomy* **2**, 1018.

Jaki, S. L. (1969): *The Paradox of Olbers Paradox*. New York: Herder and Herder 1969.

Jansky, K. G. (1933a): Electrical disturbances apparently of extraterrestrial origin. *Proceedings of the Institute of Radio Engineers* **21**, 1387-1398.

Jansky, K. G. (1933b): Radio waves from outside the solar system. *Nature* **132**, 66.

Jansky, K. G. (1935): A note on the source of interstellar interference. *Proceedings of the Institute of Radio Engineers* **23**, 158-163. Reproduced in Lang and Gingerich (1979).

Janssen, P. J. C. (1868): Éclipse de soleil du 18 Aout 1868. *Annales de Chimie et de Physique* **15**, 414-426. Paris: G. Masson 1878.

Janssen, P. J. C. (1868): *Summary of some of the Results Obtained at Cocanada, during the Eclipse last August.* Letter to the French Academy of Sciences reproduced in Meadows (1970).

Janssen, P. J. C. (1872): Observations of the solar eclipse of 12 December 1871. *Nature* **5**, 249. Reproduced in Meadows (1970).

Jarosik, N. et al. (2011): Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: sky maps, systematic errors, and basic results. *Astrophysical Journal Supplement* **192**, 14.

Jeans, J. H. (1902): On the stability of a spherical nebula. *Philosophical Transactions of the Royal Society of London* **199**, 1-53. Reproduced in Lang and Gingerich (1979).

Jeans, J. H. (1905): On the partition of energy between matter and aether. *Philosophical Magazine* **10**, 91-98.
Jeans, J. H. (1909): Temperature-radiation and the partition of energy in continuous media. *Philosophical Magazine* **17**, 229-254.
Jeans, J. H. (1916, 1925): *The Dynamical Theory of Gases*. London: Cambridge University Press 1916, 1925.
Jeans, J. H. (1917): The motion of tidally-distorted masses, with special reference to theories of cosmogony. *Memoirs of the Royal Astronomical Society London* **62**, 1-48.
Jeffreys, H. (1920): Tidal friction in the shallow seas. *Philosophical Transactions of the Royal Society of London, Series A* **221**, 239.
Jeffreys, H. (1926): The stability of a layer of fluid heated below. *Philosophical Magazine* **2**, 833.
Jeffreys, H. (1975): Tidal friction. *Quarterly Journal of the Royal Astronomical Society* **16**, 145.
Joss, P. C., Rappaport, S. A. (1984): Neutron stars in interacting binary systems. *Annual Review of Astronomy and Astrophysics* **22**, 537-592.
Joule, J. D. (1847): On matter, living force, and heat. In *The Manchester Courier* 5 and 12 May, 1847. Reproduced in Brush (1948).
Joule, J. D. (1850): On the mechanical equivalent of heat. *Philosophical Transactions of the Royal Society of London* **140**, 61-82.
Joy, A. H. (1939): Rotation effects, interstellar absorption, and certain dynamical constants of the Galaxy determined from Cepheid variables. *The Astrophysical Journal* **89**, 356. Reproduced in Lang and Gingerich (1979).
Joy, A. H. (1954): Variable stars of low luminosity. *Publications of the Astronomical Society of the Pacific* **16**, 230.
Joy, A. H. (1956): Radial-velocity measures of SS Cygni at minimum light. *The Astrophysical Journal* **124**, 317.
Kalas, P. et al. (2005): A planetary system as the origin of structure in Fomalhaut's dust belt. *Nature* **435** (7045), 1067-1070.
Kalas, P. et al. (2008): Optical images of an exosolar planet 25 light-years from Earth. *Science* **322** (5906), 1345-1348.
Kalberla, P. M. W., Kerp, J. (2009): The HI distribution of the Milky Way. *Annual Review of Astronomy and Astrophysics* **47**, 27-61.
Kaler, J. B. (1985): Planetary nebulae and their central stars. *Annual Review of Astronomy and Astrophysics* **23**, 89-117.
Kamionkowski, M., Bahcall, J. N. (1994): The rate of the proton-proton reaction. *The Astrophysical Journal* **420**, 884.
Kant, I. (1755): *Universal Natural History and Theory of the Heavens: An Essay on the Constitution and Mechanical Origin of the Whole Universe Treated According to Newton's Principles*. English translation of German original by W. Hastie reproduced by Ann Arbor: University of Michigan Press 1969.
Kaplan, G. H. (2005): The IAU resolutions on astronomical reference systems, time scales, and Earth rotation models. *U. S. Naval Observatory Circular No. 179*, 1-104.
Kaplan, S. A., Pikelner, S. B. (1970): *The Interstellar Medium*. Cambridge, Massachusetts: Harvard University Press 1970.
Kapteyn, J. C. (1898): Stern mit grosster bislang bekannter Eigenbewegung. *Astronomische Nachrichten* **145** (9-10), 159-160.
Kapteyn, J. C. (1905): Star streaming. *Report of the British Association for the Advancement of Science* 237-265. Reproduced in Lang and Gingerich (1979).
Kapteyn, J. C. (1922): First attempt at a theory of the arrangement and motion of the sidereal system. *The Astrophysical Journal* **55**, 302-327. Reproduced in Lang and Gingerich (1979).
Keeling, C. D. (1960): The concentration and isotopic abundances of carbon dioxide in the atmosphere. *Tellus* **12**, Issue 2, 200-203.
Keeling, C. D. (1978): Atmospheric carbon dioxide in the nineteenth century. *Science* **202** (4372), 1109.

Keeling, C. D. (1997): Climate change and carbon dioxide: an introduction. *Proceedings of the National Academy of Sciences* **94** (16), 8273-8274.
Kellermann, K. I., Pauliny-Toth, I. I. K. (1981): Compact radio sources. *Annual Review of Astronomy and Astrophysics* **19**, 373-410.
Kelvin, Lord (W. Thomson) (1848): On an absolute thermometric scale founded on Carnot's theory of the motive power of heat. *Proceedings of the Cambridge Philosophical Society* **1**, 66.
Kelvin, Lord (W. Thomson) (1862): Physical considerations regarding the possible age of the Sun's heat. *Philosophical Magazine* **23**, 158.
Kelvin, Lord (W. Thomson) (1863): On the secular cooling of the Earth. *Philosophical Magazine* **25**, 1-14.
Kelvin, Lord (W. Thomson) (1899): The age of the Earth as an abode fitted for life. *Journal of the Victoria Institute* **31**, 11-35.
Kemp, J. C., Swedlund, J. B., Landstreet, J. D., Angel, J. R. P. (1970): Discovery of circularly polarized light from a white dwarf star. *Astrophysical Journal Letters* **161**, L77-L79. Reproduced in Lang and Gingerich (1979).
Kennedy, R. J. (1926): A refinement of the Michelson-Morley experiment. *Proceedings of the National Academy of Sciences* **12**, 621-629.
Kennedy, R. J., Thorndike, E. M. (1932): Experimental establishment of the relativity of time. *Physical Review* **42**, 400.
Kennelly, A. E. (1902): On the elevation of the electrically-conducting strata in the Earth's atmosphere. *Electrical World and Engineer* **39**, 473.
Kennicutt, R. C. Jr. (1998): Star formation in galaxies along the Hubble sequence. *Annual Review of Astronomy and Astrophysics* **36**, 189-231.
Kennicutt, R. C. Jr., Evans, N. J. (2012): Star formation in the Milky Way and nearby galaxies. *Annual Review of Astronomy and Astrophysics* **50**, 531-608.
Kepler, J. (1609): *Astronomia Nova*. Full title in English: *New Astronomy, Based upon Causes, or Celestial Physics, Treated by Means of Commentaries on the Motions of the Star Mars, From the Observations of Tycho Brahe*. English translation by Donahue, W. H., *New Astronomy*. Cambridge, England: Cambridge University Press 1992.
Kepler, J. (1619): *De Harmonice Monde* (*The Harmony of the World*). Lintz, Austria: Johann Blancken 1619.
Kerr, F. J. (1969): The large-scale distribution of hydrogen in the Galaxy. *Annual Review of Astronomy and Astrophysics* **7**, 39-66.
Kerr, F. J., Lynden-Bell, D. (1986): Review of galactic constants. *Monthly Notices of the Royal Astronomical Society* **221**, 1023.
Kippenhahn, R., Weigert, A., Achim, W. (2012): *Stellar Structure and Evolution, Second Edition*. Heidelberg: Springer Verlag 2012.
Kirchhoff, G. R. (1860): On the simultaneous emission and absorption of rays of the same definite refrangibility. *Philosophical Magazine* **19**, 193.
Kirchhoff, G. R. (1861a): On a new proposition in the theory of heat. *Philosophical Magazine* **21** (4), 241-247. Reproduced in Meadows (1970).
Kirchhoff, G. R. (1861b): On the chemical analysis of the solar atmosphere. *Philosophical Magazine* **21** (4), 185-188. Reproduced in Meadows (1970).
Kirchhoff, G. R., Bunsen, R. (1860): Chemical analysis by observation of spectra. *Annalen der Physik und der Chemie* **110**, 161-189.
Kirchhoff, G. R., Bunsen, R. (1860/61): Chemical analysis of spectrum – observations. *Philosophical Magazine and Journal of Science* **20**, 89-109, **22**, 329-249, 498-510.
Kitchin, C. R. (2013): *Telescopes and Techniques*. New York: Springer 2013.
Klebesadel, R. W., Strong, I. B., Olson, R. A. (1973): Observations of gamma-ray bursts of cosmic origin. *Astrophysical Journal Letters* **182**, L85-L88.
Kley, W., Nelson, R. P. (2012): Planet-disk interaction and orbital evolution. *Annual Review of Astronomy and Astrophysics* **50**, 211-249.

Knop, R. A. et al. (2003): New constaraits on omega matter omega lambda and w from an independent set of high-redshift supernovae observed with the *Hubble Space Telescope*. *The Astrophysical Journal* **598**, 102-137.
Komatsu, E. et al. (2011): Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological interpretation. *Astrophysical Journal Supplement* **192**, 18.
Kopp, G., Lawrence, G., Rottman, G. (2005): The total irradiance monitor (TIM): Science results. *Solar Physics* **230**, 129-139.
Kopp, R. A., Pneuman, G. W. (1976): Magnetic reconnection in the corona and the loop prominence phenomenon. *Solar Physics* **50**, 85-98.
Kormendy, J. (1988): Evidence for a central dark mass in NGC 4594 (the sombrero galaxy). *The Astrophysical Journal* **335**, 40.
Kormendy, J. (1988): Evidence for a supermassive black hole in the nucleus of M 31. *The Astrophysical Journal* **325**, 128.
Kormendy, J. (2013): Supermassive black holes in the *HST* era. *Annual review of Astronomy and Astrophysics* **51**.
Kormendy, J., Richstone, D. (1995): Inward bound – the search for supermassive black holes in galactic nuclei. *Annual Review of Astronomy and Astrophysics* **33**, 581-624.
Kotera, K., Olinto, A. V. (2011): The astrophysics of ultrahigh-energy cosmic rays. *Annual Review of Astronomy and Astrophysics* **49**, 119-153.
Kowalski, M. et al. (2008): Improved cosmological constraints from new, old and combined supernova datasets. *The Astrophysical Journal* **686**, 749-778.
Kraft, R. P. (1964): Binary stars among cataclysmic variables III. Ten old novae. *The Astrophysical Journal* **139**, 457. Reproduced in Lang and Gingerich (1979).
Kragh, H. (1997): *Cosmology and Controversy: The Historical Development of Two Theories of the Universe*. Princeton New Jersey: Princeton Univesity Press 1992.
Kragh, H. (2002): *Quantum Generations: A History of Physics in the Twentieth Century*. Princeton, New Jersey, Princeton Univeristy Press 2002.
Kramer, M. et al. (2006): Tests of general relativity from timing the double pulsar. *Science* **314** (5796), 97-102.
Kramer, M., Stairs, I. H. (2008): The double pulsar. *Annual Review of Astronomy and Astrophysics* **46**, 541-572.
Kramers, H. A. (1924): The law of dispersion and Bohr's theory of spectra. *Nature* **113**, 783.
Kravtsov, A. V., Borgani, S. (2012): Formation of galaxy clusters. *Annual Review of Astronomy and Astrophysics* **50**, 353-409.
Kronig, A. K. (1856): Grundziigc einer Theorie der Gase (A general theory of gases). *Annalen der Physik* **99**, 315-322.
Kudritzki, R.-P., Puls, J. (2000): Winds from hot stars. *Annual Review of Astronomy and Astrophysics* **38**, 613-666.
Kulessa, A. S., Lynden-Bell, D. (1992): The mass of the Milky Way Galaxy. *Monthly Notices of the Royal Astronomical Society* **255**, 105-118.
Kulkarni, S. R., et al. (1998): Identification of a host galaxy at redshft z = 3.42 for the gamma ray burst of 14 December 1997. *Nature* **393**, 35-39.
Kulsrud, R. M., Ostriker, J. P., Gunn, J. E. (1972): Acceleration of cosmic rays in supernova remnants. *Physical Review Letters* **28**, 636.
Lada, C. J., Lada, E. A. (2003): Embedded clusters in molecular clouds. *Annual Review of Astronomy and Astrophysics* **41**, 57-115.
Lambeck, K. (1978): Tidal dissipation in the oceans - astronomical, geophysical and oceanographic consequences. *Philosophical Transactions of the Royal Society of London, Series A* **287**, 545.
Lambeck, K. (1980): *The Earth's Variable Rotation*. Cambridge, England: Cambridge University Press 1980.
Lammer, H. et al. (2008): Atmospheric escape and evolution of terrestrial planets and satellites. *Space Science Reviews* **139,** 399-436.

