Sun's gravitational sphere of influence extends out much farther, to approximately $2 \times 10^7$ AU, and there are bodies in orbit around the Sun at those distances. These include the Kuiper belt, which may extend out to $\sim 10^8$ AU, and the Oort cloud, which is populated to the limits of the Sun's gravitational field.

### III. The Origin of the Solar System

Our knowledge of the origin of the Sun and the planetary system comes from two sources: study of the solar system itself, and study of star formation in nearby giant molecular clouds. The two sources are radically different. In the case of the solar system, we have an abundance of detailed information on the planets, their satellites, and numerous small bodies. But the solar system we see today is a highly evolved system that has undergone massive changes since it first condensed from the natal cloud, and we must learn to recognize which qualities reflect that often violent evolution and which truly record conditions at the time of solar system formation.

In contrast, when studying even the closest star-forming regions (which are about 140 pc from the Sun), we are handicapped by a lack of adequate resolution and detail. In addition, we are forced to take a “snapshot” view of many young stars at different stages in their formation, and from that attempt to generate a time-ordered sequence of the many different stages and processes involved. When we observe the formation of other stars we need also to recognize that some of the observed processes or events may not be applicable to the formation of our own Sun and planetary system.

Still, a coherent picture has emerged of the major events and processes in the formation of the solar system. That picture assumes that the Sun is a typical star and that it formed in a similar way to many of the low-mass protostars we see today.

The birthplace of stars is giant molecular clouds in the galaxy. These huge clouds of molecular hydrogen have masses of $10^4$ to $10^6$ solar masses, $M_\odot$. Within these clouds are denser regions or “cores” where star formation actually takes place. Some process, perhaps the shock wave from a nearby supernova, triggers the gravitational collapse of a cloud core. Material falls toward the center of the core under its own self-gravity and a massive object begins to grow at the center of the cloud. Heated by the gravitational potential energy of the infalling matter, the object becomes self-luminous, and is then described as a “protostar.” Although central pressures and temperatures are not yet high enough to ignite nuclear fusion, the protostar begins to heat the growing nebula around it. The timescale of the infall of the cloud material for a solar-mass cloud is about $10^5$ years.

The infalling cloud material consists of both gas and dust. The gas is mostly hydrogen (77%) with helium (21%) and other gases. The dust is a mix of interstellar grains, including silicates, organics, and condensed ices. A popular model suggests that the silicate grains are coated with icy-organic mantles. As the dust grains fall inward, they experience a pressure from the increasing density of gas toward the center of the nebula. This slows and even halts the inward radial component of their motion. However, the dust grains can still move vertically with respect to the central plane of the nebula, as defined by the rotational angular momentum vector of the original cloud core. As a result, the grains settle toward the central plane.

As the grains settle, they begin to collide with one another. The grains stick and quickly grow from microscopic to macroscopic objects, perhaps meters in size (initial agglomerations of grains may look very much like the suspected cometary IDP in Fig. 9). This process continues and even increases as the grains reach the denser environment at the central plane of the nebula. The meter-sized bodies grow to kilometer-sized bodies, and these bodies grow to 100 km-sized bodies. These bodies are known as planetesimals. As a planetesimal begins to acquire significant mass, its cross section for accretion grows beyond its physical cross section because it is now capable of gravitationally deflecting smaller planetesimals toward it. These larger planetesimals then “run away” from the others, growing at an ever-increasing rate.

The actual process is far more complex than described here, and many details of this scenario still need to be worked out. For example, the role of turbulence in the nebula is not well quantified. Turbulence would tend to slow or even prevent the accretion of grains into larger objects. Also, the role of electrostatic and magnetic effects in the nebula are not understood.

