Astronomy

Astronomy, science that encompasses the study of all extraterrestrial objects and phenomena. Until the invention of the telescope and the discovery of the laws of motion and gravity in the 17th century, astronomy was primarily concerned with noting and predicting the positions of the Sun, Moon, and planets, originally for calendrical and astrological purposes and later for navigational uses and scientific interest. The catalog of objects now studied is much broader and includes, in order of increasing distance, the solar system, the stars that make up the Milky Way Galaxy, and other, more distant galaxies. With the advent of scientific space probes, Earth also has come to be studied as one of the planets, though its more-detailed investigation remains the domain of the Earth sciences.



Hubble Space Telescope

Hubble Space Telescope, photographed by the space shuttle Discovery.

**The Scope Of Astronomy**

Since the late 19th century astronomy has expanded to include astrophysics, the application of physical and chemical knowledge to an understanding of the nature of celestial objects and the physical processes that control their formation, evolution, and emission of radiation. In addition, the gases and dust particles around and between the stars have become the subjects of much research. Study of the nuclear reactions that provide the energy radiated by stars has shown how the diversity of atoms found in nature can be derived from a universe that, following the first few minutes of its existence, consisted only of hydrogen, helium, and a trace of lithium. Concerned with phenomena on the largest scale is cosmology, the study of the evolution of the universe. Astrophysics has transformed cosmology from a purely speculative activity to a modern science capable of predictions that can be tested.

Its great advances notwithstanding, astronomy is still subject to a major constraint: it is inherently an observational rather than an experimental science. Almost all measurements must be performed at great distances from the objects of interest, with no control over such quantities as their temperature, pressure, or chemical composition. There are a few exceptions to this limitation—namely, meteorites (most of which are from the asteroid belt, though some are from the Moon or Mars), rock and soil samples brought back from the Moon, samples of comet and asteroid dust returned by robotic spacecraft, and interplanetary dust particles collected in or above the stratosphere. These can be examined with laboratory techniques to provide information that cannot be obtained in any other way. In the future, space missions may return surface materials from Mars, or other objects, but much of astronomy appears otherwise confined to Earth-based observations augmented by observations from orbiting satellites and long-range space probes and supplemented by theory.

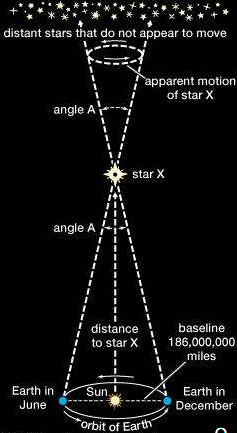


nickel-iron meteorite

Nickel-iron meteorite, from Canyon Diablo, Arizona.

**Determining astronomical distances**

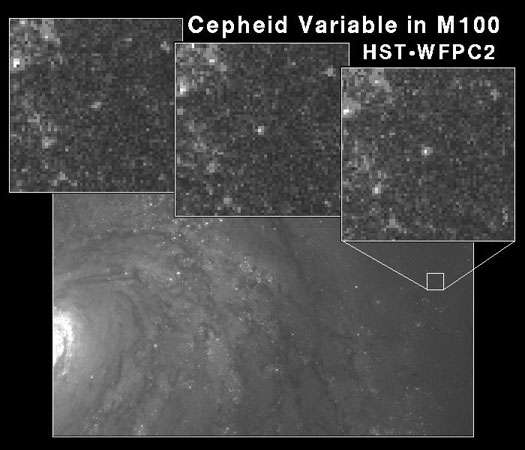
A central undertaking in astronomy is the determination of distances. Without a knowledge of astronomical distances, the size of an observed object in space would remain nothing more than an angular diameter and the brightness of a star could not be converted into its true radiated power, or luminosity. Astronomical distance measurement began with a knowledge of Earth’s diameter, which provided a base for triangulation. Within the inner solar system, some distances can now be better determined through the timing of radar reflections or, in the case of the Moon, through laser ranging. For the outer planets, triangulation is still used. Beyond the solar system, distances to the closest stars are determined through triangulation, in which the diameter of Earth’s orbit serves as the baseline and shifts in stellar parallax are the measured quantities. Stellar distances are commonly expressed by astronomers in parsecs (pc), kiloparsecs, or megaparsecs. (1 pc = 3.086 × 1018 cm, or about 3.26 light-years [1.92 × 1013 miles].) Distances can be measured out to around a kiloparsec by trigonometric parallax. The accuracy of measurements made from Earth’s surface is limited by atmospheric effects, but measurements made from the Hipparcos satellite in the 1990s extended the scale to stars as far as 650 parsecs, with an accuracy of about a thousandth of an arc second. The Gaia satellite is expected to measure stars as far away as 10 kiloparsecs to an accuracy of 20 percent. Less-direct measurements must be used for more-distant stars and for galaxies.



stellar distances

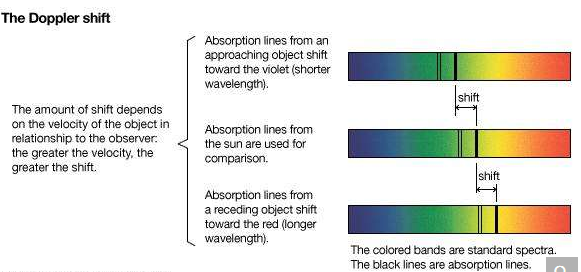
**Calculating stellar distances**

Two general methods for determining galactic distances are described here. In the first, a clearly identifiable type of star is used as a reference standard because its luminosity has been well determined. This requires observation of such stars that are close enough to Earth that their distances and luminosities have been reliably measured. Such a star is termed a “standard candle.” Examples are Cepheid variables, whose brightness varies periodically in well-documented ways, and certain types of supernova explosions that have enormous brilliance and can thus be seen out to very great distances. Once the luminosities of such nearer standard candles have been calibrated, the distance to a farther standard candle can be calculated from its calibrated luminosity and its actual measured intensity. (The measured intensity [I] is related to the luminosity [L] and distance [d] by the formula I = L/4πd2.) A standard candle can be identified by means of its spectrum or the pattern of regular variations in brightness. (Corrections may have to be made for the absorption of starlight by interstellar gas and dust over great distances.) This method forms the basis of measurements of distances to the closest galaxies.



A region of the spiral galaxy M100 (bottom), with three frames (top) showing a Cepheid variable increasing in brightness. These images were taken with the Wide Field Planetary Camera 2 (WFPC2) on board the Hubble Space Telescope (HST).

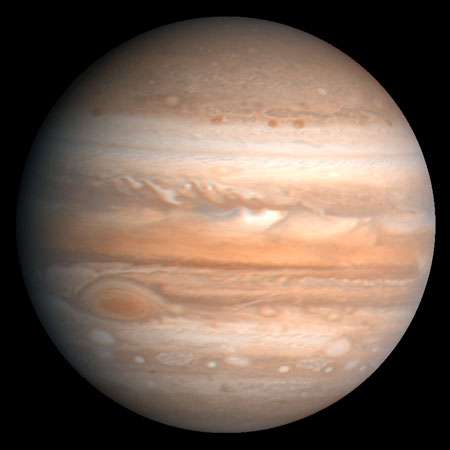
The second method for galactic distance measurements makes use of the observation that the distances to galaxies generally correlate with the speeds with which those galaxies are receding from Earth (as determined from the Doppler shift in the wavelengths of their emitted light). This correlation is expressed in the Hubble law: velocity = H × distance, in which H denotes Hubble’s constant, which must be determined from observations of the rate at which the galaxies are receding. There is widespread agreement that H lies between 67 and 73 kilometres per second per megaparsec (km/sec/Mpc). H has been used to determine distances to remote galaxies in which standard candles have not been found. (For additional discussion of the recession of galaxies, the Hubble law, and galactic distance determination)



Doppler shift.