Lampland, C. O. (1921): Observed changes in the structure of the "Crab" Nebula (N.G.C. 1952). *Publications of the Astronomical Society of the Pacific* **33**, 79.
Landau, L. (1938): Origin of stellar energy. *Nature* **141**, 333-334.
Lane, J. H. (1870): On the theoretical temperature of the sun; under the hypothesis of a gaseous mass maintaining its volume by its internal heat, and depending on the laws of gases as known to terrestrial experiment. *American Journal of Science and Arts* (2^{nd} series) **50**, 57-74. Reproduced in Meadows (1970).
Lang, K. R. (1978, 1992): *Astrophysical Formulae. First, Second Editions*. New York: Springer 1978, 1992.
Lang, K. R. (1985): The uncertainties of space and time. *Vistas in Astronomy* **28**, 277-288.
Lang, K. R. (1999): *Astrophysical Formulae, Third Edition*. Heidelberg: Springer Verlag 1999.
Lang, K. R. (2009): *The Sun From Space, Second Edition*. Heidelberg: Springer Verlag 2009.
Lang, K. R. et al. (1975): The composite Hubble diagram. *The Astrophysical Journal* **202**, 583.
Lang, K. R., Gingerich, O. (1979): *A Source Book in Astronomy and Astrophysics 1900-1975*. Cambridge, Massachusetts: Harvard University Press 1979.
Langer, N. (2012): Pre-supernova evolution of massive single and binary stars. *Annual Review of Astronomy and Astrophysics* **50**, 107-164.
Laplace, P. S. Marquis de (1796): *Exposition du Systeme du Monde* (Cercie-Social: Paris, 1796). *Exposition of the System of the World*. Also see English translation of Laplace's proof of the existence of black holes in S. W. Hawking and G. F. R. Ellis: *The large scale structure of space-time*. Cambridge, England: Cambridge University Press 1973.
Laplace, P. S. Marquis de (1816): Sur la vitesse du son dans l'air et dan l'eau (On the velocity of sound in the air and the water). *Ann. Chem. Phys.* **3**, 238.
Lapparent, V. de, Geller, M. J., Huchra, J. P. (1986): A slice of the universe. *Astrophysical Journal Letters* **302**, L1.
Large, M. I., Vaughan, A. E., Mills, B. Y. (168). A pulsar supernova association? *Nature* **220**, 340-341.
Larmor, J. (1897): On the theory of the magnetic influence on spectra; and on the radiation from moving ions. *Philosophical Magazine* **44**, 503.
Lattes, C. M. G., Occhialini, G. P. S., Powell, C. F. (1947): Observations on the tracks of slow mesons in photographic emulsions. *Nature* **160**, 453-456, 492.
Lawrence, E. O., Cooksey, D. (1936): On the apparatus for the multiple acceleration of light ions to high speed. *Physical Review* **50**, 1131-1140.
Lawrence, E. O., Livingston, M. S. (1932): The production of high speed light ions without the use of high voltages. *Physical Review* **40**, 19-35.
Lawrence, E. O., Livingston, M. S. (1934): The multiple acceleration of ions to very high speeds. *Physical Review* **45**, 608-612.
Lawrence, E. O., Livingston, S., White, M. C. (1932): The disintegration of lithium by swiftly moving protons. *Physical Review* **40,** 150-151.
Leavitt, H. S. (1908): 1777 variables in the Magellanic Clouds. *Annals of the Harvard College Observatory* **60**, 87-108.
Leavitt, H. S. (1912): Periods of twenty-five variable stars in the Small Magellanic Cloud. *Harvard College Observatory Circular No. 173*, 1-3. Reproduced in Lang and Gingerich (1979).
Lebreton, Y. (2000): Stellar structure and evolution: deductions from *HIPPARCOS*. *Annual Review of Astronomy and Astrophysics* **38**, 35-77.
Ledoux, P., Renson, P. (1966): Magnetic stars. *Annual Review of Astronomy and Astrophysics* **4**, 293-352.
Lehnert, M. D. et al. (2010): Spectroscopic confirmation of a galaxy at redshift $z = 8.6$. *Nature* **467**, 940-942.
Leibacher, J. W., Stein, R. F. (1971): A new description of the solar five-minute oscillation. *Astrophysical Letters* **7**, 191-192.
Leibundgut, B. (2001): Cosmological implications from observations of Type Ia supernovae. *Annual Review of Astronomy and Astrophysics* **39**, 67-98.

Leighton, R. B. (1961): Considerations on localized velocity fields in stellar atmospheres: Prototype – The solar atmosphere. In: *Aerodynamic Phenomena in Stellar Atmospheres. Proceedings of the Fourth Symposium on Cosmical Gas Dynamics*. *Supplemento del Nuovo Cimento* **22**, 321-325.

Leighton, R. B. (1963): The solar granulation. *Annual Review of Astronomy and Astrophysics* **1**, 19-40.

Leighton, R. B. (1969): A magneto-kinematic model or the solar cycle. *The Astrophysical Journal* **156**, 1-26.

Leighton, R. B., Noyes, R. W., Simon, G. W. (1962): Velocity fields in the solar atmosphere I. Preliminary report. *The Astrophysical Journal* **135**, 474-499.

Lemaître, G. (1927): Un univers homogene de masse constante et de rayon croissant, redant compte de la vitesse radiale des nebuleuses extra-galactiques (A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extragalactic nebulae*). Annales de la Société de Bruxelles* **A47**, 49. English translation in *Monthly Notices of the Royal Astronomical Society* **91**, 483 (1931) and reproduced in Lang and Gingerich (1979).

Lemaître, G. (1931a): A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extragalactic nebulae*). Monthly Notices of the Royal Astronomical Society* **91**, 483 (1931), reproduced in Lang and Gingerich (1979).

Lemaître, G. (1931b): L'expansion de l'espace. *La Revue des Questions Scientifiques*, *4e Série*, November 1931. English translation in Lemaître (1950).

Lemaître, G. (1931c): The beginning of the world from the point of view of quantum theory. *Nature* **127** (3210), 706.

Lemaître, G. (1934): Evolution of the expanding universe. *Proceedings of the National Academy of Science* **20**, 12.

Lemaître, G. (1950): *The Primeval Atom: An Essay on Cosmogony*. New York: D. Van Nostrand Co. 1950.

Leverrier, U. J. J. (1859) : Lettre de M. Le Verrier à M. Faye sur la théorie de Mercure et sur le mouvement du périhélie de cette planète. *Compte Rendus de l'Académie des sciences (Paris)* **49**, 379-383; Theorie du Mouvement de Mercure. *Annales de l'Observatoire de Paris* **5**.

Lewis, J. S. (1974): The temperature gradient in the solar nebula. *Science* **186** (4162), 440-443.

Lewis, J. S. (2004): *Physics and Chemistry of the Solar System*. Burlington, Massachusetts: Elsevier Academic Press 2004.

Liebert, J. (1980): White dwarf stars. *Annual Review of Astronomy and Astrophysics* **18**, 363-398.

Lifshitz, E. M. (1946): On the gravitational stability of the expanding universe. *J. Phys. USSR (Zhurn. Erksp. Theor. Fiz)* **10**, 116.

Lighthill, M. J. (1952): On sound generated aerodynamically: I. General theory. *Proceedings of the Royal Society London* **A211**, 564.

Lighthill, M. J. (1954): On sound generated aerodynamically: II. Turbulence as a source of sound. *Proceedings of the Royal Society London* **A222**, 1.

Lin, C. C., Shu, F. H. (1967): Density waves in disk galaxies. *International Astronomical Union Symposium No. 31*, 313-317. Reproduced in Lang and Gingerich (1979).

Lindblad, B. (1925): Star-streaming and the structure of the stellar system. *Arkiv för matematik, astronomi och fysik* **19A**, no. 21, 1-8. Reproduced in Lang and Gingerich (1979).

Linde, A. D. (1982): A new inflationary universe scenario – a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters* **108B**, 389.

Lockyer, J. N. (1866): Spectroscopic observations of the Sun. *Proceedings of the Royal Society* **15**, 256-258. Reproduced in Meadows (1970).

Lockyer, J. N. (1869): Spectroscopic observations of the Sun. III, IV. *Proceedings of the Royal Society* **17**, 350-356, 415-418. Reproduced in Meadows (1970).

Lockyer, J. N. (1887): *The Chemistry of the Sun*. London: Macmillan and Co. 1887.

Lodge, O. (1919): Gravitation and light. *Nature* **104**, 354.

Loeb, A., Barkana, R. (2001): The reionization of the universe by the first stars and quasars. *Annual Review of Astronomy and Astrophysics* **39**, 29-66.
Longair, M. S. (2011): *High Energy Astrophysics, Third Edition*. Cambridge, England: Cambridge University Press 2011.
Lorentz, H. A. (1898): Influence du champ magnétique sur l'emission lumineuse (On the influence of magnetic forces on light emission). *Rev. d'electricité* **14**, 435 (1898) and also *Annalen der Physik* **63**, 278 (1897).
Lorentz, H. A. (1904): Electromagnetic phenomena in a system moving with any velocity smaller than that of light. *Proc. Am. Acad. Sci.* **6**, 809. Reproduced in: *The principle of relativity* (ed. A. Sommerfeld). New York: Dover 1952.
Loveday, J. et al. (1992): The stromlo-apm redshift survey I. the luminosity function and space density of galaxies. *The Astrophysical Journal* **390**, 338-344.
Lu, F. J. et al. (2011): The single-degenerate binary origin of Tycho's supernova as traced by the stripped envelope of the companion. *The Astrophysical Journal* **732** (1), 11.
Lucretius, T. (55 B.C.): *De Rerum Natura* (*On the Nature of the Universe*). English translation by R. E. Latham, London: Penguin Books 1951.
Luhman, K. L. (2012): The formation and early evolution of low-mass stars and brown dwarfs. *Annual Review of Astronomy and Astrophysics* **50**, 65-106.
Luther, G. G., Towler, W. R. (1982): Redetermination of the Newtonian gravitational constant G. *Physical Review Letters* **48**, 121.
Lyman, T. (1906): The spectrum of hydrogen in the region of extremely short wave-length. *The Astrophysical Journal* **23**, 181.
Lynden-Bell, D. (1969): Galactic nuclei as collapsed old quasars. *Nature* **223**, 690.
Lynds, R., Petrosian, V. (1989): Luminous arcs in clusters of galaxies. *The Astrophysical Journal* **336**, 1.
Lyne, A. (2004): A double-pulsar system: a rare laboratory for relativistic gravity and plasma physics *Science* **303** (5661), 1153-1157.
Lyne, A., Graham-Smith, F. (2012): *Pulsar Astronomy, Fourth Edition*. Cambridge, England: Cambridge University Press 2012.
Mac Dougal, D. W. (2013): *Newton's Gravity: A Student's Guide to the Mechanics of the Universe*. New York: Springer 2013.
Macchetto, F. et al. (1997): The supermassive black hole of M 87 and the kinematics of its associated gaseous disk. *The Astrophyical Journal* **489** (2), 579.
Maclaurin, C. (1742): A treatise on fluxions. (1742) Cf.: *History of the Mathematical Theories of Attraction, and the Figure of the Earth* by I. Todhunter. Macmillan 1873. Reproduced New York: Dover 1962.
Madore, B. F., Freedman, W. C. (1991): The Cepheid distance scale. *Publications of the Astronomical Society of the Pacific* **103**, 933.
Maeder, A., Meynet, G. (2000): The evolution of rotating stars. *Annual Review of Astronomy and Astrophysics* **38**, 143-190.
Majaess, D. (2010): Concerning the distance to the center of the Milky Way and its structure. *Acta Astronomica* **60**, 55-74.
Majaess, D., Turner, D., Gieren, W. (2012): New evidence supporting cluster membership for the keystone calibrator Delta Cephei. *The Astrophysical Journal* **747**, 145.
Mann, I. (2010): Interstellar dust in the solar system. *Annual Review of Astronomy and Astrophysics* **48**, 173-203.
Marcy, G. W., Butler, R. P. (1996): A planetary companion to 70 Virginis. *Astrophysical Journal Letters* **464**, L147.
Marcy, G. W., Butler, R. P. (1998): Detection of extrasolar giant planets. *Annual Review of Astronomy and Astrophysics* **36**, 57-98.
Marois, C. et al. (2008): Direct imaging of multiple planets orbiting the star HR 8799. *Science* **322** (5906), 1348-1352.
Mateo, M. (1998): Dwarf galaxies of the local group. *Annual Review of Astronomy and Astrophysics* **36**, 435-506.

Mather, J. C., *et al*. (1994): Measurement of the cosmic microwave background spectrum by the *COBE* FIRAS instrument. *The Astrophysical Journal* **420**, 439-512.
Mathewson, D. S., Ford, V. L. (1970): Polarization observations of 1800 stars. *Memoirs of the Royal Astronomical Society* **74**, 139.
Mathis, J. S. (1990): Interstellar dust and extinction. *Annual Review of Astronomy and Astrophysics* **28**, 37.
Maxwell, J. C. (1860): Illustrations of the dynamical theory of gases: Part I. On the motions and collisions of perfectly elastic spheres; Part II. On the process of diffusion of two or more kinds of moving particles among one another. *Philosophical Magazine* **19**, 19-32, **20**, 21-37. Reproduced in Brush (1948).
Maxwell, J. C. (1865): A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London* **155**, 459-512.
Mayer, R. (1842): The forces of inorganic nature. *Annalen der Chemie und Pharmacie*. English translation in *Philosophical Magazine* **24**, 371 (1862), reproduced in Brush (1948).
Mayor, M., Queloz, D. (1995): A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355-359.
Mazzullo, S. J. (1971): Length of year during the Silurian and Devonian periods: new values. *Geological Society of America Bulletin* **82**, 1085-1086.
McComas, D. J., et al. (2000): Solar wind observations over *Ulysses*' first full polar orbit. *Journal of Geophysical Research* **105**, A5, 10419-10434.
McComas, D. J., et al. (2003): The three-dimensional solar wind around solar maximum. *Geophysical Research Letters* **30**, No. 10, 1517.
McConnell, N. J. et al. (2011): Two ten-billion-solar-mass black holes at the centres of giant elliptical galaxies. *Nature* **480,** 215-218.
McCray, R. (1993): Supernova 1987A revisited. *Annual Review of Astronomy and Astrophysics* **31**, 175-216.
McCray, R., Snow, T. P. Jr. (1979): The violent interstellar medium. *Annual Review of Astronomy and Astrophysics* **17**, 213-240.
McCrea, W. H. (1929): The hydrogen chromosphere. *Monthly Notices of the Royal Astronomical Society* **89**, 483-497.
McDonald, A. B. (2005): Sudbury Neutrino Observatory results. *Physica Scripta* **T121**, 29-32.
McKee, C. F., Hollenbach, D. J. (1980): Interstellar shock waves. *Annual Review of Astronomy and Astrophysics* **18**, 219-262.
McKee, C. F., Ostricker, E. C. (2007): Theory of star formation. *Annual Review of Astronomy and Astrophysics* **45**, 565-687.
McWilliam, A. (1997): Abundance ratios and galactic chemical evolution. *Annual Review of Astronomy and Astrophysics* **35**, 503-556.
Meadows, A. J. (1970): *Early Solar Physics*. Oxford, England: Pergamon Press 1970.
Meitner, L., Frisch, O. R. (1939): Disintegration of uranium by neutrons: a new type of nuclear reaction. *Nature* **143**, 239-240.
Melia, F., Falcke, H. (2001): The supermassive black hole at the galactic center. *Annual Review of Astronomy and Astrophysics* **39**, 309-352.
Menzel, D. H. (1926): The planetary nebulae. *Publications of the Astronomical Society of the Pacific* **38**, 295. Reproduced in Lang and Gingerich (1979).
Menzel, D. H. (1962): *Selected Papers on Physical Processes in Ionized Plasmas*. New York: Dover Publications, 1962.
Menzel, D. H., Pekeris, C. L. (1935): Absorption coefficients and hydrogen line intensities. *Monthly Notices of the Royal Astronomical Society* **96**, 77 (1935). Reproduced in: *Selected Papers on Physical Processes in Ionized Plasmas* (ed. D. H. Menzel). New York: Dover 1962.
Mermilliod, J. C. (1981): Comparative studies of young open clusters III. Empirical isochronous curves and the zero age main sequence. *Astronomy and Astrophysics* **97**, 235.
Messier, C. (1781): Catalogue des nébuleuses et des amas d'étoiles, que l'on decouvre parmi les etoiles fixes. *Mem. Mat. Phys. Acad. des Sci.* for 1771. Final form: Connoissance des Temps

for 1784, published in 1781. English translation by K. G. Jones in *J. British Astron. Ass.* **79**, 357 (1969).