Nevertheless, it appears that accretion in the central plane of the solar nebula can account for the growth of planets from interstellar grains. An artist’s concept of the accretion disk in the solar nebula is shown in Fig. 12. In the inner region of the solar nebula, close to the forming Sun, the higher temperatures would vaporize icy and organic grains, leaving only silicate grains to form the planetesimals, which eventually merged to form the terrestrial planets. At larger distances where the nebula was cooler, organic and icy
grains would condense and these would combine with
the silicates to form the cores of the giant planets.
Because the total mass of ice and organics may have
been several times the mass of silicates, the cores of
the giant planets may actually have grown faster than
the terrestrial planets interior to them.

At some point, the growing cores of the giant
planets became sufficiently massive to begin capturing
hydrogen and helium directly from the nebula gas.
Because of the lower temperatures in the outer
planets zone, the giant planets were able to retain
the gas and continue to grow even larger. The
terrestrial planets close to the Sun may have acquired
some nebula gas, but likely could not hold on to it
at their higher temperatures.

Observations of protostars in nearby molecular
clouds have found substantial evidence for accretionary
disks and gas nebulae surrounding these stars. The
relative ages of these protostars can be estimated by
comparing their luminosity and color with theoretical
predictions of their location in the Hertzsprung–
Russell diagram. One of the more interesting observa-
tions is that the nebula dust and gas around solar-mass
protostars seem to dissipate after about 10^7 years. It
appears that the nebula and dust may be swept away by
mass outflows, essentially superpowerful solar winds,
from the protostars. If the Sun formed similarly to the
protostars we see today, then these observations set
strong limits on the likely formation times of Jupiter
and Saturn.

An interesting process that must have occurred
during the late stages of planetary accretion is “giant
impacts,” that is, collisions between very large pro-
toplanetary objects. As noted in Section II. C, a
giant impact between a Mars-sized protoplanet and
the proto-Earth is now the accepted explanation for
the origin of the Earth’s Moon. Giant impacts have
similarly been invoked to explain the high mean
density of Mercury, the retrograde rotation of Venus,
the high obliquity of Uranus, and possibly even the
formation of the Pluto–Charon binary. Although it
was previously thought that such giant impacts were
low-probability events, they are now recognized to
be a natural consequence of the final stages of
planetary accretion.

Another interesting process late in the accretion of
the planets is the clearing of debris from the planetary
zones. At some point in the growth of the planets,
their gravitational spheres of influence grew suffi-
ciently large that an encounter with a planetesimal
would more likely lead to the planetesimal being scat-
tered into a different orbit, rather than an actual colli-
sion. This would be particularly true for the massive
Jovian planets, both because of their stronger gravita-
tional fields and because of their larger distances from the Sun.

Since it is just as likely that a planet will scatter objects inward as outward, the clearing of the planetary zones resulted in planetesimals being flung throughout the solar system, and in a massive bombardment of all planets and satellites. Many planetesimals were also flung out of the planetary system to interstellar space, or to distant orbits in the Oort cloud. Although the terrestrial planets are generally too small to eject objects out of the solar system, they can scatter objects to Jupiter-encountering orbits where Jupiter will quickly dispose of them.

The clearing of the planetary zones has several interesting consequences. The dynamical interaction between the planets and the remaining planetesimals results in an exchange of angular momentum. Computer-based dynamical simulations have shown that this causes the semimajor axes of the planets to migrate radially. In general, Saturn, Uranus, and Neptune are expected to first move inward and then later outward as the ejection of material progresses. Jupiter, which ejects the most material because of its huge mass, migrates inward, but only by a few tenths of an astronomical unit.

This migration of the giant planets has significant consequences for the populations of small bodies in the planetary region. As the planets move, the locations of their mean motion and secular resonances will move with them. This will result in some small bodies being captured into resonances while others will be thrown into chaotic orbits, leading to their eventual ejection from the system or possibly to impacts on the planets and the Sun. The radial migration of the giant planets has been invoked in the clearing of both the outer regions of the main asteroid belt and the inner regions of the Kuiper belt.