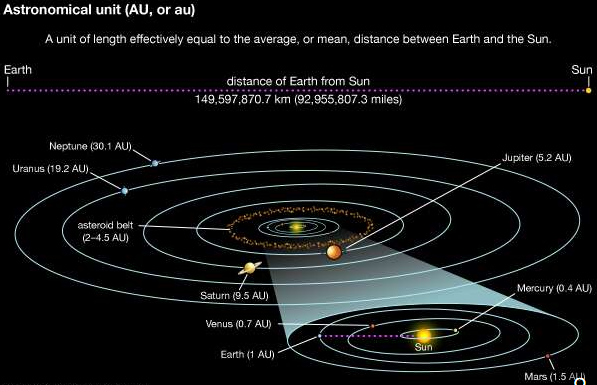
**Study of the solar system**

The solar system took shape 4.57 billion years ago, when it condensed within a large cloud of gas and dust. Gravitational attraction holds the planets in their elliptical orbits around the Sun. In addition to Earth, five major planets (Mercury, Venus, Mars, Jupiter, and Saturn) have been known from ancient times. Since then only two more have been discovered: Uranus by accident in 1781 and Neptune in 1846 after a deliberate search following a theoretical prediction based on observed irregularities in the orbit of Uranus. Pluto, discovered in 1930 after a search for a planet predicted to lie beyond Neptune, was considered a major planet until 2006, when it was redesignated a dwarf planet by the International Astronomical Union.



Jupiter, the fifth planet from the Sun and largest planet in the solar system. The Great Red Spot is visible in the lower left. This image is based on observations made by the Voyager 1 spacecraft in 1979.

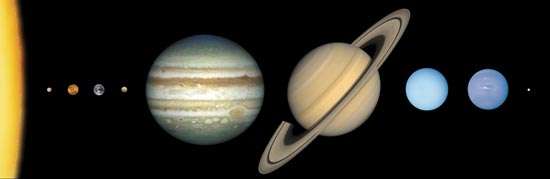
The average Earth-Sun distance, which originally defined the astronomical unit (AU), provides a convenient measure for distances within the solar system. The astronomical unit was originally defined by observations of the mean radius of Earth’s orbit but is now defined as 149,597,870.7 km (about 93 million miles). Mercury, at 0.4 AU, is the closest planet to the Sun, while Neptune, at 30.1 AU, is the farthest. Pluto’s orbit, with a mean radius of 39.5 AU, is sufficiently eccentric that at times it is closer to the Sun than is Neptune. The planes of the planetary orbits are all within a few degrees of the ecliptic, the plane that contains Earth’s orbit around the Sun. As viewed from far above Earth’s North Pole, all planets move in the same (counterclockwise) direction in their orbits.



scale of the solar system

Distances in the solar system as measured in astronomical units (AU).

Most of the mass of the solar system is concentrated in the Sun, with its 1.99 × 1033 grams. Together, all of the planets amount to 2.7 × 1030 grams (i.e., about one-thousandth of the Sun’s mass), and Jupiter alone accounts for 71 percent of this amount. The solar system also contains five known objects of intermediate size classified as dwarf planets and a very large number of much smaller objects collectively called small bodies. The small bodies, roughly in order of decreasing size, are the asteroids, or minor planets; comets, including Kuiper belt, Centaur, and Oort cloud objects; meteoroids; and interplanetary dust particles. Because of their starlike appearance when discovered, the largest of these bodies were termed asteroids, and that name is widely used, but, now that the rocky nature of these bodies is understood, their more descriptive name is minor planets.



The eight planets of the solar system and Pluto, in a montage of images scaled to show the approximate sizes of the bodies relative to one another. Outward from the Sun, which is represented to scale by the yellow segment at the extreme left, are the four rocky terrestrial planets (Mercury, Venus, Earth, and Mars), the four hydrogen-rich giant planets (Jupiter, Saturn, Uranus, and Neptune), and icy, comparatively tiny Pluto.

The eight planets of the solar system and Pluto, in a montage of images scaled to show the approximate sizes of the bodies relative to one another. Outward from the Sun, which is represented to scale by the yellow segment at the extreme left, are the four rocky terrestrial planets (Mercury, Venus, Earth, and Mars), the four hydrogen-rich giant planets (Jupiter, Saturn, Uranus, and Neptune), and icy, comparatively tiny Pluto.

The four inner, terrestrial planets—Mercury, Venus, Earth, and Mars—along with the Moon have average densities in the range of 3.9–5.5 grams per cubic cm, setting them apart from the four outer, giant planets—Jupiter, Saturn, Uranus, and Neptune—whose densities are all close to 1 gram per cubic cm, the density of water. The compositions of these two groups of planets must therefore be significantly different. This dissimilarity is thought to be attributable to conditions that prevailed during the early development of the solar system. Planetary temperatures now range from around 170 °C (330 °F, 440 K) on Mercury’s surface through the typical 15 °C (60 °F, 290 K) on Earth to −135 °C (−210 °F, 140 K) on Jupiter near its cloud tops and down to −210 °C (−350 °F, 60 K) near Neptune’s cloud tops. These are average temperatures; large variations exist between dayside and nightside for planets closest to the Sun, except for Venus with its thick atmosphere.



Saturn and its spectacular rings, in a natural-colour composite of 126 images taken by the Cassini spacecraft on October 6, 2004. The view is directed toward Saturn's southern hemisphere, which is tipped toward the Sun. Shadows cast by the rings are visible against the bluish northern hemisphere, while the planet's shadow is projected on the rings to the left.

The surfaces of the terrestrial planets and many satellites show extensive cratering, produced by high-speed impacts (see meteorite crater). On Earth, with its large quantities of water and an active atmosphere, many of these cosmic footprints have eroded, but remnants of very large craters can be seen in aerial and spacecraft photographs of the terrestrial surface. On Mercury, Mars, and the Moon, the absence of water and any significant atmosphere has left the craters unchanged for billions of years, apart from disturbances produced by infrequent later impacts. Volcanic activity has been an important force in the shaping of the surfaces of the Moon and the terrestrial planets. Seismic activity on the Moon has been monitored by means of seismometers left on its surface by Apollo astronauts and by Lunokhod robotic rovers. Cratering on the largest scale seems to have ceased about three billion years ago, although on the Moon there is clear evidence for a continued cosmic drizzle of small particles, with the larger objects churning (“gardening”) the lunar surface and the smallest producing microscopic impact pits in crystals in the lunar rocks.



Mercury

Meteorite crater surrounded by rays of ejected material on Mercury, in a photograph taken by the Messenger probe, January 14, 2008. A chain of craters crosses the centre of the rayed crater.

All of the planets apart from the two closest to the Sun (Mercury and Venus) have natural satellites (moons) that are very diverse in appearance, size, and structure, as revealed in close-up observations from long-range space probes. The four outer dwarf planets have moons; Pluto has at least five moons, including one, Charon, fully half the size of Pluto itself. Over 200 asteroids and 80 Kuiper belt objects also have moons. Four planets (Jupiter, Saturn, Uranus, and Neptune), one dwarf planet (Haumea), and one Centaur object (Chariklo) have rings, disklike systems of small rocks and particles that orbit their parent bodies.



Charon

Pluto's moon Charon, in a photo taken by the New Horizons spacecraft, July 13, 2015.

**Lunar exploration**

During the U.S. Apollo missions a total weight of 381.7 kg (841.5 pounds) of lunar material was collected; an additional 300 grams (0.66 pounds) was brought back by unmanned Soviet Luna vehicles. About 15 percent of the Apollo samples have been distributed for analysis, with the remainder stored at the NASA Johnson Space Center, Houston, Texas. The opportunity to employ a wide range of laboratory techniques on these lunar samples has revolutionized planetary science. The results of the analyses have enabled investigators to determine the composition and age of the lunar surface. Seismic observations have made it possible to probe the lunar interior. In addition, retroreflectors left on the Moon’s surface by Apollo astronauts have allowed high-power laser beams to be sent from Earth to the Moon and back, permitting scientists to monitor the Earth-Moon distance to an accuracy of a few centimetres. This experiment, which has provided data used in calculations of the dynamics of the Earth-Moon system, has shown that the separation of the two bodies is increasing by 4.4 cm (1.7 inches) each year. (For additional information on lunar studies.)



Apollo 17 geologist-astronaut Harrison Schmitt at the foot of a huge split boulder, December 13, 1972, during the mission's third extravehicular exploration of the Taurus-Littrow Valley landing site.

**Planetary studies**

Mercury is too hot to retain an atmosphere, but Venus’s brilliant white appearance is the result of its being completely enveloped in thick clouds of carbon dioxide, impenetrable at visible wavelengths. Below the upper clouds, Venus has a hostile atmosphere containing clouds of sulfuric acid droplets. The cloud cover shields the planet’s surface from direct sunlight, but the energy that does filter through warms the surface, which then radiates at infrared wavelengths. The long-wavelength infrared radiation is trapped by the dense clouds such that an efficient greenhouse effect keeps the surface temperature near 465 °C (870 °F, 740 K). Radar, which can penetrate the thick Venusian clouds, has been used to map the planet’s surface. In contrast, the atmosphere of Mars is very thin and is composed mostly of carbon dioxide (95 percent), with very little water vapour; the planet’s surface pressure is only about 0.006 that of Earth. The outer planets have atmospheres composed largely of light gases, mainly hydrogen and helium.