Mestel, L. (1968): Magnetic braking by a stellar wind − I. *Monthly Notices of the Royal Astronomical Society* **138**, 359.

Mészáros, P. (2002): Theories of gamma-ray bursts. *Annual Review of Astronomy and Astrophysics* **40**, 137-169.

Metzger, M. R., et al. (1997): Spectral constraints on the redshift of the optical counterpart to the gamma ray burst of 8 May 1997. *Nature* **387**, 878-880.

Meyer, B. S. (1994): The r-, s-, and p-processes in nucleosynthesis. *Annual Review of Astronomy and Astrophysics* **32**, 153.

Meyer, P., Parker, E. N., Simpson, J. A. (1956): Solar cosmic rays of February, 1956 and their propagation through interplanetary space. *Physical Review* **104**, 768-783.

Michell, J. (1784): On the means of discovering the distance, magnitude, etc., of the fixed stars, in consequence of the diminution of their light, in case such a diminution should be found to take place in any of them, and such other data should be procured from observations, as would be further necessary for that purpose. *Philosophical Transactions of the Royal Society of London* **74**, 35. Reproduced in: Black holes: Selected reprints (ed. S. Detweiler). Stony Brook, New York: Am. Assoc. of Physics Teachers 1982.

Michelson, A. A. (1881): The relative motion of the Earth and the luminiferous ether. *American Journal of Science* **22**, 120-129.

Michelson, A. A. (1890): On the application of interference methods to astronomical measurements. *Philosophical Magazine* **30**, 1-21.

Michelson, A. A., Morley, E. W. (1887): On the relative motion of the earth and the luminiferous ether. *American Journal of Science* **34**, 333-345

Michelson, A. A., Pease, F. G. (1921): Measurement of the diameter of alpha Orionis with the interferometer. *The Astrophysical Journal* **53**, 249-259.

Mikheyev, S. P., Smirnov, A. Y. (1985): Resonance enhancement of oscillations in matter and solar neutrino spectroscopy. *Soviet Journal of Nuclear Physics* **42**, 913-917. *Il Nuovo Cimento* **9C**, 17-26 (1986).

Miley, G. K. (1980): The structure of extended extragalactic radio sources. *Annual Review of Astronomy and Astrophysics* **18**, 165-218.

Millikan, R. A. (1910): A new modification of the cloud method of determining the elementary electrical charge and the most probable value of that charge. *Philosophical Magazine* **19**, 209-228.

Millikan, R. A. (1913): On the elementary electric charge and the Avogadro constant. *Physical Review*, Series II, **2**, 109-143.

Millikan, R. A. (1916): A direct photoelectric determination of Planck's "h". *Physical Review* **7** (3), 366-388.

Millikan, R. A. (1926): High frequency rays of cosmic origin. *Proceedings of the National Academy of Sciences* **12**, 48-55.

Millikan, R. A., Cameron, G. H. (1926): High frequency rays of cosmic origin III. Measurements in snow-fed lakes at high altitudes. *Physical Review* **28**, 851-868.

Milne, E. A. (1921): Radiative equilibrium in the outer layers of a star: The temperature distribution and the law of darkening. *Monthly Notices of the Royal Astronomical Society* **81**, 361.

Milne, E. A. (1930): Thermodynamics of stars. *Handbuch der Astrophysik* **3**, 80. Reproduced in *Selected Papers on the Transfer of Radiation* (ed. D. H. Menzel). New York: Dover 1966.

Milne, E. A. (1935): *Relativity, Gravitation and World Structure*. Oxford: Oxford at the Clarendon Press 1935.

Minkowski, H. (1908): Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körpern (The basic equations for the electromagnetic events in moving bodies). *Gott. Nach.* **1**, 53.

Minkowski, H. (1908): *Space and Time*, an address delivered to the German Natural Scientists and Physicians, translated and reprinted in: *The principle of relativity* (ed. A. Sommerfeld). New York: Dover 1952.

Minkowski, R. (1941): Spectra of supernovae. *Publications of the Astronomical Society of the Pacific* **53**, 224-225. Reproduced in Lang and Gingerich (1979).

Minkowski, R. (1942): The Crab Nebula. *The Astrophysical Journal* **96**, 199-213 (1942). Reproduced in Lang and Gingerich (1979).

Minnaert, M., Mulders, G. (1930): Intensity measurement of the Fraunhofer lines in the wavelength region 5,150 to 5,270 A. *Zeitschrift fur Astrophysik* **1,** 192-199. English translation in Lang and Gingerich (1979).

Minton, D. A., Malhotra, R. (2007): Assessing the massive young Sun hypothesis to solve the warm young Earth puzzle. *The Astrophysical Journal* **660**, 1700-1706.

Mirabel, I. F., Rodriguez, L. F. (1999): Sources of relativistic jets in the Galaxy. *Annual Review of Astronomy and Astrophysics* **37**, 409-443.

Mitalas, R., Sills, K. R. (1992): On the photon diffusion time scale for the Sun. *The Astrophysical Journal* **401**, 759-760.

Morgan, W. W., Keenan, P. C., Kellman, E. (1943): *An Atlas of Stellar Spectra, with an Outline of Spectral Classification*. Chicago: University of Chicago Press.

Morgan, W. W., Sharpless, S., Osterbrock, D. (1952): Some features of galactic structure in the neighborhood of the Sun. *Astronomical Journal*, **57**, 3. Reproduced in Lang and Gingerich (1979).

Mössbauer, R. L. (1958): Kernresonanzfluoreszenz von Gammastrahlung in Ir^{191} (Nuclear resonance fluorescence by gamma radiation in Ir^{191}). *Zeitshrift für Physik* **151** (2), 124-143.

Muhleman, D. O. (1969): On the radio method of determining the astronomical unit. *Monthly Notices of the Royal Astronomical Society* **144**, 151-157.

Müller, C. A., Oort, J. H. (1951): The interstellar hydrogen line at 1,420 MHz and an estimate of galactic rotation. *Nature* **168**, 356-358. Reproduced in Lang and Gingerich (1979).

Munk, W. H., Mac Donald, G. J. F. (1960, 1975): *The Rotation of the Earth: A Geophysical Discussion*. Cambridge, England: Cambridge University Press 1960, 1975.

Narlikar, J. V., Padmanabhan, T. (1991): Inflation for astronomers. *Annual Review of Astronomy and Astrophysics* **29**, 325-362.

Natário, J. (2011): *General Relativity without Calculus: A Concise Introduction to the Geometry of Relativity*. New York: Springer 2011.

Nather, R. E., Warner, B., Macfarlane, M. (1969): Optical pulsations in the Crab Nebula pulsar. *Nature* **221** (5180), 527.

Neddermeyer, S. H., Anderson, C. D. (1937): Note on the nature of cosmic-ray particles. *Physical Review* **51** (10), 884-886.

Neugebauer, M. (1997): Pioneers of space physics: a career in the solar wind. *Journal of Geophysical Research* **102**, A12, 26,887-26,894.

Neugebauer, M., Snyder, C. W. (1966): *Mariner 2* observations of the solar wind. *Journal of Geophysical Research* **71**, 4469-4484.

Newcomb, S. (1895): *The Elements of the Four Inner planets and the Fundamental Constants*. Washington, D.C.: U.S. Government Printing Office, 1895.

Newton, H. W., Nunn, M. L. (1951): The Sun's rotation derived from sunspots 1934−1944 and additional results. *Monthly Notices of the Royal Astronomical Society* **111**, 413-421.

Newton, I. (1671): A letter of Mr. Isaac Newton, Professor of the Mathematicks in the University of Cambridge; containing his new theory about light and colors. *Philosophical Transactions of the Royal Society* **6**, 3075-3087.

Newton, I. (1687): *Philosophiae Naturalis Principia Mathematica* (*Mathematical Principles of Natural Philosophy*). Three editions London, 1687, Cambridge, 1713, London, 1726. English translation of the third edition assembled and edited by Alexandre Koyré and I. Bernard Cohen, Cambridge, Massachusetts: Harvard University Press 1971.

Newton, I. (1704): *Opticks: or, a Treatise of the Reflexions, Refractons, Inflexions and Colours of Light*. London: S. Smith and B. Walford, Printers to the Royal Society.

Nobili, A. M., Will, C. M. (1986): The real value of Mercury's perihelion advance. *Nature* **320**, 39.
Noyes, R. W., et al. (1984): Rotation, convection, and magnetic activity in lower main sequence stars. *The Astrophysical Journal* **279**, 763-777.
Noyes, R. W., Leighton, R. B. (1963): Velocity fields in the solar atmosphere II. The oscillation field. *The Astrophysical Journal* **138**, 631-647.
Olbers, W. (1826): *Über die Durchsichtigkeit des Weltraums* (*About the Transparency of the Universe*). Bode Jb. 15. Reproduced in Dickson (1968) and Jaki (1969).
Oort, J. H. (1927): Observational evidence confirming Lindblad's hypothesis of a rotation of the galactic system. *Bulletin of the Astronomical Institutes of the Netherlands* **3**, 275. Reproduced in Lang and Gingerich (1979).
Oort, J. H. (1928): Dynamics of the galactic system in the vicinity of the Sun. *Bulletin of the Astronomical Institutes of the Netherlands* **4**, 269.
Oort, J. H. (1983): Superclusters. *Annual Review of Astronomy and Astrophysics* **21**, 373-428.
Oort, J. H., Hulst, H. C. van de (1946): Gas and smoke in interstellar space. *Bulletin of the Astronomical Institues of the Netherlands* **10**, 187.
Oort, J. H., Kerr, F. J., Westerhout, G. (1958): The galactic system as a spiral nebula. *Monthly Notices of the Royal Astronomical Society* **118**, 379-389. Reproduced in Lang and Gingerich (1979).
Öpik, E. (1922): An estimate of the distance of the Andromeda nebula. *The Astrophysical Journal* **55**, 406-410.
Öpik, E. (1938): Stellar structure, source of energy and evolution. *Publications de l'Observatoire astronomique de l'Université de Tartu* **30**, No. 3, 1-115. Reproduced in Lang and Gingerich (1979).
Öpik, E. J. (1951): Stellar models with variable composition II. Sequences of models with energy generation proportional to the fifteenth power of temperature. *Proceedings of the Royal Irish Academy* **54**, No. 4, 7 and *Contributions from the Armagh Observatory* **1**, No. 3, 1.
Öpik, E. J. (1953): Stellar associations and supernovae. *Irish Astronomical Journal* **2**, 219.
Oppenheimer, J. R., Snyder, H. (1939): On continued gravitational contraction. *Physical Review* **56**, 455-459. Reproduced in Lang and Gingerich (1979).
Oppenheimer, J. R., Volkoff, G. M. (1939): On massive neutron cores. *Physical Review* **55**, 374. Reproduced in Lang and Gingerich (1979).
Orosz, J. A. et al. (2011): The mass of the black hole in Cygnus X-1. *The Astrophysical Journal* **742** (2), 84.
Osterbrock, D. E. (1961): The heating of the solar chromosphere, plages, and corona by magnetohydrodynamic waves. *The Astrophysical Journal* **134**, 347-388.
Osterbrock, D. E. (1965): Temperature in H II regions and planetary nebulae. *The Astrophysical Journal* **142**, 1423.
Osterbrock, D. E. (1991): Active galactic nuclei. *Reports on Progress in Physics* **54**, 579-633.
Osterbrock, D. E., Mathews, W. G. (1986): Emission-line regions of active galaxies and QSOs. *Annual Review of Astronomy and Astrophysics* **24**, 171-203.
Ostriker, J. P., Gunn, J. E. (1969): On the nature of pulsars I. Theory. *The Astrophysical Journal* **157**, 1395.
Pacini, F. (1967): Energy emission from a neutron star. *Nature* **216**, 567-568. Reproduced in Lang and Gingerich (1979).
Paczynski, B. (1971): Evolutionary processes in close binary systems. *Annual Review of Astronomy and Astrophysics* **9**, 183-208.
Page, D. et al. (2011): Rapid cooling of the neutron star in Cassiopeia A triggered by neutron superfluidity in dense matter. *Physical Review Letters* **106**, 081101.
Paladini, R., Davies, R. D., De Zotti, G. (2004): Spatial distribution of galactic H II regions. *Monthly Notices of the Royal Astronomical Society* **347**, 237-245.
Pallavicini, R., et al. (1981): Relations among stellar X-ray emission observed from Einstein, stellar rotation and bolometric luminosity. *The Astrophysical Journal* **248**, 279.