Another consequence of the clearing of the planetary zones is that rocky planetesimals formed in the terrestrial planets zone will be scattered throughout the Jovian planets region, and vice versa for icy planetesimals formed in the outer planets zone. The bombardment of the terrestrial planets by icy planetesimals is of particular interest, both in explaining the Late Heavy Bombardment and as a means of delivering the volatile reservoirs of the terrestrial planets. Isotopic studies suggest that some fraction of the water in the Earth’s oceans may have come from comets, though not all of it. Also, the recent discovery of an asteroidal-appearing object, 1996 PW, on a long-period comet orbit has provided evidence that asteroids may indeed have been ejected to the Oort cloud, where they may make up 1–3% of the population there.

The Milky Way is a large, spiral galaxy, about 30 kpc in diameter. Some parts of the galactic disk can be traced out to 25 kpc from the galactic center, and the halo can be traced to 50 kpc. The galaxy contains approximately 100 billion stars and the total mass of the galaxy is estimated to be about $4 \times 10^{11}$ solar masses ($M_\odot$). Approximately 25% of the mass of the galaxy is estimated to be in visible stars, about 15% in stellar remnants (white dwarfs, neutron stars, and black holes), 25% in interstellar clouds and interstellar material, and 35% in “dark matter.” Dark matter is a general term used to describe unseen mass in the galaxy, which is needed to explain the observed dynamics of the galaxy (i.e., stellar motions, galactic rotation) but which has not been detected through any available means. There is considerable speculation about the nature of the dark matter, which includes everything from exotic nuclear particles to brown dwarfs (substellar objects, not capable of nuclear burning) and dark stars (the burned-out remnants of old stars) to massive black holes. The galaxy is estimated to have an age of 10 to 15 billion years, equal to the age of the universe.

The Milky Way galaxy consists of four major structures: the galactic disk, the central bulge, the halo, and the corona (Fig. 13). As the name implies, the disk is a highly flattened, rotating structure about 15 kpc in radius and about 0.5–0.8 kpc thick, depending on which population of stars is used to trace the disk. The disk contains relatively young stars and interstellar clouds, arranged in a multi-arm spiral structure (Fig. 14). At the center of the disk is the bulge, an oblate spheroid about 3 kpc in radius in the plane of the disk, and with a radius of about 1.5 kpc perpendicular to the disk. The bulge rotates more slowly than the disk, and consists largely of densely packed older stars and interstellar clouds. It does not display spiral structure. At the center of the bulge is the nucleus, a complex region only 4–5 pc across, which appears to have a massive black hole at its center. The mass of the central black hole has been estimated at 2.6 million $M_\odot$.

The halo surrounds both of these structures and extends ~20 kpc from the galactic center. The halo has an oblate spheroid shape and contains older stars and globular clusters of stars. The corona appears to be a yet more distant halo at 60–100 kpc and consists of dark matter, unobservable except for the effect it has on the dynamics of observable bodies in the galaxy. The corona may be several times more massive than
the other three galactic components combined. Many descriptions of the galaxy include the halo and the corona as a single component.

The galactic disk is visible in the night sky as the Milky Way, a bright band of light extending around the celestial sphere. When examined with a small telescope, the Milky Way is resolved into thousands or even millions of individual stars, and numerous nebulae and star clusters. The direction to the center of the galaxy is in the constellation Sagittarius (best seen from the Southern Hemisphere in June) and the disk appears visibly wider in that direction, which is the view of the central bulge.

The disk is not perfectly flat; there is evidence for warping in the outer reaches of the disk, between 15 and 25 kpc. The warp may be the result of gravitational

FIGURE 13 An image of the sky at infrared wavelengths as constructed from IRAS satellite data. The Milky Way galaxy is visible as the bright horizontal band through the image, with the galactic bulge at the center of the image. The fainter, S-shaped structure extending from lower left to upper right is the zodiacal dust cloud in the ecliptic plane. The plane of the ecliptic is tilted 63° to the plane of the galaxy. Dark gorges are gaps in the data caused by incomplete scans by IRAS.