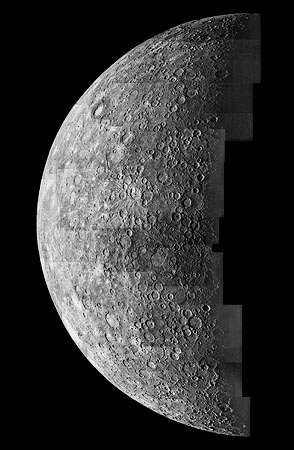
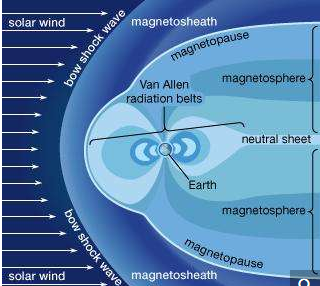


Photo mosaic of Mercury, taken by the Mariner 10 spacecraft, 1974.

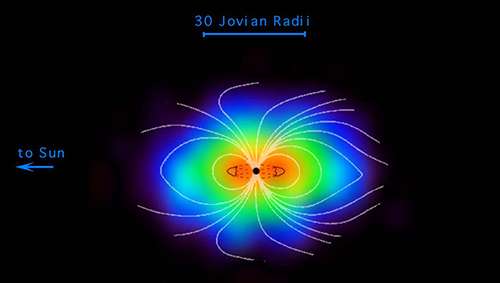
Each planet rotates on its axis, and nearly all of them rotate in the same direction—counterclockwise as viewed from above the ecliptic. The two exceptions are Venus, which rotates in the clockwise direction beneath its cloud cover, and Uranus, which has its rotation axis very nearly in the plane of the ecliptic.

Some of the planets have magnetic fields. Earth’s field extends outward until it is disturbed by the solar wind—an outward flow of protons and electrons from the Sun—which carries a magnetic field along with it. Through processes not yet fully understood, particles from the solar wind and galactic cosmic rays (high-speed particles from outside the solar system) populate two doughnut-shaped regions called the Van Allen radiation belts. The inner belt extends from about 1,000 to 5,000 km (600 to 3,000 miles) above Earth’s surface, and the outer from roughly 15,000 to 25,000 km (9,300 to 15,500 miles). In these belts, trapped particles spiral along paths that take them around Earth while bouncing back and forth between the Northern and Southern hemispheres, with their orbits controlled by Earth’s magnetic field. During periods of increased solar activity, these regions of trapped particles are disturbed, and some of the particles move down into Earth’s atmosphere, where they collide with atoms and molecules to produce auroras.



The Van Allen radiation belts contained within Earth's magnetosphere. Pressure from the solar wind is responsible for the asymmetrical shape of the magnetosphere and the belts.

Jupiter has a magnetic field far stronger than Earth’s and many more trapped electrons, whose synchrotron radiation (electromagnetic radiation emitted by high-speed charged particles that are forced to move in curved paths, as under the influence of a magnetic field) is detectable from Earth. Bursts of increased radio emission are correlated with the position of Io, the innermost of the four Galilean moons of Jupiter. Saturn has a magnetic field that is much weaker than Jupiter’s, but it too has a region of trapped particles. Mercury has a weak magnetic field that is only about 1 percent as strong as Earth’s and shows no evidence of trapped particles. Uranus and Neptune have fields that are less than one-tenth the strength of Saturn’s and appear much more complex than that of Earth. No field has been detected around Venus or Mars.



Jupiter's magnetosphere

Jupiter's magnetosphere as observed by the Cassini spacecraft in 2000. The magnetosphere is the largest object in the solar system.

**Investigations of the smaller bodies**

More than 500,000 asteroids with well-established orbits are known, and thousands of additional objects are discovered each year. Hundreds of thousands more have been seen, but their orbits have not been as well determined. It is estimated that several million asteroids exist, but most are small, and their combined mass is estimated to be less than a thousandth that of Earth. Most of the asteroids have orbits close to the ecliptic and move in the asteroid belt, between 2.3 and 3.3 AU from the Sun. Because some asteroids travel in orbits that can bring them close to Earth, there is a possibility of a collision that could have devastating results (see Earth impact hazard).



Gaspra

Asteroid Gaspra, composite of two images taken by the Galileo spacecraft on October 29, 1991. Galileo observed some 600 impact craters on Gaspra, from the large concavity visible on the lower right to craters as small as 100 metres (330 feet) in diameter. The asteroid's irregular shape and fracture lines suggest that it was once part of a larger body.

Comets are considered to come from a vast reservoir, the Oort cloud, orbiting the Sun at distances of 20,000–50,000 AU or more and containing trillions of icy objects—latent comet nuclei—with the potential to become active comets. Many comets have been observed over the centuries. Most make only a single pass through the inner solar system, but some are deflected by Jupiter or Saturn into orbits that allow them to return at predictable times. Halley’s Comet is the best known of these periodic comets; its next return into the inner solar system is predicted for 2061. Many short-period comets are thought to come from the Kuiper belt, a region lying mainly between 30 AU and 50 AU from the Sun—beyond Neptune’s orbit but including part of Pluto’s—and housing perhaps hundreds of millions of comet nuclei. Very few comet masses have been well determined, but most are probably less than 1018 grams, one-billionth the mass of Earth.



Composite image of the fragments of Comet Shoemaker-Levy 9, taken by the Hubble Space Telescope, 1994.

Since the 1990s more than a thousand comet nuclei in the Kuiper belt have been observed with large telescopes; a few are about half the size of Pluto, and Pluto is the largest Kuiper belt object. Pluto’s orbital and physical characteristics had long caused it to be regarded as an anomaly among the planets. However, after the discovery of numerous other Pluto-like objects beyond Neptune, Pluto was seen to be no longer unique in its “neighbourhood” but rather a giant member of the local population. Consequently, in 2006 astronomers at the general assembly of the International Astronomical Union elected to create the new category of dwarf planets for objects with such qualifications. Pluto, Eris, and Ceres, the latter being the largest member of the asteroid belt, were given this distinction. Two other Kuiper belt objects, Makemake and Haumea, were also designated as dwarf planets.

**Pluto**



Pluto as seen by the New Horizons spacecraft, July 14, 2015.

Smaller than the observed asteroids and comets are the meteoroids, lumps of stony or metallic material believed to be mostly fragments of asteroids. Meteoroids vary from small rocks to boulders weighing a ton or more. A relative few have orbits that bring them into Earth’s atmosphere and down to the surface as meteorites. Most meteorites that have been collected on Earth are probably from asteroids. A few have been identified as being from the Moon, Mars, or the asteroid Vesta.



Chelyabinsk meteorite of 2013

A cloud trail left behind by a meteorite that later exploded over Chelyabinsk province, Russia, February 15, 2013.

Meteorites are classified into three broad groups: stony (chondrites and achondrites; about 94 percent), iron (5 percent), and stony-iron (1 percent). Most meteoroids that enter the atmosphere heat up sufficiently to glow and appear as meteors, and the great majority of these vaporize completely or break up before they reach the surface. Many, perhaps most, meteors occur in showers (see meteor shower) and follow orbits that seem to be identical with those of certain comets, thus pointing to a cometary origin. For example, each May, when Earth crosses the orbit of Halley’s Comet, the Eta Aquarid meteor shower occurs. Micrometeorites (interplanetary dust particles), the smallest meteoroidal particles, can be detected from Earth-orbiting satellites or collected by specially equipped aircraft flying in the stratosphere and returned for laboratory inspection. Since the late 1960s numerous meteorites have been found in the Antarctic on the surface of stranded ice flows (see Antarctic meteorites). Some meteorites contain microscopic crystals whose isotopic proportions are unique and appear to be dust grains that formed in the atmospheres of different stars.



Hoba meteorite, lying where it was discovered in 1920 in Grootfontein, Namibia. The object, the largest meteorite known and an iron meteorite by classification, is made of nickel-iron alloy and estimated to weigh nearly 60 tons.