Paradijs, J. Van, *et al*. (1997): Transient optical emission from the error box of the γ-ray burst of 28 February 1997. *Nature* **386**, 686.

Paradijs, J. Van, Kouveliotou, C., Wjers, R. A. M. J. (2000): Gamma-ray burst afterglows *Annual Review of Astronomy and Astrophysics* **38**, 379-425.

Parker, E. N. (1958): Origin and dynamics of cosmic rays. *Physical Review* **109**, 1328.

Parker, E. N. (1960): The hydrodynamic theory of solar corpuscular radiation and stellar winds. *The Astrophysical Journal* **132**, 821-866.

Parker, E. N. (1963): *Interplanetary Dynamical Processes*. New York: John Wiley 1963.

Parker, P. D., Bahcall, J. N., Fowler, W. A. (1964): Termination of the proton-proton chain in stellar interiors. *The Astrophysical Journal* **139**, 602.

Paschen, F. (1908): Zur Kenntnis ultraroter Linienspektra I (On the knowledge of infrared line spectra I). *Annalen der Physik* **27**, 537.

Patterson, C. (1956): Age of the meteorites and the Earth. *Geochim. et Cosmochim. Acta* **10**, 230.

Pauli, W. (1927): Zur Quantenmechanik des magnetischen Elektrons (On the quantum mechanics of the magnetic electron). *Zeitschrift für Physik* **43**, 601.

Pauli, W. (1933): *Remarks at the Seventh Solvay Conference*, October 1933.

Pawsey, J. L. (1946): Observation of million degree thermal radiation from the sun at a wavelength of 1.5 meters. *Nature* **158**, 633-634.

Payne, C. H. (1925): The relative abundances of the elements. In: *Stellar atmospheres* (C. H. Payne). Cambridge, Massachusetts: Harvard University Press 1925. Reproduced in Lang and Gingerich (1979).

Peale, S. J. (1999): Origin and evolution of the natural satellites. *Annual Review of Astronomy and Astrophysics* **37**, 533-602.

Pease, F. G. (1918): The rotation and radial velocity of the central part of the Andromeda Nebula. *Proceedings of the National Academy of Sciences* **4**, 21-24.

Peebles, P. J. E. (1995): Mass of the Milky Way and redshifts of the nearby galaxies. *The Astrophysical Journal* **449**, 52.

Peebles, P. J. E. (2003): The cosmological constant and dark energy. *Reviews of Modern Physics* **75**, 559-599.

Peebles, P. J. E., Ratra, B. (2003): The cosmological constant and dark energy. *Review of Modern Physis* **75**, 559-606.

Peebles, P. J. E., Schramm, D. N., Turner, E. L., Kron, R. G. (1991): The case for the relativistic hot big bang cosmology. *Nature* **352**, 769.

Peebles, P. J. E., Yu, J. T. (1970): Primeval adiabatic perturbation in an expanding universe. *The Astrophysical Journal* **162**, 815.

Penrose, R. (1965): Gravitational collapse and space-time singularities. *Physical Review Letters* **14**, 57.

Penzias, A. A. (1979): The origin of the elements. *Reviews of Modern Physics* **51**, 425-431.

Penzias, A. A., Wilson, R. W. (1965): A measurement of excess antenna temperature at 4080 MHz. *The Astrophysical Journal* **142**, 419-421. Reproduced in Lang and Gingerich (1979).

Perl, M. L., *et al*. (1975): Evidence for anomalous lepton production on $e^+ - e^-$ annihilation. *Physical Review Letters* **35** (22), 1489.

Perl, M. L., *et al*. (1976): Properties of anomalous $e\mu$ events produced in $e^+ e^-$ annihilation. *Physics Letters* **63B**, 466.

Perlmutter, S. et al. (1998): Discovery of a supernova explosion at half the age of the universe. *Nature* **391**, 51.

Perlmutter, S. et al. (1999): Measurement of omega and lambda from 42 high-redshift supernovae. *The Astrophysical Journal* **517**, 565-586.

Perrin, J. (1895): New experiments on the cathode rays. Read before *Paris Academy of Sciences*, December 30, 1895. English translation in *Nature* **53**, 298-299 (1896).

Peterson, D. M. et al. (2006): Vega is a rapidly rotating star. *Nature* **440** (7096), 896-899.

Pfund, A. H. (1924): The emission of nitrogen and hydrogen in the infrared. *Journal of the Optical Society of America* **9**, 193.

Phinney, E. S., Kulkarni, S. R. (1994): Binary and millisecond pulsars. *Annual Review of Astronomy and Astrophysics* **32**, 591-639.
Piddington, J. H. (1956): Solar atmospheric heating by hydromagnetic waves. *Monthly Notices of the Royal Astronomical Society* **116**, 314.
Pigott, E. (1785): Observations of a new variable star. *Philosophical Transactions of the Royal Society* **75**, 127-136.
Pilkington, J. D. et al. (1968): Observations of some further pulsed radio sources. *Nature* **218**, 126.
Planck, M. (1901): Über das Gesetz der Energieverteilung im Normalspectrum (On the law of energy distribution in a normal spectrum). *Annalen der Physik* **4**, 553.
Planck, M. (1910): Zur Theorie der Wärmestrahlung (On the theory of thermal radiation). *Annalen der Physik* **31**, 758.
Planck, M. (1913): *The Theory of Heat Radiation*. Reproduced New York: Dover 1959.
Pogson, N. (1856): Magnitude of 36 of the minor planets. *Monthly Notices of the Royal Astronomical Society* **17**, 12.
Popper, D. M. (1954): Red shift in the spectrum of 40 Eridani B. *The Astrophysical Journal* **120**, 316.
Popper, D. M. (1980): Stellar masses. *Annual Review of Astronomy and Astrophysics* **18**, 115-164.
Pound, R. V., Rebka, G. A. (1959): Gravitational redshift in nuclear resonance. *Physical Review Letters* **3**, 439.
Pound, R. V., Rebka, G. A. (1960): Apparent weight of photons. *Physical Review Letters* **4**, 337.
Pound, R. V., Snider, J. L. (1964): Effect of gravity on nuclear resonance. *Physical Review Letters* **13**, 539.
Pound, R. V., Snider, J. L. (1965): Effect of gravity on gamma radiation. *Physical Review* **140**, B788.
Poynting, J. H. (1904): Radiation in the solar system: its effect on temperature and its pressure on small bodies. *Philosophical Transactions of the Royal Society of London*, *Series A* **202**, 346-398.
Pringle, J. E. (1981): Accretion discs in astrophysics. *Annual Review of Astronomy and Astrophysics* **19**, 137-160.
Proudman, J. (1952): The generation of noise by isotropic turbulence. *Proceedings of the Royal Society of London* **A214**, 119.
Provencal, J. L. (1998): Testing the white dwarf mass-radius relation with *HIPPARCOS*. *The Astrophysical Journal* **494**, 759-767.
Provencal, J. L., Shipman, H. L., Hog, E., Thejll, P. (1998): Testing the white dwarf mass-radius relation with *HIPPARCOS*. *Astrophysical Journal* **494**, 759-767.
Putman, M., Peek, J. E. G., Joung, M. R. (2012): Gaseous galaxy halos. *Annual Review of Astronomy and Astrophysics* **50**, 491-529.
Quirrenbach, A. (2001): Optical interferometery. *Annual Review of Astronomy and Astrophysics* **39**, 353-401.
Rabe, E. (1950): Derivation of the fundamental astronomical constants from the observation of Eros during 1926-1945. *Astronomical Journal* **55**, 112-126.
Ramsay, W. (1895): Helium, a gaseous constituent of certain minerals. Part I, II. *Proceedings of the Royal Society of London* **58**, 80-89, **59**, 325-330.
Ramsay, W., Soddy, F. (1903): Experiments in radioactivity and the production of helium from radium. *Proceedings of the Royal Society* **72**, 204-207.
Rana, N. C. (1991): Chemical evolution of the Galaxy. *Annual Review of Astronomy and Astrophysics* **29**, 129-162.
Ratra, B., Vogeley, M. S. (2008): The beginning and evolution of the universe. *Publications of the Astronomical Society of the Pacific* **120**, 235-265.
Rayleigh, Lord (John Strutt) (1871): On the light from the sky, its polarization and colour. *Philosophical Magazine* **41**, 107-120, 274-279.

Rayleigh, Lord (John Strutt) (1899): On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky. *Philosophical Magazine*, **47**, 375-394.
Rayleigh, Lord (John Strutt) (1900): Remarks upon the law of complete radiation. *Philosophical Magazine* **49**, 539-540.
Rayleigh, Lord (John Strutt) (1905): The dynamical theory of gases and of radiation. *Nature* **72**, 54.
Rayleigh, Lord (John Strutt) (1905): The rate of formation of radium. *Nature* **72**, 365.
Rayleigh, Lord (John Strutt) (1916): On convective currents in a horizontal layer of fluid when the higher temperature is on the under side. *Philosophical Magazine* **32**, 529.
Reasenberg, R. D., *et al*. (1979): *Viking* relativity experiment: Verification of signal retardation by solar gravity. *Astrophysical Journal Letters* **234**, L219.
Reber, G. (1944): Cosmic static. *The Astrophysical Journal* **100**, 279-287. Reproduced in Lang and Gingerich (1979).
Rees, M. J. (1984): Black hole models for active galactic nuclei. *Annual Review of Astronomy and Astrophysics* **22**, 471-506.
Reeves, H. (1994): On the origin of the light elements ($Z < 6$). *Reviews of Modern Physics* **66**, 193-216.
Reeves, H., Audouze, J., Fowler, W. A., Schramm, D. N. (1973): On the origin of the light elements. *The Astrophysical Journal* **179**, 909.
Reeves, H., Fowler, W. A., Hoyle, F. (1970): Galactic cosmic ray origin of Li, Be and B in stars. *Nature* **226**, 727.
Reid, I. et al. (2002): The Palomar/MSU nearby star spectroscopic survey. IV. The luminosity function in the solar neighborhood and M dwarf kinematic. *The Astronomical Journal* **124**, 2721-2738.
Reid, M. J. (1993): The distance to the center of the Galaxy. *Annual Review of Astronomy and Astrophysics* **31**, 345-372,
Reid, M. J. et al. (2011): The trigonometric parallax of Cygnus X-1. *The Astrophysical Journal* **742** (2), 83.
Reid, M. J., Moran, J. M. (1981): Masers. *Annual Review of Astronomy and Astrophysics* **19**, 231-276.
Reid, N. I. (1999): The H−R diagram and the galactic distance scale after *HIPPARCOS*. *Annual Review of Astronomy and Astrophysics* **37**, 191-237.
Reifenstein, E. C. et al. (1969): Crab Nebula pulsar NP0527. *Physical Review Letters* **22** (7), 311.
Reines, F., Cowan, C. L. Jr. (1953): Detection of the free neutrino. *Physical Review* **92**, 830-831.
Remillard, R. A., McClintock, J. E. (2006): X-ray properties of black-hole binaries. *Annual Review of Astronomy and Astrophysics* **44**, 49-92.
Renn, J., Sauer, T., Stachel, J. (1997): The origin of gravitational lensing: A postcript to Einstein's 1936 Science paper. *Science* **275**, 184.
Revelle, R., Suess, H. E. (1957): Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO2 during the past decades. *Tellus* **9**, 18-27.
Reynolds, C. S. (2008): Bringing black holes into focus. *Nature* **455** (7209), 39.
Reynolds, S. P. (2008): Supernova remnants at high energy. *Annual Review of Astronomy and Astrophysics* **46**, 89-126.
Richards, D. W., Comella, J. M. (1969): The period of the pulsar NP 0532. *Nature* **222**, 551.
Riess, A. G. et al. (1998): Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal* **116**, 1009-1038.
Riess, A. G. et al. (2011): A 3 percent solution: determination of the Hubble constant with the Hubble Space Telescope and wide field camera 3. *The Astrophysical Journal* **730**, 119.
Riess, A. G., Press, W. H., Kirshner, R. P. (1995): Using Type Ia supernova light curve shapes to measure the Hubble constant. *Astrophysical Journal Letters* **438**, L17.
Riess, A. G., Press, W. H., Kirshner, R. P. (1996): A precise distance indictor: Type Ia supernova multicolor light-curve shapes. *The Astrophysical Journal* **473**, 88.

Ritter, A. (1898): On the constitution of gaseous celestial bodies. *The Astrophysical Journal* **8**, 293.

Robertson, D. S., Carter, W. E., Dillinger, W. H. (1991): New measurement of solar gravitational deflection of radio signals using VLBI. *Nature* **349**, 768.

Robertson, H. P. (1935): Kinematics and world-structure. *The Astrophysical Journal* **82**, 284.

Robertson, H. P. (1936): Kinematics and world-structure II, III. *The Astrophysical Journal* **83**, 187, 257.

Robertson, H. P. (1937) : Dynamical effects of radiation in the solar system. *Monthly Notices of the Royal Astronomical Society* **97**, 423-438.

Roche, E. (1849/50/51): Mémoire sur la figure d'une masse fluide (Soumise à l'attraction d'un point éloigné), Memoir on the figure of a fluid mass (subject to the attraction of a distant point). *Academie des sciences de Montpellier: Memoires de la section des sciences* **1**, 243-262, 333-348, **2**, 21-32.

Rolfs, C. E., Rodney, W. S. (1988, 2005): *Cauldrons in the Cosmos: Nuclear Astrophysics*. Chicago, Illinois: University of Chicago Press 1988, 2005.

Rømer, O. (1677): A demonstration concerning the motion of light. *Philosophical Transaction of the Royal Society* No. 136, 893-894 (June 25, 1677).

Röntgen, W. C. (1896): On a new kind of rays. *Würzburg Physical and Medical Society*, 1895. Translated by A. Stanton *Nature* **53**, 274 (1876).

Rood, H. J. (1981): Clusters of galaxies. *Reports Progress Physics* **44**, 1077 .

Rood, H. J. (1988): Voids. *Annual Review of Astronomy and Astrophysics* **26**, 245-294.

Rose, R. D., et al. (1969): Determination of the gravitational constant G. *Physical Review Letters* **23**, 655.

Roseveare, N. T.: *Mercury's perihelion from Le Verrier to Einstein*. Oxford: Clarendon Press 1982.

Rosse, Earl (William Parsons) (1850): Observations on the Nebulae. *Philosophical Transactions of the Royal Society* **140**, 499-514.

Rubin, V. C. (1983): The rotation of spiral galaxies. *Science* **220**, 1339.

Russell, H. N. (1913): "Giant" and "dwarf' stars. *The Observatory* **36**, 324-329.

Russell, H. N. (1914): Relations between the spectra and other characteristics of stars. *Popular Astronomy* **22**, 275-294. Reproduced in Lang and Gingerich (1979).

Russell, H. N. (1929): On the composition of the sun's atmosphere. *The Astrophysical Journal* **70**, 11-82. Reproduced in Lang and Gingerich (1979).

Russell, H. N., Dugan, R. S., Stewart, J. W. (1927): *Astronomy II: Astrophysics and Stellar Astronomy*. Boston: Ginn and Co.

Rutherford, E. (1899): Uranium radiation and the electrical conduction produced by it. *Philosophical Magazine* **47**, 109-163.

Rutherford, E. (1900): A radioactive substance emitted from thorium compounds. *Philosophical Magazine* **49**, 1-14.

Rutherford, E. (1904): *Radioactivity*. Cambridge, England: Cambridge University Press 1904.

Rutherford, E. (1905): Radium - the cause of the earth's heat. *Harper's Magazine February*, 390-396.

Rutherford, E. (1911): The scattering of α and β particles by matter and the structure of the atom. *Philosophical Magazine* **21**, 669.

Rutherford, E. (1914): The structure of the atom. *Philosophical Magazine* **27**, 488.

Rutherford, E. (1929): Origin of actinium and age of the Earth. *Nature* **123**, 313-314.

Rutherford, E., Ardrade, E. N. da C. (1914): The wavelength of soft gamma rays from Radium B. *Philosophical Magazine* **27**, 854-868.

Rutherford, E., Chadwick, J. (1921): The disintegration of elements by α-particles. *Nature* **107,** 41.

Rutherford, E., Soddy, F. (1902): The cause and nature of radioactivity, part I, part II. *Philosophical Magazine* **4**, 370, 569.

Rutherford, E., Soddy, F. (1903): The radioactivity of uranium. A comparative study of the radioactivity of radium and thorium. Condensation of the radioactive emanations. The

products of radioactive change and their specific material nature. *Philosophical Magazine* **5**, 441, 445, 561, 576.
Rydberg, J. R. (1890): On the structure of the line-spectra of the chemical elements. *Philosophical Magazine* **29**, 331.
Sackmann, I.-J., Boothroyd, A. I. (2003): Our Sun. V. A bright young Sun consistent with helioseismology and warm temperatures on ancient Earth and Mars. *The Astrophysical Journal* **583**, 1024-1039.
Sackmann, I.-J., Boothroyd, A. I., Kraemer, K. E. (1993): Our Sun III. Present and future. *The Astrophysical Journal* **418**, 457-468.
Sagan, C., Chyba, C. (1997): The early faint Sun paradox: Organic shielding of ultraviolet-labile greenhouse gases. *Science* **276**, 1217-1221.
Saha, M. N. (1920): Ionization in the solar chromosphere. *Philosophical Magazine* **40**, 479 -488. Reproduced in Lang and Gingerich (1979).
Saha, M. N. (1921): On the physical theory of stellar spectra. *Proceedings of the Royal Society of London* **A99**, 13-153..
Salpeter, E. E. (1952): Nuclear reactions in the stars: I. Proton-proton chain. *Physical Review* **88**, 547.
Salpeter, E. E. (1952): Nuclear reactions in stars without hydrogen. *The Astrophysical Journal* **115**, 326-328. Reproduced in Lang and Gingerich (1979).
Salpeter, E. E. (1955): The luminosity function and stellar evolution. *The Astrophysical Journal* **121**, 161-167.
Salpeter, E. E. (1957): Nuclear reactions in stars. Buildup from helium. *Physical Review* **107**, 516-525.
Salpeter, E. E. (1964): Accretion of interstellar matter by massive objects. *The Astrophysical Journal* **140**, 796.
Salpeter, E. E. (1977): Formation and destruction of dust grains. *Annual Review of Astronomy and Astrophysics* **15**, 267-293.
Sampson, R. A. (1894): On the rotation and mechanical state of the Sun. *Memoirs of the Royal Astronomical Society* **51**, 123.
Sandage, A. R. (1953): The color-magnitude diagram for the globular cluster M 3. *Astronomical Journal* **58**, 61.
Sandage, A. R. (1957): Observational approach to evolution III. semiempirical evolution tracks for M 67 and M 3. *The Astrophysical Journal* **126**, 326.
Sandage, A. R. (1958): Current problems in the extragalactic distance scale. *The Astrophysical Journal* **127**, 513.
Sandage, A. R. (1970): Main-sequence photometry, color-magnitude diagrams, and ages for the globular clusters M 3, M 13, M 15 and M 92. *The Astrophysical Journal* **162**, 841.
Sandage, A. R. (1972): Classical cepheids - cornerstone to extragalactic distances. *Quarterly Journal of the Royal Astronomical Society* **13**, 202.
Sandage, A. R. (1986): The population concept, globular clusters, subdwarfs, ages, and the collapse of the Galaxy. *Annual Review of Astronomy and Astrophyics*. **24**, 421-458.
Sandage, A. R. (1988): Observational tests of world models. *Annual Review of Astrononomy and Astrophysics* **26**, 561-630.
Sandage, A .R. (1988): The case for Ho = 55 from the 21 centimeter linewidth absolute magnitude relation for field galaxies. *The Astrophysical Journal* **331**, 605.
Sandage, A. R., Schwarzschild, M. (1952): Inhomogeneous stellar models II. models with exhausted cores in gravitational contraction. *The Astrophysical Journal* **116**, 463-476. Reproduced in Lang and Gingerich (1979).
Sandage, A. R., Tammann, G. A. (2006): Absolute magnitude calibrations of Population I and II Cepheids and other pulsating variables in the Hertzsprung−Russell diagram. *Annual Review of Astronomy and Astrophysics* **44**, 93-140.
Sanders, D. B., Mirabel, I. F. (1996): Luminous infrared galaxies. *Annual Review of Astronomy and Astrophysics* **34**, 749-792.

Sargent, W. L. W. et al. (1978): Dynamical evidence for a central mass concentration in the galaxy M 87. *The Astrophysical Journal* **221**, 731.
Savage, B. D., Mathis, J. S. (1979): Observed properties of interstellar dust. *Annual Review of Astronomy and Astrophysics* **17**, 73-111.
Savage, B. D., Sembach, K. R. (1996): Interstellar abundances from absorption-line observations with the *Hubble Space Telescope*. *Annual Review of Astronomy and Astrophysics* **34**, 279-329.
Schatzman, E. (1945): Théorie du debit d'énergie des naines blanches *Annales d'Astrophysique* **8**, 143-209. Also see Koester, D., Chanmugam, G. (1990): Physics of white dwarf stars. *Reports on Progress in Physics* **53**, 837-915.
Schatzman, E. (1949): The heating of the solar corona and chromosphere. *Annales d'Astrophysique* **12**, 203-218.
Schatzman, E. (1958): *White Dwarfs*. Amsterdam: North Holland 1958.
Schechter, P. (1976): An analytic expression for the luminosity function for galaxties. *The Astrophysical Journal* **203**, 297.
Scheuer, P. A. G. (1960): The absorption coefficient of a plasma at radio frequencies. *Monthly Notices of the Royal Astronomical Society* **120**, 231.
Schiaparelli, G. (1889): Sulla rotazione di Mercurio. *Astronomische Nachrichten* **123**, 241.
Schild, R. E. (1977): Interstellar reddening law. *Astronomical Journal* **82**, 337.
Schilling, G. (2001): New model shows Sun was a hot young star. *Science* **293**, 2188-2189.
Schmidt, B. P. et al. (1998): The high-z supernova search: measuring cosmic deceleration and global curvature of the universe using Type Ia supernova. *The Astrophysical Journal* **507**, 46.
Schmidt, M. (1963): 3C 273 A star-like object with large red-shift. *Nature* **197**, 1040. Reproduced in Lang and Gingerich (1979).
Schneider, P. (1984): The amplification caused by gravitational bending of light. *Astronomy and Astrophysics* **140**, 119.
Schneider, P. (2006): *Extragalactic Astronomy and Cosmology*. New York: Springer 2006.
Schönberg, M., Chandrasekhar, S. (1942): On the evolution of main-sequence stars. *The Astrophysical Journal* **96**, 161.
Schott, G. A. (1912): *Electromagnetic Radiation and the Mechanical Reactions Arising From It*. Cambridge, England: Cambridge University Press 1912.
Schrijver, C. J., Title, A. M. (2003): The magnetic connection between the solar photosphere and the corona. *Astrophysical Journal (Letters)* **597**, L165-L168.
Schuster, A. (1905): Radiation through a foggy atmosphere. *Astrophysical Journal* **21**, 1. Reproduced in *Selected Papers on the Transfer of Radiation* (ed. D. H. Menzel). New York: Dover 1966.
Schwabe, S. H. (1844): Sonnen-Beobachtungem im Jahre 1843. (Solar observations during 1843). *Astronomische Nachrichten* **21**, No. 495, 233-236. Reprinted in *Kosmos* (Ed. A. Von Humboldt). English translation Meadows (1970).
Schwarzschild, K. (1906): Über das Gleichgewicht der Sonnenatmosphäre (On the equilibrium of the solar atmosphere.) *Nach. Ges. Gott.* **195**, 41. English translation in: *Selected Papers on the Transfer of Radiation* (ed. D. H. Menzel). New York: Dover 1966.
Schwarzschild, K. (1916): Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie (The gravitational field of a point mass according to Einstein's theory). *Sitz. Acad. Wiss., Physik-Math Kl.* **1**, 189. English translation in Lang and Gingerich (1979).
Schwarzschild, M. (1948): On noise arising from the solar granulation. *The Astrophysical Journal* **107**, 1.
Schwarzschild, M. (1958): *Structure and Evolution of Stars*. Princeton, New Jersey: Princeton University Press 1958.
Schwarzschild, M., Härm, R. (1962): Red giants of population II. *The Astrophysical Journal* **136**, 158.
Schwinger, J. (1949): The classical radiation of accelerated electrons. *Physical Review* **75**, 1912-1915.

Seager, S., Deming, D. (2010) Exoplanet atmospheres. *Annual Review of Astronomy and Astrophysics* **48**, 631-672.
Seeger, P. A., Fowler, W. A., Clayton, D. D. (1965): Nucleosynthesis of heavy elements by neutron capture. *Astrophysical Journal Supplement* **11**, 121.
Seidelmann, P. K. (2005): *Explanatory Supplement to the Astronomical Almanac*. Mill Valley, California: University Science Books 2005.
Seyfert, C. K. (1943): Nuclear emission in spiral nebulae. *The Astrophysical Journal* **97**, 28-40. Reproduced in Lang and Gingerich (1979).
Shajn, G., Struve, O. (1929): On the rotation of the stars. *Monthly Notices of the Royal Astronomical Society* **89**, 222-239 . Reproduced in Lang and Gingerich (1979).
Shapiro, I. I. (1964): Fourth test of general relativity. *Physical Review Letters* **13** (26) 789-791.
Shapiro, I. I. *et al*. (1971): Fourth test of general relativity − new radar result. *Physical Review Letters* **26**, 1132-1135. Reproduced in Lang and Gingerich (1979).
Shapiro, S. S. et al. (2004): Measurement of the solar gravitational deflection of radio waves using geodetic very-long-baseline interferometry data 1979-1999. *Physical Review Letters* **92**, 121101.
Shapley, A. E. (2011): Physical properties of galaxies from z = 2 to 4. *Annual Review of Astronomy and Astrophysics* **49**, 525-580.
Shapley, H. (1914): On the nature and cause of Cepheid variation. *The Astrophysical Journal* **40**, 448.
Shapley, H. (1918): On the determination of the distances of globular clusters. *The Astrophysical Journal* **48**, 89.
Shapley, H., Curtis, H. D. (1921): The Scale of the Universe. *Bulletin of the National Research Council of the National Academy of Sciences* (Washington, D.C.) **2**, 171-217. Part 1, Harlow Shapley, 171-193, Reproduced in Lang and Gingerich (1979) Part 2, Heber D. Curtis, 194-217.
Shapley, H., Howarth, H. E. (1929): *A Source Book in Astronomy*. New York: McGraw-Hill 1929.
Share, G. H., et al. (2004*): RHESSI* e+-e- annihilation radiation observations: Implications for conditions in the flaring solar chromosphere. *The Astrophysical Journal (Letters)* **615**, L169-L172.
Sheeley, N. R. Jr., Wang, Y.-M., Nash, A. G. (1992): A new determination of the solar rotation rate. *The Astrophysical Journal* **401**, 378-385.
Shipman, H. L. (1972): Masses and radii of white dwarfs. *The Astrophysical Journal* **177**, 723.
Shipman, H. L., Provencal, J. L., Høg, E., Thejll, P. (1997): The mass and radius of 40 Eridani B from *HIPPARCOS*: An accurate test of stellar interior theory. *Astrophysical Journal Letters* **488**, L43.
Shklovskii, I. S. (1949): In Russian. *Astronomicheskii Zhurnal* **26**, 10-14.
Shklovskii, I. S. (1953): On the nature of the luminescence of the Crab Nebula. *Doklady Akad. Nauk SSSR* **90**, 983-986. English translation in Lang and Gingerich (1979).
Shklovskii, I. S. (1953): The possibility of observing monochromatic radio emissions from interstellar molecules. *Doklady Akad. Nauk SSSR* **92**, No. 1, 25.
Shternin, P. S., et al. (2011): Cooling neutron star in the Cassiopeia A supernova remnant: Evidence for superfluidity in the core. *Monthly Notices of the Royal Astronomical Society* **412**, L108-L112.
Shu, F. H., Adams, F. C., Lizano, S. (1987): Star formation in molecular clouds: Observation and theory. *Annual Review of Astronomy and Astrophysics* **25**, 23-81.
Shull, J. M., Beckwith, S. (1982): Interstellar molecular hydrogen. *Annual Review of Astronomy and Astrophysics* **20**, 163-190.
Slettebak, A. (1949): On the axial rotation of the brighter O and B stars. *The Astrophysical Journal* **110**, 498.
Slettebak, A. (1954): The spectra and rotational velocities of the bright stars of Draper types B8-A2. *The Astrophysical Journal* **119**, 146.