**Determinations of age and chemical composition**

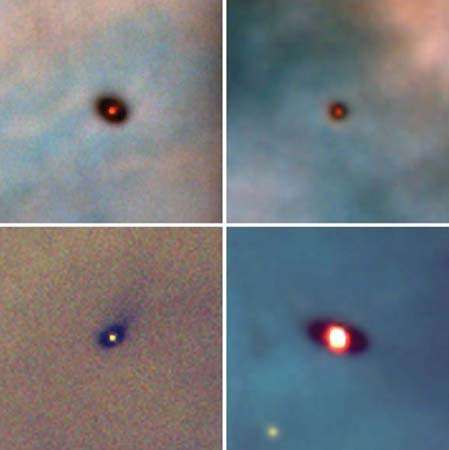
The age of the solar system, taken to be close to 4.6 billion years, has been derived from measurements of radioactivity in meteorites, lunar samples, and Earth’s crust. Abundances of isotopes of uranium, thorium, and rubidium and their decay products, lead and strontium, are the measured quantities.

Assessment of the chemical composition of the solar system is based on data from Earth, the Moon, and meteorites as well as on the spectral analysis of light from the Sun and planets. In broad outline, the solar system abundances of the chemical elements decrease with increasing atomic weight. Hydrogen atoms are by far the most abundant, constituting 91 percent; helium is next, with 8.9 percent; and all other types of atoms together amount to only 0.1 percent.

**Theories of origin**

The origin of Earth, the Moon, and the solar system as a whole is a problem that has not yet been settled in detail. The Sun probably formed by condensation of the central region of a large cloud of gas and dust, with the planets and other bodies of the solar system forming soon after, their composition strongly influenced by the temperature and pressure gradients in the evolving solar nebula. Less-volatile materials could condense into solids relatively close to the Sun to form the terrestrial planets. The abundant, volatile lighter elements could condense only at much greater distances to form the giant gas planets.

Four recently formed stars in the Orion nebula show what the solar system might have looked like at its birth some 4.6 billion years ago. Gravity collapsed dense globules of dust and gas into each of the four stars. The outer material formed into the surrounding protoplanetary disks, which are believed to be the precursors of planets. The bright glow in each image is the star, and the dark oval is the protoplanetary disk seen at an angle.



Four recently formed stars in the Orion nebula show what the solar system might have looked like at its birth some 4.6 billion years ago. Gravity collapsed dense globules of dust and gas into each of the four stars. The outer material formed into the surrounding protoplanetary disks, which are believed to be the precursors of planets. The bright glow in each image is the star, and the dark oval is the protoplanetary disk seen at an angle.

In the1990s astronomers confirmed that other stars have one or more planets revolving around them. Studies of these planetary systems have both supported and challenged astronomers’ theoretical models of how Earth’s solar system formed. Unlike the solar system, many extrasolar planetary systems have large gas giants like Jupiter orbiting very close to their stars, and in some cases these “hot Jupiters” are closer to their star than Mercury is to the Sun.

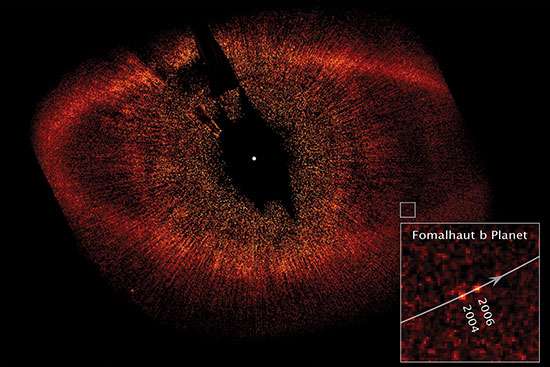
That so many gas giants, which form in the outer regions of their system, end up so close to their stars suggests that gas giants migrate and that such migration may have happened in the solar system’s history. According to the Grand Tack hypothesis, Jupiter may have done so within a few million years of the solar system’s formation. In this scenario, Jupiter is the first giant planet to form, at about 3 AU from the Sun. Drag from the protoplanetary disk causes it to fall inward to about 1.5 AU. However, by this time, Saturn begins to form at about 3 AU and captures Jupiter in a 3:2 resonance. (That is, for every three revolutions Jupiter makes, Saturn makes two.) The two planets migrate outward and clear away any material that would have gone to making Mars bigger. Mars should be bigger than Venus or Earth, but it is only half their size. The Grand Tack, in which Jupiter moves inward and then outward, explains Mars’s small size.

About 500 million years after the Grand Tack, according to the Nice Model (named after the French city where it was first proposed), after the four giant planets—Jupiter, Saturn, Uranus, and Neptune—formed, they orbited 5–17 AU from the Sun. These planets were in a disk of smaller bodies called planetesimals and in orbital resonances with each other. About four billion years ago, gravitational interactions with the planetesimals increased the eccentricity of the planets’ orbits, driving them out of resonance. Saturn, Uranus and Neptune migrated outward, and Jupiter migrated slightly inward. (Uranus and Neptune may even have switched places.) This migration scattered the disk, causing the Late Heavy Bombardment. The final remnant of the disk became the Kuiper belt.

The origin of the planetary satellites is not entirely settled. As to the origin of the Moon, the opinion of astronomers long oscillated between theories that saw its origin and condensation as simultaneous with the formation of Earth and those that posited a separate origin for the Moon and its later capture by Earth’s gravitational field. Similarities and differences in abundances of the chemical elements and their isotopes on Earth and the Moon challenged each group of theories. Finally, in the 1980s a model emerged that gained the support of most lunar scientists—that of a large impact on Earth and the expulsion of material that subsequently formed the Moon. For the outer planets, with their multiple satellites, many very small and quite unlike one another, the picture is less clear. Some of these moons have relatively smooth icy surfaces, whereas others are heavily cratered; at least one, Jupiter’s Io, is volcanic. Some of the moons may have formed along with their parent planets, and others may have formed elsewhere and been captured.

**Study of extrasolar planetary systems**

The first extrasolar planets were discovered in 1992, and more than 3,700 such planets are known. Over 600 of these systems have more than one planet. Because planets are much fainter than their stars, fewer than 100 have been imaged directly. Most extrasolar planets have been found through their transit, the small dimming of a star’s light when a planet passes in front of it.



The extrasolar planet Fomalhaut b in images taken by the Hubble Space Telescope in 2004 and 2006. The black spot at the centre of the image is a coronagraph used to block the light from Fomalhaut, which is located at the white dot. The oval ring is Fomalhaut's dust belt, and the lines radiating from the centre of the image are scattered starlight.

Many of these planets are unlike those of the solar system. Hot Jupiters are large gas giants that orbit very close to their star. For example, HD 209458b is 0.69 times the mass of Jupiter and orbits its star every 3.52 days. Hot Neptunes are large ice giants about 10 percent of Jupiter’s mass that also orbit very close to their star. Super-Earths are planets that are likely rocky like Earth but several times larger.



Artist's conception of the extrasolar planet HD 209458 b, some 150 light-years from Earth.

A primary goal of extrasolar planet research has been finding another planet that could support life. A useful guide for finding a life-supporting planet has been the concept of a habitable zone, the distance from a star where liquid water could survive on a planet’s surface. About 40 planets have been found that are roughly Earth-sized and orbit in a habitable zone.



Kepler-452b

**Study of the stars**

**Measuring observable stellar properties**

The measurable quantities in stellar astrophysics include the externally observable features of the stars: distance, temperature, radiation spectrum and luminosity, composition (of the outer layers), diameter, mass, and variability in any of these. Theoretical astrophysicists use these observations to model the structure of stars and to devise theories for their formation and evolution. Positional information can be used for dynamical analysis, which yields estimates of stellar masses.



Pleiades

Bright nebulosity in the Pleiades (M45, NGC 1432), distance 490 light-years. Cluster stars provide the light, and surrounding clouds of dust reflect and scatter the rays from the stars.

In a system dating back at least to the Greek astronomer-mathematician Hipparchus in the 2nd century BCE, apparent stellar brightness (m) is measured in magnitudes. Magnitudes are now defined such that a first-magnitude star is 100 times brighter than a star of sixth magnitude. The human eye cannot see stars fainter than about sixth magnitude, but modern instruments used with large telescopes can record stars as faint as about 30th magnitude. By convention, the absolute magnitude (M) is defined as the magnitude that a star would appear to have if it were located at a standard distance of 10 parsecs. These quantities are related through the expression m − M = 5 log10 r − 5, in which r is the star’s distance in parsecs.

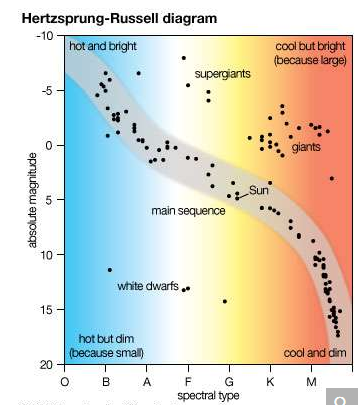
The magnitude scale is anchored on a group of standard stars. An absolute measure of radiant power is luminosity, which is related to the absolute magnitude and usually expressed in ergs per second (ergs/sec). (Sometimes the luminosity is stated in terms of the solar luminosity, 3.86 × 1033 ergs/sec.) Luminosity can be calculated when m and r are known. Correction might be necessary for the interstellar absorption of starlight.

There are several methods for measuring a star’s diameter. From the brightness and distance, the luminosity (L) can be calculated, and, from observations of the brightness at different wavelengths, the temperature (T) can be calculated. Because the radiation from many stars can be well approximated by a Planck blackbody spectrum, these measured quantities can be related through the expression L = 4πR2σT4, thus providing a means of calculating R, the star’s radius. In this expression, σ is the Stefan-Boltzmann constant, 5.67 × 10−5 ergs/cm2K4sec, in which K is the temperature in kelvins. (The radius R refers to the star’s photosphere, the region where the star becomes effectively opaque to outside observation.) Stellar angular diameters can be measured through interferometry—that is, the combining of several telescopes together to form a larger instrument that can resolve sizes smaller than those that an individual telescope can resolve. Alternatively, the intensity of the starlight can be monitored during occultation by the Moon, which produces diffraction fringes whose pattern depends on the angular diameter of the star. Stellar angular diameters of several milliarcseconds can be measured.

Many stars occur in binary systems, in which the two partners orbit their mutual centre of mass. Such a system provides the best measurement of stellar masses. The period (P) of a binary system is related to the masses of the two stars (m1 and m2) and the orbital semimajor axis (mean radius; a) via Kepler’s third law: P2 = 4π2a3/G(m1 + m2). (G is the universal gravitational constant.) From diameters and masses, average values of the stellar density can be calculated and thence the central pressure. With the assumption of an equation of state, the central temperature can then be calculated. For example, in the Sun the central density is 158 grams per cubic cm; the pressure is calculated to be more than one billion times the pressure of Earth’s atmosphere at sea level and the temperature around 15 million K (27 million °F). At this temperature, all atoms are ionized, and so the solar interior consists of a plasma, an ionized gas with hydrogen nuclei (i.e., protons), helium nuclei, and electrons as major constituents. A small fraction of the hydrogen nuclei possess sufficiently high speeds that, on colliding, their electrostatic repulsion is overcome, resulting in the formation, by means of a set of fusion reactions, of helium nuclei and a release of energy. Some of this energy is carried away by neutrinos, but most of it is carried by photons to the surface of the Sun to maintain its luminosity.

Other stars, both more and less massive than the Sun, have broadly similar structures, but the size, central pressure and temperature, and fusion rate are functions of the star’s mass and composition. The stars and their internal fusion (and resulting luminosity) are held stable against collapse through a delicate balance between the inward pressure produced by gravitational attraction and the outward pressure supplied by the photons produced in the fusion reactions.

Stars that are in this condition of hydrostatic equilibrium are termed main-sequence stars, and they occupy a well-defined band on the Hertzsprung-Russell (H-R) diagram, in which luminosity is plotted against colour index or temperature. Spectral classification, based initially on the colour index, includes the major spectral types O, B, A, F, G, K and M, each subdivided into 10 parts. Temperature is deduced from broadband spectral measurements in several standard wavelength intervals. Measurement of apparent magnitudes in two spectral regions, the B and V bands (centred on 4350 and 5550 angstroms, respectively), permits calculation of the colour index, CI = mB − mV, from which the temperature can be calculated.

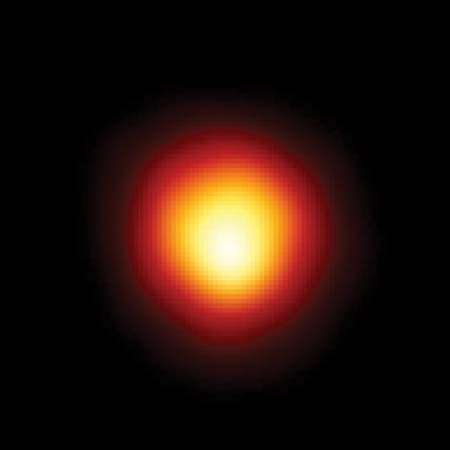


Hertzsprung-Russell diagram

Hertzsprung-Russell diagram. Spectral type (a measure of a star's temperature), following the order introduced by American astronomer Annie Jump Cannon, is plotted on the horizontal axis, and absolute magnitude (the intrinsic brightness of a star) is plotted on the vertical axis.

For a given temperature, there are stars that are much more luminous than main-sequence stars. Given the dependence of luminosity on the square of the radius and the fourth power of the temperature (R2T4 of the luminosity expression above), greater luminosity implies larger radius, and such stars are termed giant stars or supergiant stars. Conversely, stars with luminosities much less than those of main-sequence stars of the same temperature must be smaller and are termed white dwarf stars. Surface temperatures of white dwarfs typically range from 10,000 to 12,000 K (18,000 to 21,000 °F), and they appear visually as white or blue-white.

The strength of spectral lines of the more abundant elements in a star’s atmosphere allows additional subdivisions within a class. Thus, the Sun, a main-sequence star, is classified as G2 V, in which the V denotes main sequence. Betelgeuse, a red giant with a surface temperature about half that of the Sun but with a luminosity of about 10,000 solar units, is classified as M2 Iab. In this classification, the spectral type is M2, and the Iab indicates a giant, well above the main sequence on the H-R diagram.



Betelgeuse

Betelgeuse imaged in ultraviolet light by the Hubble Space Telescope.

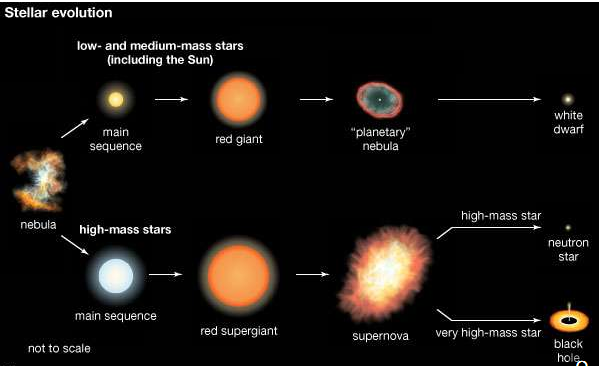
**Star formation and evolution**

The range of physically allowable masses for stars is very narrow. If the star’s mass is too small, the central temperature will be too low to sustain fusion reactions. The theoretical minimum stellar mass is about 0.08 solar mass. An upper theoretical bound called the Eddington limit, of several hundred solar masses, has been suggested, but this value is not firmly defined. Stars as massive as this will have luminosities about one million times greater than that of the Sun.



Horsehead Nebula

A general model of star formation and evolution has been developed, and the major features seem to be established. A large cloud of gas and dust can contract under its own gravitational attraction if its temperature is sufficiently low. As gravitational energy is released, the contracting central material heats up until a point is reached at which the outward radiation pressure balances the inward gravitational pressure, and contraction ceases. Fusion reactions take over as the star’s primary source of energy, and the star is then on the main sequence. The time to pass through these formative stages and onto the main sequence is less than 100 million years for a star with as much mass as the Sun. It takes longer for less massive stars and a much shorter time for those much more massive.



stellar evolution

Once a star has reached its main-sequence stage, it evolves relatively slowly, fusing hydrogen nuclei in its core to form helium nuclei. Continued fusion not only releases the energy that is radiated but also results in nucleosynthesis, the production of heavier nuclei.