Slettebak, A. (1955): The spectra and rotational velocities of the bright stars of Draper types A3-G0. *The Astrophysical Journal* **121**, 653.
Slettebak, A., Howard, R. F. (1955): Axial rotation in the brighter stars of Draper types B2-B5. *The Astrophysical Journal* **121**, 102.
Slipher, V. M. (1914): The radial velocity of the Andromeda Nebula. *Popular Astronomy* **22**, 19-21, *Lowell Observatory Bulletin No. 58* **2**, 56-57 (1913).
Slipher, V. M. (1915): Spectrographic observations of nebulae. *Popular Astronomy* **23**, 21-24.
Slipher, V. M. (1917): A spectrographic investigation of spiral nebulae. *Proceedings of the American Philosophical Society* **56**, 403-409. Reproduced in Lang and Gingerich (1979).
Smartt, S. J. (2009): Progenitors of core-collapse supernovae. *Annual Review of Astronomy and Astrophysics* **47**, 63-106.
Smith, B. A., Terrile, R. J. (1984): A circumstellar disk around Beta Pictoris. *Science* **226** (4681), 1421-1424.
Smith, I. W. M. (2011): Laboratory astrochemistry: gas-phase processes. *Annual Review of Astronomy and Astrophysics* **49**, 29-66.
Smoot, G. F. (1999): Summary of results from COBE. *AIP Conference Proceedings* **476**, 1-10.
Smoot, G. F. *et al*. (1992): Structure in the *COBE* differential microwave radiometer first-year maps. *Astrophysical Journal Letters* **396**, L1-L5.
Snider, J. L. (1972): New measurement of the solar gravitational redshift. *Physical Review Letters* **28**, 853.
Snodgrass, H. B. (1983): Magnetic rotation of the solar photosphere. *The Astrophysical Journal* **270**, 288-299.
Snow, T. P., McCall, B. J. (2006): Diffuse atomic and molecular clouds. *Annual Review of Astronomy and Astrophysics* **44**, 367-414.
Snyder, L. E., Buhl, D., Zuckerman, B., Palmer, P. (1969): Microwave detection of interstellar formaldehyde. *Physical Review Letters* **22**, 679.
Sobel, D. (1995): *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. New York: Warlker and Co. 1995.
Soderblom, D. R. (2010): The ages of stars. *Annual Review of Astronomy and Astrophysics* **48**, 581-629.
Sofue, Y., Fujimoto, M., Wielebinski, R. (1986): Global structure of magnetic fields in spiral galaxies. *Annual Review of Astronomy and Astrophysics* **24**, 459-497.
Sofue, Y., Rubin, V. (2001): Rotation curves of spiral galaxies. *Annual Review of Astronomy and Astrophysics* **39**, 137-174.
Soldner, J. G. (1801, 1804) Attraktion eines Weltkörpers, an welchem er nahe vorbeigeht (Concerning the deflection of a light ray from its straight path due to the attraction of a heavenly body which it passes closely). *Berliner Astronomische Jahrbuch* 161-172. Article written in 1801, published in 1804.
Solomon, P. M., Vanden Bout, P. A. (2005): Molecular gas at high redshift. *Annual Review of Astronomy and Astrophysics* **43**, 677-725.
Spiegel, E. A. (1971): Convection in stars: Part. I. Basic Boussinesq convection. *Annual Review of Astronomy and Astrophysics* **9**, 323-352.
Spitzer, L. (1948): The temperature of interstellar matter I. *The Astrophysical Journal* **107**, 6.
Spitzer, L. (1949): The temperature of interstellar matter II. *The Astrophysical Journal* **109**, 337.
Spitzer, L. (1990): Theories of the hot interstellar gas. *Annual Review of Astronomy and Astrophysics* **28**, 71-101.
Spitzer, L., Savedoff, M. P. (1950): The temperature of interstellar matter III. *The Astrophysical Journal* **111**, 593-608. Reproduced in Lang and Gingerich (1979).
Staelin, D. H., Reifenstein, E. C. (1968): Pulsating radio sources near the Crab Nebula. *Science* **162** (3861), 1481.
Standish, E. M. (1998): Time scales in the JPL and CfA ephemerides. *Astronomy and Astrophysics* **336**, 381-384.

Starrfield, S., Sparks, W. M., Truran, J. W. (1974): CNO abundances and hydrodynamic models of the nova outburst II. 1.0 solar mass models with enhanced carbon and oxygen. *Astrophysical Journal Supplement* **28**, 247-270.

Starrfield, S., Sparks, W. M., Truran, J. W. (1985): Recurrent novae as a consequence of the accretion of solar material onto a 1.38 solar mass white dwarf. *The Astrophysical Journal* **291**, 136-146.

Stecher, P., Williams, D. A. (1968): Interstellar molecule formation. *The Astrophysical Journal* **146,** 88.

Stefan, A. J. (1879): Beziehung zwischen Wärmestrahlung und Temperatur (Relation between thermal radiation and temperature). *Wien. Ber.* **79**, 397.

Stein, W. A., Soifer, B. T. (1983): Dust in galaxies. *Annual Review of Astronomy and Astrophysics* **21**, 177-207.

Stephenson, F. R., Morrison, L. V. (1984): Long-term changes in the rotation of the Earth - 700 B. C. to A. D. 1980. *Philosophical Transactions of the Royal Society of London*, *Series A* **313**, 47-70.

Stone, E. C., et al. (2005): *Voyager 1* explores the termination shock region and the heliosheath beyond. *Science* **309**, 2017-2020.

Stone, E. C., et al. (2008): An asymmetric solar wind termination shock. *Nature* **454**, 71-74.

Stoner, E. C. (1929): The limiting density of white dwarf stars. *Philosophical Magazine* **7** (7), 63-70.

Stoner, E. C. (1930): The equilibrium of dense stars. *Philosophical Magazine* **9** (7), 944-963.

Stoney, G. J. (1881): On the physical units of nature. *Philosophical Magazine* **11** (5), 384.

Stoney, G. J. (1894): Of the "electron," or atom of electricity. *Philosophical Magazine*, **38** (5), 418-420.

Stoney, G. J. (1898): Of atmospheres upon planets and satellites. *Astrophysical Journal* **7**, 25-55. Reproduced in Lang and Gingericah (1979).

Stoney, G. J. (1900): Escape of gases from planetary atmospheres. *Astrophysical Journal* **11**, 251.

Størmer, C. (1907): Sur les trajectories des corpuscles, electrisés dans l'espace sous l'action du magnétisme terrestre avec l'application aux aurores boréales. *Archives des sciences physiques et naturelles (Geneva)* **24**, 5, 113, 221, 317; **32**, 117-123, 190-219, 277-314, 415-436, 505-509 (1911); **33**, 51-69, 113-150 (1912).

Størmer, C. (1955): *The Polar Aurora*. Oxford, England: The Clarendon Press 1955.

Street, J. C., Stevenson, E. C. (1937): New evidence for the existence of a particle of mass intermediate between the proton and electron. *Physical Review* **52**, 1003-1004.

Strömgren, B. (1931): The possible solutions of the equations of fit on the standard model. *Monthly Notices of the Royal Astronomical Society* **91**, 466-482.

Strömgren, B. (1932): The opacity of stellar matter and the hydrogen content of the stars. *Zeitschrift für Astrophysik* **4**, 118-152.

Strömgren, B. (1939): The physical state of interstellar hydrogen. *The Astrophysical Journal* **89**, 526-547. Reproduced in Lang and Gingerich (1979).

Struve, O. (1930): On the axial rotation of stars. *The Astrophysical Journal* **72**, 1.

Suess, H. E., Urey, H. C. (1956): Abundances of the elements. *Reviews of Modern Physics* **28**, 53-74.

Sulentic, J. W., Marziani, P., Dultzin-Hacyan, D. (2000): Phenomenology of broad emission lines in active galactic nuclei. *Annual Review of Astronomy and Astrophysics* **38**, 521-571.

Sunyaev, R. A., Zeldovich, Y. B. (1980): Microwave background radiation as a probe of the contemporary structure and history of the universe. *Annual Review of Astronomy and Astrophysics* **18**, 537-560.

Suzuki, N. et al. (2012): The *Hubble Space Telescope* cluster supernova survey V. Improving the dark-energy constraints above z greater than 1 and building an early-type-hosted supernova sample. *The Astrophysical Journal* **746**, 85-109.

Swain, M. R., Vasisht, G., Tinetti, G. (2008): The presence of methane in the atmosphere of an extrasolar planet. *Nature* **452** (7185), 329-331.

Sweet, P. A. (1969): Mechanisms of solar flares. *Annual Review of Astronomy and Astrophysics* **7**, 149-176.
Swings, P. (1943): Edlén's identification of the coronal liens with forbidden lines of Fe, Ni, Ca. *The Astrophysical Journal* **98**, 116-128.
Tassoul, J. L. (1978): *Theory of Rotating Stars*. Princeton, New Jersey: Princeton University Press 1978.
Taylor, G. I. (1919): Tidal friction in the Irish sea. *Philosophical Transactions of the Royal Society of London, Series A* **220**, 1-33.
Taylor, J. H. (1992): Pulsar timing and relativistic gravity. *Philosophical Transactions of the Royal Society of London* **A341**, 117.
Taylor, J. H. (1994): Binary pulsars and relativistic gravity. *Review of Modern Physics* **66**, 711-719.
Taylor, J. H., Stinebring, D. R. (1986): Recent progress in the understanding of pulsars. *Annual Review of Astronomy and Astrophysics* **24**, 285-327.
Taylor, J. H., Weisberg, J. M. (1982): A new test of general relativity – Gravitational radiation and the binary pulsar PSR 1913+16. *The Astrophysical Journal* **253**, 908-920.
Taylor, J. H., Weisberg, J. M. (1989): Further experimental tests of relativistic gravity using the binary pulsar PSR 1913 + 16. *The Astrophysical Journal* **345**, 434.
Thomas, L. H. (1926): The motion of the spinning electron. *Natur*e **117**, 514.
Thomas, L. H. (1927): The kinematics of an electron with an axis. *Philosophical Magazine* **1,** 1.
Thompson, M. J. et al. (2003): The internal rotation of the sun. *Annual Review of Astronomy and Astrophysics* **41,** 599-643.
Thomson, J. J. (1897a): Cathode rays. *Philosophical Magazine* **44**, 293-316.
Thomson, J. J. (1897b): Conductivity of a gas through which cathode rays are passing. *Philosophical Magazine* **44**, 298.
Thomson, J. J. (1903, 1906): *Conduction of Electricity Through Gases*. Cambridge, England: Cambridge University Press 1903, 1906.
Thomson, W.: see Kelvin, Lord.
Tice, M. M., Lowe, D. R. (2004): Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* **431**, 549-552.
Tinetti, G. et al. (2007): Water vapour in the atmosphere of a transiting extrasolar planet. *Nature* **448**, 169-171.
Tinsley, B. M. (1975): Nucleochronology and chemical evolution. *The Astrophysical Journal* **198**, 145.
Todhunter, I. (1962): *A History of the Mathematical Theories of Attraction and the Figure of the Earth.* New York: Dover 1962.
Tolman, R. C. (1934): *Relativity, Thermodynamics and Cosmology*. Oxford: At the Clarendon Press 1934.
Tonks, L., Langmuir, I. (1929): Oscillations in ionized gases. *Physical Review* **33**, 195.
Totsuka, Y. (1991): Recent results on solar neutrinos from Kamiokande. *Nuclear Physics* **B19**, 69-76.
Townes, C. H. (1957): Microwave and radiofrequency resonance lines of interest to radio astronomy. In: *Radio Astronomy: I.A.U. Symposium No. 4* (ed H. C. Van De Hulst). Cambridge, England: Cambridge University Press 1957.
Townes, C. H. et al. (2009): A systematic change with time in the size of Betelgeuse. *The Astrophysical Journal Letters* **697**, L127.
Treu, T. (2010): Strong lensing by galaxies. *Annual Review of Astronomy and Astrophysics* **48**, 87-125.
Trimble, V. (1975): The origin and abundances of the chemical elements. *Reviews of Modern Physics* **47**, 877-976.
Trimble, V. (1982): Supernovae. part 1 − the events. *Reviews of Modern Physics* **54**, 1183-1224.
Trimble, V. (1983): Supernovae. part 2 − the aftermath. *Reviews of Modern Physics* **55**, 511-563.
Trimble, V. (1987): Existence and nature of dark matter in the universe. *Annual Review of Astronomy and Astrophysics* **25**, 425-472.