Stellar evolution has of necessity been followed through computer modeling, because the timescales for most stages are generally too extended for measurable changes to be observed, even over a period of many years. One exception is the supernova, the violently explosive finale of certain stars. Different types of supernovas can be distinguished by their spectral lines and by changes in luminosity during and after the outburst. In Type Ia, a white dwarf star attracts matter from a nearby companion; when the white dwarf’s mass exceeds about 1.4 solar masses, the star implodes and is completely destroyed. Type II supernovas are not as luminous as Type Ia and are the final evolutionary stage of stars more massive than about eight solar masses. Type Ib and Ic supernovas are like Type II in that they are from the collapse of a massive star, but they do not retain their hydrogen envelope.

The nature of the final products of stellar evolution depends on stellar mass. Some stars pass through an unstable stage in which their dimensions, temperature, and luminosity change cyclically over periods of hours or days. These so-called Cepheid variables serve as standard candles for distance measurements. Some stars blow off their outer layers to produce planetary nebulas. The expanding material can be seen glowing in a thin shell as it disperses into the interstellar medium while the remnant core, initially with a surface temperature as high as 100,000 K (180,000 °F), cools to become a white dwarf. The maximum stellar mass that can exist as a white dwarf is about 1.4 solar masses and is known as the Chandrasekhar limit. More-massive stars may end up as either neutron stars or black holes.

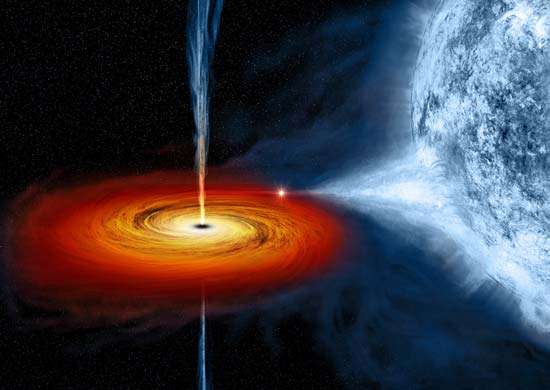


Nebula NGC 6751 formed several thousands of years ago as a dying star in the constellation Aquila threw off its outer layers of gas, which then fluoresced into a shell. The image, captured via NASA's Hubble Space Telescope, combines views taken through three different color filters in order to show the different temperatures of the nebular gases, with blue representing the hottest and orange and red the coolest. Our own sun will undergo a similar process in about 6 billion years, as it nears the end of its life.

The average density of a white dwarf is calculated to exceed one million grams per cubic cm. Further compression is limited by a quantum condition called degeneracy, in which only certain energies are allowed for the electrons in the star’s interior. Under sufficiently great pressure, the electrons are forced to combine with protons to form neutrons. The resulting neutron star will have a density in the range of 1014–1015 grams per cubic cm, comparable to the density within atomic nuclei. The behaviour of large masses having nuclear densities is not yet sufficiently understood to be able to set a limit on the maximum size of a neutron star, but it is thought to be less than three solar masses.

Still more-massive remnants of stellar evolution would have smaller dimensions and would be even denser that neutron stars. Such remnants are conceived to be black holes, objects so compact that no radiation can escape from within a characteristic distance called the Schwarzschild radius. This critical dimension is defined by Rs = 2GM/c2. (Rs is the Schwarzschild radius, G is the gravitational constant, M is the object’s mass, and c is the speed of light.) For an object of three solar masses, the Schwarzschild radius would be about three kilometres. Radiation emitted from beyond the Schwarzschild radius can still escape and be detected.

Although no light can be detected coming from within a black hole, the presence of a black hole may be manifested through the effects of its gravitational field, as, for example, in a binary star system. If a black hole is paired with a normal visible star, it may pull matter from its companion toward itself. This matter is accelerated as it approaches the black hole and becomes so intensely heated that it radiates large amounts of X-rays from the periphery of the black hole before reaching the Schwarzschild radius. Some candidates for stellar black holes have been found—e.g., the X-ray source Cygnus X-1. Each of them has an estimated mass clearly exceeding that allowable for a neutron star, a factor crucial in the identification of possible black holes. (Supermassive black holes that do not originate as individual stars are thought to exist at the centre of active galaxies; see below Study of other galaxies and related phenomena.)



Cygnus X-1

the X-ray source Cygnus X-1. A black hole pulls in matter from a stellar companion.

Whereas the existence of stellar black holes has been strongly indicated, the existence of neutron stars was confirmed in 1968 when they were identified with the then newly discovered pulsars, objects characterized by the emission of radiation at short and extremely regular intervals, generally between 1 and 1,000 pulses per second and stable to better than a part per billion. Pulsars are considered to be rotating neutron stars, remnants of some supernovas.

**Study of the Milky Way Galaxy**

Stars are not distributed randomly throughout space. Many stars are in systems consisting of two or three members separated by less than 1,000 AU. On a larger scale, star clusters may contain many thousands of stars. Galaxies are much larger systems of stars and usually include clouds of gas and dust.



Milky Way Galaxy as seen from Earth

The solar system is located within the Milky Way Galaxy, close to its equatorial plane and about 8 kiloparsecs from the galactic centre. The galactic diameter is about 30 kiloparsecs, as indicated by luminous matter. There is evidence, however, for nonluminous matter—so-called dark matter—extending out nearly twice this distance. The entire system is rotating such that, at the position of the Sun, the orbital speed is about 220 km per second (almost 500,000 miles per hour) and a complete circuit takes roughly 240 million years. Application of Kepler’s third law leads to an estimate for the galactic mass of about 100 billion solar masses. The rotational velocity can be measured from the Doppler shifts observed in the 21-cm emission line of neutral hydrogen and the lines of millimetre wavelengths from various molecules, especially carbon monoxide. At great distances from the galactic centre, the rotational velocity does not drop off as expected but rather increases slightly. This behaviour appears to require a much larger galactic mass than can be accounted for by the known (luminous) matter. Additional evidence for the presence of dark matter comes from a variety of other observations. The nature and extent of the dark matter (or missing mass) constitutes one of today’s major astronomical puzzles.

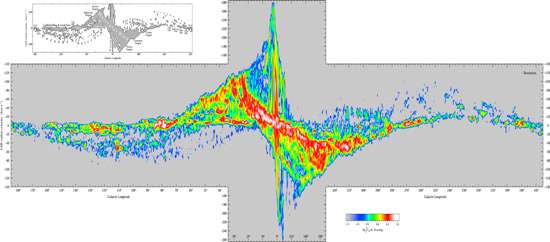
There are about 100 billion stars in the Milky Way Galaxy. Star concentrations within the galaxy fall into three types: open clusters, globular clusters, and associations. Open clusters lie primarily in the disk of the galaxy; most contain between 50 and 1,000 stars within a region no more than 10 parsecs in diameter. Stellar associations tend to have somewhat fewer stars; moreover, the constituent stars are not as closely grouped as those in the clusters and are for the most part hotter. Globular clusters, which are widely scattered around the galaxy, may extend up to about 100 parsecs in diameter and may have as many as a million stars. The importance to astronomers of globular clusters lies in their use as indicators of the age of the galaxy. Because massive stars evolve more rapidly than do smaller stars, the age of a cluster can be estimated from its H-R diagram. In a young cluster the main sequence will be well populated, but in an old cluster the heavier stars will have evolved away from the main sequence. The extent of the depopulation of the main sequence provides an index of age. In this way, the oldest globular clusters have been found to be about 12.5 billion years old, which should therefore be the minimum age for the galaxy.



Globular cluster M80 (also known as NGC 6093) in an optical image taken by the Hubble Space Telescope. M80 is located 28,000 light-years from Earth and contains hundreds of thousands of stars.

**Investigations of interstellar matter**

The interstellar medium, composed primarily of gas and dust, occupies the regions between the stars. On average, it contains less than one atom in each cubic centimetre, with about 1 percent of its mass in the form of minute dust grains. The gas, mostly hydrogen, has been mapped by means of its 21-cm emission line. The gas also contains numerous molecules. Some of these have been detected by the visible-wavelength absorption lines that they impose on the spectra of more-distant stars, while others have been identified by their own emission lines at millimetre wavelengths. Many of the interstellar molecules are found in giant molecular clouds, wherein complex organic molecules have been discovered.



Longitude-velocity map of the Milky Way Galaxy as shown by spectral line emission of carbon monoxide in molecular clouds. The vertical axis represents velocity and the horizontal axis longitude. The gentle curves in the left and right portions of the map trace the spiral arms of the Milky Way Galaxy. The vertical structure in the middle of the map is the centre of the Galaxy. The emission stretching from the upper left to the lower right in the middle portion of the map is the “molecular ring,” a ring of gas and dust in orbit between 4 and 8 kiloparsecs from the centre of the Galaxy.

In the vicinity of a very hot O- or B-type star, the intensity of ultraviolet radiation is sufficiently high to ionize the surrounding hydrogen out to a distance as great as 100 parsecs to produce an H II region, known as a Strömgren sphere. Such regions are strong and characteristic emitters of radiation at radio wavelengths, and their dimensions are well calibrated in terms of the luminosity of the central star. Using radio interferometers, astronomers are able to measure the angular diameters of H II regions even in some external galaxies and can thereby deduce the great distances to those remote systems. This method can be used for distances up to about 30 megaparsecs. (For additional information on H II regions, see nebula: Diffuse nebulae (H II regions).)



H II region

NGC 604, H II region inside the Triangulum Galaxy (M33), photographed by Hubble Space Telescope. NGC 604 is one of the largest H II regions; it contains more than 200 stars and is about 1,500 light-years in diameter.

Interstellar dust grains scatter and absorb starlight, the effect being roughly inversely proportional to wavelength from the infrared to the near ultraviolet. As a result, stellar spectra tend to be reddened. Absorption typically amounts to about one magnitude per kiloparsec but varies considerably in different directions. Some dusty regions contain silicate materials, identified by a broad absorption feature around a wavelength of 10 μm. Other prominent spectral features in the infrared range have been sometimes, but not conclusively, attributed to graphite grains and polycyclic aromatic hydrocarbons (PAHs).

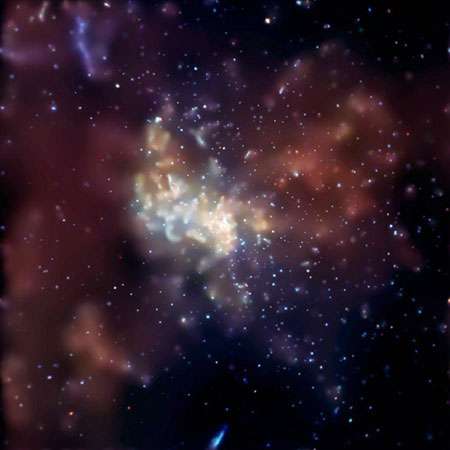
Starlight often shows a small degree of polarization (a few percent), with the effect increasing with stellar distance. This is attributed to the scattering of the starlight from dust grains that have been partially aligned in a weak interstellar magnetic field. The strength of this field is estimated to be a few microgauss, very close to the strength inferred from observations of nonthermal cosmic radio noise. This radio background has been identified as synchrotron radiation, emitted by cosmic-ray electrons traveling at nearly the speed of light and moving along curved paths in the interstellar magnetic field. The spectrum of the cosmic radio noise is close to what is calculated on the basis of measurements of the cosmic rays near Earth.

Cosmic rays constitute another component of the interstellar medium. Cosmic rays that are detected in the vicinity of Earth comprise high-speed nuclei and electrons. Individual particle energies, expressed in electron volts (eV; 1 eV = 1.6 × 10−12 erg), range with decreasing numbers from about 106 eV to more than 1020 eV. Among the nuclei, hydrogen nuclei are the most plentiful at 86 percent, helium nuclei next at 13 percent, and all other nuclei together at about 1 percent. Electrons are about 2 percent as abundant as the nuclear component. (The relative numbers of different nuclei vary somewhat with kinetic energy, while the electron proportion is strongly energy-dependent.)

A minority of cosmic rays detected in Earth’s vicinity are produced in the Sun, especially at times of increased solar activity (as indicated by sunspots and solar flares). The origin of galactic cosmic rays has not yet been conclusively identified, but they are thought to be produced in stellar processes such as supernova explosions, perhaps with additional acceleration occurring in the interstellar regions. (For additional information on interstellar matter, see Milky Way Galaxy: The general interstellar medium.)

**Observations of the galactic centre**

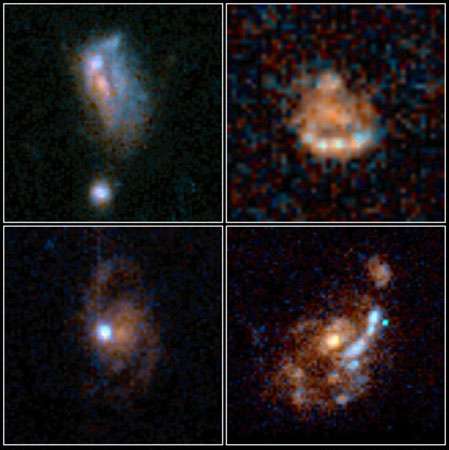
The central region of the Milky Way Galaxy is so heavily obscured by dust that direct observation has become possible only with the development of astronomy at nonvisual wavelengths—namely, radio, infrared, and, more recently, X-ray and gamma-ray wavelengths. Together, these observations have revealed a nuclear region of intense activity, with a large number of separate sources of emission and a great deal of dust. Detection of gamma-ray emission at a line energy of 511,000 eV, which corresponds to the annihilation of electrons and positrons (the antimatter counterpart of electrons), along with radio mapping of a region no more than 20 AU across, points to a very compact and energetic source, designated Sagittarius A\*, at the centre of the galaxy. Sagittarius A\* is a supermassive black hole with a mass equivalent to 4,310,000 Suns.



Cosmic radio-wave source Sagittarius A\* in an image from the Chandra X-ray Observatory. Sagittarius A\* is an extremely bright source within the larger Sagittarius A complex and contains the black hole at the Milky Way Galaxy's centre.

**Study of other galaxies and related phenomena**

Galaxies are normally classified into three principal types according to their appearance: spiral, elliptical, and irregular. Galactic diameters are typically in the tens of kiloparsecs and the distances between galaxies typically in megaparsecs.



Four irregular galaxies, as observed by the Hubble Space Telescope.

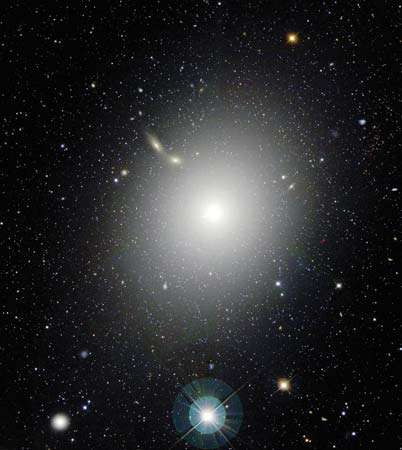
Spiral galaxies—of which the Milky Way system is a characteristic example—tend to be flattened, roughly circular systems with their constituent stars strongly concentrated along spiral arms. These arms are thought to be produced by traveling density waves, which compress and expand the galactic material. Between the spiral arms exists a diffuse interstellar medium of gas and dust, mostly at very low temperatures (below 100 K [−280 °F, −170 °C]). Spiral galaxies are typically a few kiloparsecs in thickness; they have a central bulge and taper gradually toward the outer edges.



Whirlpool Galaxy (M51); NGC 5195

The Whirlpool Galaxy (left), also known as M51, an Sc galaxy accompanied by a small irregular companion galaxy, NGC 5195 (right).

Ellipticals show none of the spiral features but are more densely packed stellar systems. They range in shape from nearly spherical to very flattened and contain little interstellar matter. Irregular galaxies number only a few percent of all stellar systems and exhibit none of the regular features associated with spirals or ellipticals.



The giant elliptical galaxy M87, also known as Virgo A, in an optical image taken by the Canada-France-Hawaii Telescope on Mauna Kea, Hawaii. M87 appears near the centre of the Virgo Cluster of galaxies.

Properties vary considerably among the different types of galaxies. Spirals typically have masses in the range of a billion to a trillion solar masses, with ellipticals having values from 10 times smaller to 10 times larger and the irregulars generally 10–100 times smaller. Visual galactic luminosities show similar spreads among the three types, but the irregulars tend to be less luminous. In contrast, at radio wavelengths the maximum luminosity for spirals is usually 100,000 times less than for ellipticals or irregulars.