Trimble, V. (1988): 1987A – the greatest supernova since Kepler. *Reviews of Modern Physics* **60**, 859-871.
Trimble, V. L., Greenstein, J. L. (1972): The Einstein redshift in white dwarfs III. *The Astrophysical Journal* **177**, 441.
Troland, T. H., Heiles, C. (1977): The Zeeman effect in radio frequency recombination lines. *The Astrophysical Journal* **214**, 703.
Trumpler, R. J. (1930): Preliminary results on the distances, dimensions, and space distribution of open star clusters. *Lick Observatory Bulletin* **14**, no. 420, 154-188. Reproduced in Lang and Gingerich (1979).
Tsuji, T. (1986): Molecules in stars. *Annual Review of Astronomy and Astrophysics* **24**, 89-125.
Tully, R. B. (1993): The Hubble constant. *Proceedings of the National Academy of Science* **90**, 4806.
Tully, R. B., Fisher, J. R. (1977): A new method of determining distances to galaxies. *Astronomy and Astrophysics* **54**, 661.
Turck-Chièze, S, et al. (1988): Revisiting the solar model. *The Astrophysical Journal* **335**, 415-424.
Turck-Chièze, S. et al. (2001): Solar neutrino emission deduced from a seismic model. *Astrophysical Journal Letters* **555**, L69-L73.
Tyndall, J. (1861): The Bakerian Lecture: On the absorption and radiation of heat by gases and vapours, and on the physical connexion of radiation, absorption, and conduction. *Philosophical Transactions of the Royal Society of London* **151**, 1-36. *Philosophical Magazine, Series 4* **22**, 168-194, 273-285.
Tyndall, J. (1863): *On Radiation Through the Earth's Atmosphere*. Public Lecture January 1863. *Philosophical Magazine*, *Series 4*, **25**, 200-206.
Tyndall, J. (1863): On the relation of radiant heat to aqueous vapor. *Philsophical Magazine, Series 4*, **26**, 30-54.
Udry, S., Santos, N. C. (2007): Statistical properties of exoplanets. *Annual Review of Astronomy and Astrophysics* **45**, 397-439.
Ulrich, M.-H., Maraschi, L., Urry, C. M. (1997): Variability of active galactic nuclei. *Annual Review of Astronomy and Astrophysics* **35**, 445-502.
Ulrich, R. K. (1970): The five-minute oscillations on the solar surface. *Astrophysical Journal* **162**, 993-1002.
Unsöld, A. (1928): Über die Struktur der Fraunhoferschen Linien und die quantitative Spektralanalyse der Sonnenatmosphäre. *Zeitschrift fur Physik* **46**, 765.
Unsöld, A. (1969): Stellar abundances and the origin of the elements. *Science* **163**, 1015.
Vaiana, G. S., Rosner, R. (1978): Recent advances in coronal physics. *Annual Review of Astronomy and Astrophysics* **16**, 393-428.
Van Allen, J. A., McIlwain, C. E., Ludwig, G. H. (1959): Radiation observations with satellite 1958ε. *Journal of Geophysical Research* **64**, 271-286. Reproduced in Lang and Gingerich (1973).
Van de Hulst, H. C. (1945): Radio waves from space: Origin of radiowaves. *Nederlands tijdschrift voor natuurkunde* **11**, 210-221. English translation in Lang and Gingerich (1979).
Van de Hulst, H. C. (1949): The solid particles of interstellar space. *Recherches astronomiques de l'Observatoire d'Utrecht* **11**, pt. 2, 1-50. Reproduced in Lang and Gingerich (1979).
Van den Berg, D. A., Bolte, M., Stetson, P. B. (1996): The age of the galactic globular cluster system. *Annual Review of Astronomy and Astrophysics* **34**, 461-510.
Van den Bergh, S. (1993): Was Tycho's supernova a subluminous supernova of Type Ia? *The Astrophysical Journal* **413**, 67-69.
Van den Bergh, S. (1999): The early history of dark matter. *Publications of the Astronomical Society of the Pacific* **111**, 657-660.
Van Der Kruit, P. C., Freeman, K. C. (2011): Galaxy disks. *Annual Review of Astronomy and Astrophysics* **49**, 301-371.
Van Helden, A. (1985): *Measuring the Universe: Cosmic dimensions from Aristarchus to Halley.* Chicago, Illinois: University of Chicago Press 1985.

Van Helden, A. (1989): *Sidereus Nuncius, or The Sidereal Messenger*. Chicago, Illinois: University of Chicago Press 1989.
Van Paradijs, J. et al. (1997): Transient optical emission from the error box of the gamma ray burst of 28 February 1997. *Nature* **386**, 686-689.
Van Vleck. J. H., Huber, D. L. (1977): Absorption, emission and linebreadths: a semihistorical perspective. *Reviews of Modern Physics* **49**, 939-959.
Vegard, L. (1913): On spectra of the aurora borealis. *Physikalishe Zeitschrift* **14**, 677.
Verbunt, F. (1993): Origin and evolution of x-ray binaries and binary radio pulsars. *Annual Review of Astronomy and Astrophysics* **31**, 93-127.
Verschuur, G. L. (1968): Positive determination of an interstellar magnetic field by measurement of the Zeeman splitting of the 21-cm hydrogen line. *Physical Review Letters* **21**, 775.
Verschuur, G. L. (1971): Recent measurements of the Zeeman effect at 21-centimeter wavelength. *The Astrophysical Journal* **165**, 651.
Vessot, R. F. C., *et. al*. (1980): Test of relativistic gravitation with a space-borne hydrogen maser. *Physical Review Letters* **45**, 2081.
Villard, M. P. (1900): Sur le rayonnement du radium (On the radiation of radium). *Comptes Rendus* **130**, 1178.
Vogt, H. (1926): The relationship between the masses and luminosities of the stars. *Astronomische Nachrichten* **226**, 301-304. English translation in Lang and Gingerich (1979).
Vogt, S. S. et al. (2010): The Lick-Carnegie exoplanet survey: A 3.1 Earth-mass planet in the habitable zone of the nearby M3V star Gliese 581. *The Astrophysical Journal* **723**, 954.
Wagner, W. J. (1984): Coronal mass ejections. *Annual Review of Astronomy and Astrophysics* **22**, 267-289.
Wagoner, R. V., Fowler, W. A., Hoyle, F. (1967): On the synthesis of elements at very high temperature. *Astrophysical Journal* **148**, 3.
Walker, M. F. (1954): Nova DQ Herculis (1934): an eclipsing binary with very short period. *Publications of the Astronomical Society of the Pacific* **66**, 230-232.
Walker, M. F. (1956): Studies of extremely young clusters I: NGC 2264. *Astrophysical Journal Supplement* **2**, 365-387. Reproduced in Lang and Gingerich (1979).
Walker, M. F. (1963): Nova T Aurigae 1891: a new short-period eclipsing binary. *The Astrophysical Journal* **138**, 313.
Wallerstein, G., Conti, P. S. (1969): Lithium and beryllium in stars. *Annual Review of Astronomy and Astrophysic*s **7**, 99-120.
Wasserburg, G. J. et al. (1977): Outline of a lunar chronology. *Philosophical Transactions of the Royal Society of London, Series A* **285**, 7-22.
Weart, S. R. (2008): *The Discovery of Global Warming, Revised and Expanded Edition*. Cambridge, Mass.: Harvard University Press 2008.
Weaver, T. A., Zimmerman, G. B., Woosley, S. E. (1978): Presupernova evolution of massive stars. *The Astrophysical Journal* **225**, 1021.
Webster, B. L., Murdin, P. (1972): Cygnus X-1: A spectroscopic binary with a heavy companion? *Nature* **235**, 37-38. Reproduced in Lang and Gingerich (1979).
Weedman, D. W. (1977): Seyfert galaxies. *Annual Review of Astronomy and Astrophysics* **15**, 69-95.
Wegner, G. (1979): The gravitational redshift for the white dwarf 40 Eri B. *The Astrophysical Journal* **84**, 650.
Wegner, G. (1980): A new gravitational redshift for the white dwarf 40 Eridani B. *Astronomical Journal* **85**, 1255.
Weiler, K. W., Panagia, N., Montes, M. I., Sramek, R. A. (2002): Radio emission from supernovae and gamma-ray bursters. *Annual Review of Astronomy and Astrophysics* **40**, 387-436.
Weiler, K. W., Sramek, R. A. (1988): Supernovae and supernova remnants. *Annual Review of Astronomy and Astrophysics* **26**, 295-341.
Weinreb, S., Barrett, A. H., Meeks, M. L., Henry, J. C. (1963): Radio observations of OH in the interstellar medium. *Nature* **200**, 829-831. Reproduced in Lang and Gingerich (21979).

Weisberg, J. M., Taylor, J. H. (2005): The relativistic binary pulsar B1913+16: Thirty years of observation and analysis. *Astronomical Society of the Pacific Conference Series* **328**, 25.
Weizsäcker, C. F. (1937): Über Elementumwandlungen im Innern der Sterne I (On transformation of the elements in stellar interiors I). *Physikalische Zeitschrift* **38**, 176.
Weizsäcker, C. F. (1938): Über Elementumwandlungen in Innern der Sterne II (Element transformation inside stars II), *Physikalische Zeitschrift* **39**, 633-646. English translation in Lang and Gingerich (1979).
Wells, J. W. (1963): Coral growth and geochronometry. *Nature* **187**, 948-950.
Wesselink, A. J. (1946): The observations of brightness, colour and radial velocity of delta Cephei and the pulsation hypothesis. *Bulletin of the Astronomical Institutes of the Netherlands* **10**, 91.
Weymann, R. J., Carswell, R. F., Smith, M.G. (1981): Absorption lines in the spectra of quasistellar objects. *Annual Review of Astronomy and Astrophysics* **19**, 41-76.
Wheeler, J. C., Sneden, C., Truran, J. W. Jr. (1989): Abundance ratios as a function of metallicity. *Annual Review of Astronomy and Astrophysics* **27**, 279-349.
White, M., Scott, D., Silk, J. (1994): Anisotropies in the cosmic microwave background. *Annual Review of Astronomy and Astrophysics* **32**, 319-370.
White, S. D. M., et al. (1987): Galaxy distribution in a cold dark matter universe. *Nature* **330**, 451.
Whitford, A. E. (1958): The law of interstellar reddening. *Astronomical Journal* **63**, 201.
Wien, W. (1893): Eine neue Beziehung der Strahlung schwarzer Körper zum zweiten Hauptsatz der Wärmetheorie (One new relation between the radiation of blackbodies and the second law of thermodynamics). *Sitz. Acad. Wiss. Berlin* **1**, 55.
Wien, W. (1894): On the division of energy in the emission-spectrum of a black body. *Philosophical Magazine* **43**, 214.
Wild, J. P., Smerd, S. F. (1972): Radio bursts from the solar corona. *Annual Review of Astronomy and Astrophysics* **10**, 159-196.
Wild, J. P., Smerd, S. F., Weiss, A. A. (1963): Solar bursts. *Annual Review of Astronomy and Astrophysics* **1**, 291-366.
Wildt, R., (1939): Electron affinity in astrophysics. *The Astrophysical Journal* **89**, 295-301. Reproduced in Lang and Gingerich (1979).
Will, C. M. (1993): *Theory and Experiment in Gravitational Physics – Revised Edition.* New York: Cambridge University Press 1993.
Williams, J. P., Cieza, L. A. (2011): Protoplanetary disks and their evolution. *Annual Review of Astronomy and Astrophysics* **49**, 67-117.
Willson, L. A. (2000): Mass loss from cool stars: Impact on the evolution of stars and stellar populations. *Annual Review of Astronomy and Astrophysics* **38**, 573-611.
Wilson, C. T. R. (1911): On a method of making visible ions and tracks of ionizing particles through a gas. *Proceedings of the Royal Society of London, Series A* **95**, 285-288.
Wilson, O. C. (1966): Stellar chromospheres. *Science* **151**, 1487-1498.
Wilson, O. C. (1966): Stellar convection zones, chromospheres, and rotation. *The Astrophysical Journal* **144**, 695-708.
Wilson, O. C. (1978): Chromospheric variations in main-sequence stars. *The Astrophysical Journal* **226**, 379-396.
Wilson, R. W., Jefferts, K. B., Penzias, A. A. (1970): Carbon monoxide in the Orion Nebula. *The Astrophysical Journal* **161**, L43.
Wilson, R. W., Penzias, A. A. (1967): Isotropy of cosmic background radiation at 4080 megahertz. *Science* **156**, 1100.
Wilson, T. L., Rood, R. T. (1994): Abundances in the interstellar medium. *Annual Review of Astronomy and Astrophysics* **32**, 191-226.
Winckel, H. Van (2003): Post-AGB stars. *Annual Review of Astronomy and Astrophysics* **41**, 391-427.
Winget, D. E., *et al*. (1987): An independent method for determining the age of the universe. *Astrophysical Journal Letters* **315**, L77.

Withbroe, G. L., Noyes, R. W. (1977): Mass and energy flow in the solar chromosphere and corona. *Annual Review of Astronomy and Astrophysics* **15**, 363-387.
Wolf, M. (1923): On the dark nebula NGC 6960. *Astronomische Nachrichten* **219**, 109-116. English translation in Lang and Gingerich (1979).
Wolfenstein, L. (1978): Neutrino oscillations in matter. *Physical Review* **D17,** 2369-2374.
Wollaston, W. H. (1802): A method of examining refractive and dispersive power by prismatic reflection. *Philosophical Transactions of the Royal Society of London* **92**, 365-380.
Wood, M. A. (1992): Constraints on the age and evoution of the Galaxy from the white dwarf luminosity function. *The Astrophysical Journal* **386**, 539.
Woodward, P. R. (1978): Theoretical models of star formation. *Annual Review of Astronomy and Astrophysics* **16**, 555-584.
Woosley, S. E., Bloom, J. S. (2006): The supernova-gamma-ray burst connection. *Annual Review of Astronomy and Astrophysics* **44**, 507-556.
Woosley, S. E., Weaver, T. A. (1986): The physics of supernova explosions. *Annual Review of Astronomy and Astrophysics* **24**, 205-253.
Young, C. A. (1869): On a new method of observing contacts at the Sun's limb, and other spectroscopic observations during the recent eclipse. *American Journal of Science and Arts* **48**, 370-378. Reproduced in Meadows (1970).
Yukawa, H. (1935): On the interaction of elementary particles 1, *Proceedings of the Physico-Mathematical Society of Japan* **17**, 48.
Yukawa, H. (1937): On a possible interpretation of the penetrating component of the cosmic ray. *Proceedings of Physics and Mathematics Society of Japan* **19**, 712.
Zanstra, H. (1927): An application of the quantum theory to the luminosity of diffuse nebulae. *The Astrophysical Journal* **65**, 50-70. Reproduced in Lang and Gingerich (1979).
Zanstra H. (1928): Temperatures of stars in planetary nebulae. *Nature* **121**, 790-791.
Zaritsky, D. et al. (1989): Velocities of stars in remote galactic satellites and the mass of the Galaxy. *The Astrophysical Journal* **345**, 759.
Zeeman, P. (1896): On the influence of magnetism on the nature of the light emitted by a substance. *Philosophical Magazine* **43**, 226.
Zeeman, P. (1897a): Doublets and triplets in the spectrum produced by external magnetic forces. I., II. *Philosophical Magazine* **44**, 55, 255.
Zeeman, P. (1897b): The effect of magnetization on the nature of light emitted by a substance. *Nature* **55**, 347. English translation of an article in the *Proceedings of the Physical Society of Berlin*.
Zeldovich, Y. B. (1964): The fate of a star and the evolution of gravitational energy upon accretion. *Dokl. Akad. Nauk SSSR **155**, 67 - March*. English translation in *Sov. Phys. Doklady* **9**, 195 (1964).
Zhang, M., Low, B. C. (2005): The hydromagnetic nature of solar coronal mass ejections. *Annual Review of Astronomy and Astrophysics* **43**, 103-137.
Zhevakin, S. A. (1963): Pulsation theory of variable stars. *Annual Review of Astronomy and Astrophysics* **1**, 367-400.
Zinnecker, H., Yorke, H. W. (2007): Toward understanding massive star formation. *Annual Review of Astronomy and Astrophysics* **45**, 481-563.
Zuckerman, B. (2001): Dusty circumstellar disks. *Annual Review of Astronomy and Astrophysics* **39**, 549-580.
Zuckerman, B., Palmer, P. (1974): Radio radiation from interstellar molecules. *Annual Review of Astronomy and Astrophysics* **12**, 279-313.
Zuckerman, B., Song, I. (2004): Young stars near the Sun. *Annual Review of Astronomy and Astrophysics* **42**, 685-721.
Zurbuchen, T. H. (2007): A new view of the coupling of the Sun and heliosphere. *Annual Review of Astronomy and Astrophysics* **45**, 297-338.
Zwarf, S. F. P., McMillan, S. L.W., Gieles, M. (2010): Young massive star clusters. *Annual Review of Astronomy and Astrophysics* **48**, 431-493.