Quasars are objects whose spectra display very large redshifts, thus implying (in accordance with the Hubble law) that they lie at the greatest distances. They were discovered in 1963 but remained enigmatic for many years. They appear as starlike (i.e., very compact) sources of radio waves—hence their initial designation as quasi-stellar radio sources, a term later shortened to quasars. They are now considered to be the exceedingly luminous cores of distant galaxies. These energetic cores, which emit copious quantities of X-rays and gamma rays, are termed active galactic nuclei (AGN) and include the object Cygnus A and the nuclei of a class of galaxies called Seyfert galaxies. They may be powered by the infall of matter into supermassive black holes.



Quasar

3C 273, the brightest quasar, photographed by the Hubble Space Telescope's Advanced Camera for Surveys. The black region at the centre of the image is blocking light from the central object, revealing the host galaxy of 3C 273.

The Milky Way Galaxy is one of the Local Group of galaxies, which contains about four dozen members and extends over a volume about two megaparsecs in diameter. Two of the closest members are the Magellanic Clouds, irregular galaxies about 50 kiloparsecs away. At about 740 kiloparsecs, the Andromeda Galaxy is one of the most distant in the Local Group. Some members of the group are moving toward the Milky Way system while others are traveling away from it. At greater distances, all galaxies are moving away from the Milky Way Galaxy. Their speeds (as determined from the redshifted wavelengths in their spectra) are generally proportional to their distances. The Hubble law relates these two quantities. In the absence of any other method, the Hubble law continues to be used for distance determinations to the farthest objects—that is, galaxies and quasars for which redshifts can be measured.



The Andromeda Galaxy, also known as the Andromeda Nebula or M31. It is the closest spiral galaxy to Earth, at a distance of 2.48 million light-years.

**Cosmology**

Cosmology is the scientific study of the universe as a unified whole, from its earliest moments through its evolution to its ultimate fate. The currently accepted cosmological model is the big bang. In this picture, the expansion of the universe started in an intense explosion 13.8 billion years ago. In this primordial fireball, the temperature exceeded one trillion K, and most of the energy was in the form of radiation. As the expansion proceeded (accompanied by cooling), the role of the radiation diminished, and other physical processes dominated in turn. Thus, after about three minutes, the temperature had dropped to the one-billion-K range, making it possible for nuclear reactions of protons to take place and produce nuclei of deuterium and helium. (At the higher temperatures that prevailed earlier, these nuclei would have been promptly disrupted by high-energy photons.) With further expansion, the time between nuclear collisions had increased and the proportion of deuterium and helium nuclei had stabilized. After a few hundred thousand years, the temperature must have dropped sufficiently for electrons to remain attached to nuclei to constitute atoms. Galaxies are thought to have begun forming after a few million years, but this stage is very poorly understood. Star formation probably started much later, after at least a billion years, and the process continues today.

Observational support for this general model comes from several independent directions. The expansion has been documented by the redshifts observed in the spectra of galaxies. Furthermore, the radiation left over from the original fireball would have cooled with the expansion. Confirmation of this relic energy came in 1965 with one of the most striking cosmic discoveries of the 20th century—the observation, at short radio wavelengths, of a widespread cosmic radiation corresponding to a temperature of almost 3 K (about −270 °C [−454 °F]). The shape of the observed spectrum is an excellent fit with the theoretical Planck blackbody spectrum. (The present best value for this temperature is 2.735 K, but it is still called three-degree radiation or the cosmic microwave background.) The spectrum of this cosmic radio noise peaks at approximately a one-millimetre wavelength, which is in the far infrared, a difficult region to observe from Earth; however, the spectrum has been well mapped by the Cosmic Background Explorer (COBE), Wilkinson Microwave Anisotropy Probe, and Planck satellites. Additional support for the big bang theory comes from the observed cosmic abundances of deuterium and helium. Normal stellar nucleosynthesis cannot produce their measured quantities, which fit well with calculations of production during the early stages of the big bang.

Early surveys of the cosmic background radiation indicated that it is extremely uniform in all directions (isotropic). Calculations have shown that it is difficult to achieve this degree of isotropy unless there was a very early and rapid inflationary period before the expansion settled into its present mode. Nevertheless, the isotropy posed problems for models of galaxy formation. Galaxies originate from turbulent conditions that produce local fluctuations of density, toward which more matter would then be gravitationally attracted. Such density variations were difficult to reconcile with the isotropy required by observations of the 3 K radiation. This problem was solved when the COBE satellite was able to detect the minute fluctuations in the cosmic background from which the galaxies formed.

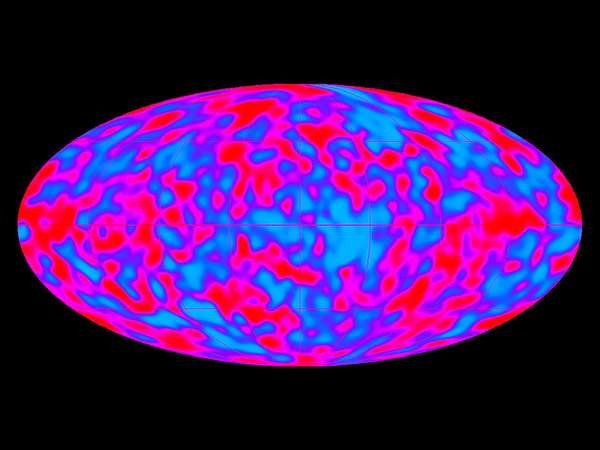


Image of the cosmic microwave background, taken by the Differential Microwave Radiometer on board the U.S. satellite Cosmic Background Explorer. The red features in the image show places where the universe was slightly denser, thus stimulating gravitational separation and, ultimately, the formation of galaxies.

The very earliest stages of the big bang are less well understood. The conditions of temperature and pressure that prevailed prior to the first microsecond require the introduction of theoretical ideas of subatomic particle physics. Subatomic particles are usually studied in laboratories with giant accelerators, but the region of particle energies of potential significance to the question at hand lies beyond the range of accelerators currently available. Fortunately, some important conclusions can be drawn from the observed cosmic helium abundance, which is dependent on conditions in the early big bang. The observed helium abundance sets a limit on the number of families of certain types of subatomic particles that can exist.

The age of the universe can be calculated in several ways. Assuming the validity of the big bang model, one attempts to answer the question: How long has the universe been expanding in order to have reached its present size? The numbers relevant to calculating an answer are Hubble’s constant (i.e., the current expansion rate), the density of matter in the universe, and the cosmological constant, which allows for change in the expansion rate. In 2003 a calculation based on a fresh determination of Hubble’s constant yielded an age of 13.7 billion ± 200 million years, although the precise value depends on certain assumed details of the model used. Independent estimates of stellar ages have yielded values less than this, as would be expected, but other estimates, based on supernova distance measurements, have arrived at values of about 15 billion years, still consistent, within the errors. In the big bang model the age is proportional to the reciprocal of Hubble’s constant, hence the importance of determining H as reliably as possible. For example, a value for H of 100 km/sec/Mpc would lead to an age less than that of many stars, a physically unacceptable result.

A small minority of astronomers have developed alternative cosmological theories that are seriously pursued. The overwhelming professional opinion, however, continues to support the big bang model.

Finally, there is the question of the future behaviour of the universe: Is it open? That is to say, will the expansion continue indefinitely? Or is it closed, such that the expansion will slow down and eventually reverse, resulting in contraction? (The final collapse of such a contracting universe is sometimes termed the “big crunch.”) The density of the universe seems to be at the critical density; that is, the universe is neither open nor closed but “flat.” So-called dark energy, a kind of repulsive force that is now believed to be a major component of the universe, appears to be the decisive factor in predictions of the long-term fate of the cosmos. If this energy is a cosmological constant (as proposed in 1917 by Albert Einstein to correct certain problems in his model of the universe), then the result would be a “big chill.” In this scenario, the universe would continue to expand, but its density would decrease. While old stars would burn out, new stars would no longer form. The universe would become cold and dark. The dark (nonluminous) matter component of the universe, whose composition remains unknown, is not considered sufficient to close the universe and cause it to collapse; it now appears to contribute only a fourth of the density needed for closure.

An additional factor in deciding the fate of the universe might be the mass of neutrinos. For decades the neutrino had been postulated to have zero mass, although there was no compelling theoretical reason for this to be so. From the observation of neutrinos generated in the Sun and other celestial sources such as supernovas, in cosmic-ray interactions with Earth’s atmosphere, and in particle accelerators, investigators have concluded that neutrinos have some mass, though only an extremely small fraction of the mass of an electron. Although there are vast numbers of neutrinos in the universe, the sum of such small neutrino masses appears insufficient to close the universe.