Zweibel, E. G., Yamada, M. (2009): Magnetic reconnection in astrophysical and laboratory plasmas. *Annual Review of Astronomy and Astrophysics* **47**, 291-332.

Zwicky, F. (1933): Die Rotverschiebund von extragalaktischen Nebeln (The redshift of extragalactic nebulae). *Helvetica Physica Acta* **6**, 110. English translation by Sydney van den Berg in The Early History of Dark Matter, *Publications of the Astronomical Society of the Pacific* **111**, 657-660 (1999).

Zwicky, F. (1937): A super-nova in NGC 4157. *Publications of the Astronomical Society of the Pacific* **49**, 204.

Zwicky, F. (1937): Nebulae as gravitational lenses. *Physical Review* **51**, 290.

Zwicky, F. (1937): On the masses of nebulae and clusters of nebulae. *The Astrophysical Journal* **86**, 217-246. Reproduced in Lang and Gingerich (1979).

Zwicky, F. (1937): On the probability on detecting nebulae which act as gravitational lenses. *Physical Review* **51**, 679.

Appendix I
Constants

Constant	Symbol	Value
Universal constants		
Speed of light in vacuum	c	2.9979×10^{8} m s^{-1}
Universal gravitational constant	G	6.674×10^{-11} N m^{2} kg^{-2}
		6.674×10^{-11} m^{3} kg^{-1} s^{-2}
Planck constant	h	6.6261×10^{-34} J s
Electric constant	ε_0	8.8542×10^{-12} F m^{-1}
Magnetic constant	μ_0	1.2566×10^{-6} N A^{-2}
Thermal radiation		
Boltzmann constant	k	1.38065×10^{-23} J K^{-1}
Stefan-Boltzmann constant	σ	5.6704×10^{-8} J s^{-1} m^{-2} K^{-4}
Radiation density constant	a	7.5657×10^{-16} J m^{-3} K^{-4}
Wien wavelength displacement law constant	b	2.8978×10^{-3} m K
Atomic and nuclear		
Electron mass	m_e	9.1094×10^{-31} kg
Elementary charge	e	1.6022×10^{-19} C
Classical electron radius	r_e	2.8179×10^{-15} m
Thomson scattering cross section	σ_T	6.65246×10^{-29} m^{2}
Atomic mass unit	m_u = u	1.660539×10^{-27} kg
Proton mass	m_p	1.6726×10^{-27} kg
Neutron mass	m_n	1.6749×10^{-27} kg
Alpha particle mass	m_α	6.644656×10^{-27} kg
Bohr radius	a_0	5.2918×10^{-11} m
Rydberg constant	R_∞	10,973,731.5685 m^{-1}
Sun		
Mass of the Sun	$M_\odot$	1.989×10^{30} kg
Luminosity of Sun	$L_\odot$	3.828×10^{26} J s^{-1}
Radius of Sun	$R_\odot$	6.955×10^{8} m
Expanding universe		
Hubble constant	H_0	75 km s^{-1} Mpc^{-1}
Age of expanding Universe	t_0	13.7×10^{9} years

(continued)

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7,

(continued)

Constant	Symbol	Value
Cosmic microwave background radiation		
Temperature	T_{CMB}	2.725 K
Photon density	N_{CMB}	4.10×10^{8} m^{-3}
Anisotropy	$\Delta T/T_{CMB}$	1.1×10^{-5}

The physical constants are accurate to the fourth decimal place. For greater accuracy with the latest values consult http://physics.nist.gov/cuu/constants/index.html

Appendix II
Units

Unit	Symbol	Value
Distance and length		
Ångström	Å	10^{-10} m
Meter	m	1 km = 10^3 m = 0.621371 mile 1 mile = 1.60934 km
Mean Earth-Sun distance	AU	1.49598×10^{11} m
Light-year	ly	9.460528×10^{15} m = 63,239.67 AU
Parsec	pc	3.08568×10^{16} m = 3.26164 light-years = 206,265 AU
Megaparsec	Mpc	10^6 pc
Angle		
Degree	°	$60' = 3600''$ $1^\circ = 0.0174532925$ radians $1'' = 4.8481368 \times 10^{-6}$ radians
(The symbol $'$ denotes minutes of arc, the symbol $''$ designates seconds of arc)		
Radian	rad	$2.06265 \times 10^5\ ''$ $57.2957795^\circ = 360^\circ/(2\pi)$
Pi	π	3.141592654
Time		
Solar day	day	24 h = 86,400 s
Sidereal day	sidereal day	23 h 56 m 04.09 s = 23.9344696 h
Year	tropical year	365.25 solar days = 3.15576×10^7 s
Energy, power, force, pressure		
Joule	J	10^7 erg
Electron-volt	ev	1.6018×10^{-19} J
Power	Watt	J s^{-1}
Force	N	kg m s^{-2}
Pressure	Pa	N m^{-2}

(continued)

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7, © Springer-Verlag Berlin Heidelberg 2013

(continued)

Unit	Symbol	Value
Mass		
Kilogram	kg	1,000 g, 1 metric ton = 10^3 kg
Solar units		
Mass of the Sun	$M_\odot$	1.989×10^{30} kg
Luminosity of Sun	$L_\odot$	3.828×10^{26} J s^{-1}
Radius of Sun	$R_\odot$	6.955×10^8 m

Appendix III
Fundamental Equations

Angular resolution, θ_r, of a telescope of diameter, D_T, at a wavelength, λ:

$$\theta_r = \frac{\lambda}{D_T} \text{ radians} \tag{A-1}$$

where 1 radian = 2.06265×10^5 seconds of arc = 2.06265×10^5 ″.

Angular source extent, θ_{size}, of a celestial source of radius, R, located at a distance, D:

$$\theta_{size} = \frac{2R}{D} \text{radians}, \tag{A-2}$$

where 1 radian = 2.06265×10^5 seconds of arc = 2.06265×10^5 ″.

Wavelength, λ, **frequency**, v, and **speed of light**, c:

$$\lambda \times v = c, \tag{A-3}$$

where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

Photon energy, E, of radiation at frequency v:

$$E = hv \tag{A-4}$$

where the Planck constant $h = 6.6261 \times 10^{-34}$ J s.

Stefan-Boltzmann law for luminosity, L, of thermal radiator with effective temperature T_{eff} and radius R:

$$L = 4\pi\sigma R^2 T_{eff}^4 \tag{A-5}$$

where $\pi = 3.14159$ and the Stefan-Boltzmann constant $\sigma = 5.6704 \times 10^{-8}$ J s^{-1} m^{-2} K^{-4}.

Apparent magnitude, absolute magnitude, and luminosity for a star. Any apparent magnitude, m, can be converted to absolute magnitude, M, through the simple formula:

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7, © Springer-Verlag Berlin Heidelberg 2013

$$\text{absolute magnitude} = M = m + 5 - 5\log D, \tag{A-6}$$

where D is the distance in parsecs, and 1 parsec = 3.26164 light-years = 3.08568×10^{16} m. The absolute magnitude can be converted into a luminosity, L, using:

$$\log\left(\frac{L}{L_\odot}\right) = 0.4(M_\odot - M), \tag{A-7}$$

or

$$L = 10^{0.4(M_\odot - M)}\, L_\odot, \tag{A-8}$$

where the absolute magnitude of the Sun in the visual range of wavelengths, where it is most intense, is $M_\odot = +4.83$ and the absolute luminosity of the Sun is $L_\odot = 3.828 \times 10^{26}$ J s^{-1}. Notice that the symbol $M_\odot$ is used to denote both the absolute magnitude of the Sun, which is used here, and the mass of the Sun, used in other equations. The Sun has an apparent magnitude of $m_\odot = -26.74$.

Wien displacement law for wavelength $\lambda_{\max}$ of maximum intensity for a thermal radiator at temperature T:

$$\lambda_{\max} = \frac{0.002898}{T}\,\text{meters}. \tag{A-9}$$

Radiant flux, f, or **apparent brightness**, of an object of luminosity L at distance D:

$$f = \frac{L}{4\pi D^2}. \tag{A-10}$$

Gravitational force, F_G, between two masses, M_1 and M_2 separated by a distance D between their centers:

$$F_G = \frac{GM_1M_2}{D^2}, \tag{A-11}$$

where the universal gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

Kepler's third law for the orbital period P of a binary system of mass M_1 and M_2 separated by distance a:

$$P^2 = \frac{4\pi^2}{G(M_1 + M_2)}a^3 \tag{A-12}$$

where the universal gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

Jeans mass, M_J, for a spherical gas cloud of radius, R, and temperature, T:

$$M_J = \frac{3kT}{Gm}R, \tag{A-13}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}, the universal gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and m is the gas particle mass. Gravitational collapse occurs if the cloud mass, M, is greater than the Jeans mass, M_J.

Escape velocity, V_{esc}, at a distance R from a mass M:

$$V_{esc} = \left(\frac{2GM}{R}\right)^{1/2}, \tag{A-14}$$

where the universal gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

Doppler effect for a change $\Delta\lambda$ in the wavelength λ due to a radial velocity V_r of a source moving away from observer:

$$\frac{\Delta\lambda}{\lambda_{emitted}} = \frac{\lambda_{observed} - \lambda_{emitted}}{\lambda_{emitted}} = \frac{V_r}{c} \quad \text{for } V_r \ll c, \tag{A-15}$$

where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

Parallax the annual parallax π_A, of a star at distance, D, is:

$$\pi_A = \text{AU/D radians}, \tag{A-16}$$

where 1 AU $= 1.49598 \times 10^{11}$ m and 1 radian $= 2.06265 \times 10^5$ s of arc. When the parallax is given in units of seconds of arc, then the distance, D, is given by:

$$D = \frac{1}{\pi_A} \text{ parsecs}, \tag{A-17}$$

where 1 parsec = 3.26164 light years = 206,265.8 AU.

Gravitational potential energy of a mass, M, with radius, R:

$$\text{Gravitational potential energy} = \frac{GM^2}{R}, \tag{A-18}$$

where the universal gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

Kinetic energy of mass, m, moving at velocity, V:

$$\text{Kinetic energy} = \frac{1}{2}mV^2. \tag{A-19}$$

Thermal energy at temperature, T:

$$\text{Thermal energy} = \frac{3}{2}kT, \tag{A-20}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}.

Thermal velocity, $V_{thermal}$, of a particle of mass, m, at temperature, T:

$$V_{thermal} = \sqrt{\frac{3kT}{m}} = \left[\frac{3kT}{m}\right]^{1/2}, \tag{A-21}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}.

Gas pressure, P_G, of particles of number density, N, and temperature T:

$$P_G = NkT, \tag{A-22}$$

where the Boltzmann constant $k = 1.38065 \times 10^{-23}$ J K^{-1}.

Radiation pressure, P_r, for a temperature T:

$$P_r = \frac{aT^4}{3} \tag{A-23}$$

where the radiation constant $a = 7.5657 \times 10^{-16}$ J m^{-3} K^{-4}.

Magnetic pressure, P_B, of a magnetic field of strength, B:

$$P_B = \frac{B^2}{2\mu_0}, \tag{A-24}$$

where the magnetic constant $\mu_0 = 1.2566 \times 10^{-6}$ N A^{-2}.

Energy radiated, ΔE, **by a mass loss**, Δm, during nuclear reactions:

$$\Delta E = \Delta m\, c^2, \tag{A-25}$$

where the speed of light $c = 2.9979 \times 10^8$ m s^{-1}.

Schwarzschild radius, R_{sch}, of a mass, M, of radius R:

$$R_{\text{sch}} = \frac{2GM}{c^2} = 2.95 \times 10^3 \left(\frac{M}{M_\odot}\right) \text{m}, \tag{A-26}$$

where the universal gravitational constant $G = 6.674 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and the Sun's mass $M_\odot = 1.989 \times 10^{30}$ kg.

Hubble law for the recession velocity V_r of a galaxy at a distance D:

$$V_r = H_0 \times D, \tag{A-27}$$

where the Hubble constant $H_0 = 75$ km s^{-1} Mpc^{-1} and 1 Mpc $= 3.08568 \times 10^{22}$ m.

Author Index

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7, © Springer-Verlag Berlin Heidelberg 2013

Subject Index

K. R. Lang, *Essential Astrophysics*, Undergraduate Lecture Notes in Physics,
DOI: 10.1007/978-3-642-35963-7, © Springer-Verlag Berlin Heidelberg 2013