FIGURE 14 Messier 100, a large spiral galaxy in the constellation Coma Berenices, as photographed by the Hubble Space Telescope. The Milky Way galaxy may appear similar to this.
perturbations due to encounters with other galaxies, and/or with the Magellanic clouds, two nearby, irregular dwarf galaxies that appear to be in orbit around the Milky Way. Similarly, evidence has been building in recent years that the bulge is not an oblate spheroid, but rather appears to have a triaxial shape. This type of structure is observed in external galaxies and is referred to as a “bar”; such galaxies are known as barred spirals. In addition, the Milky Way’s central bar appears to be tilted relative to the plane of the galactic disk. The nonspherical shape of the bulge and the tilt have important implications for understanding stellar dynamics and the long-term evolution of the galaxy.

Stars in the galactic disk have different characteristic velocities as a function of their stellar classification, and hence age. Low-mass, older stars, like the Sun, have relatively high random velocities and as a result can move farther out of the galactic plane. Younger, more massive stars have lower mean velocities and thus smaller scale heights above and below the plane. Giant molecular clouds, the birthplace of stars, also have low mean velocities and thus are confined to regions relatively close to the galactic plane. The disk rotates clockwise as viewed from “galactic north,” at a relatively constant velocity of 160–220 km sec⁻¹. This motion is distinctly non-Keplerian, the result of the very nonspherical mass distribution. The rotation velocity for a circular galactic orbit in the galactic plane defines the Local Standard of Rest (LSR). The LSR is then used as the reference frame for describing local stellar dynamics.

The Sun and the solar system are located approximately 8.5 kpc from the galactic center, and 10–20 pc above the central plane of the galactic disk. The circular orbit velocity at the Sun’s distance from the galactic center is 220 km sec⁻¹, and the Sun and the solar system are moving at approximately 17 to 22 km sec⁻¹ relative to the LSR. The Sun’s velocity vector is currently directed toward a point in the constellation of Hercules, approximately at right ascension 18h 0° and declination +30°, known as the solar apex. Because of this motion relative to the LSR, the solar system’s galactic orbit is not circular. The Sun and planets move in a quasi-elliptical orbit between about 8.4 and 9.7 kpc from the galactic center, with a period of revolution of about 240 million years. The solar system is currently close to and moving inward toward “perigalacticon,” the point in the orbit closest to the galactic center. In addition, the solar system moves perpendicular to the galactic plane in a harmonic fashion, with a period of 52 to 74 million years and an amplitude of ±49 to 93 pc out of the galactic plane. (The uncertainties in the estimates of the period and amplitude of the motion are caused by the uncertainty in the amount of dark matter in the galactic disk.) The Sun and planets passed through the galactic plane about 2–3 million years ago, moving “northward.”

The Sun and solar system are located at the inner edge of one of the spiral arms of the galaxy, known as the Orion or local arm. Nearby spiral structures can be traced by constructing a three-dimensional map of stars, star clusters, and interstellar clouds in the solar neighborhood. Two well-defined neighboring structures are the Perseus arm, farther from the galactic center than the local arm, and the Sagittarius arm, toward the galactic center. The arms are about 0.5 kpc wide and the spacing between the spiral arms is about 1.2–1.6 kpc. The local galactic spiral arm structure is illustrated in Fig. 15.

The Sun’s velocity relative to the LSR is low as compared with other G-type stars, which have typical velocities of 40–45 km sec⁻¹ relative to the LSR. Stars are accelerated by encounters with giant molecular clouds in the galactic disk. Thus, older stars can be accelerated to higher mean velocities, as noted earlier. The reason(s) for the Sun’s low velocity are not known. Velocity-altering encounters with giant molecular clouds occur with a typical frequency of once every 300–500 million years.

The local density of stars in the solar neighborhood is about 0.11 pc⁻³, though many of the stars are in binary or multiple star systems. The local density of binary and multiple star systems is 0.086 pc⁻¹. Most of these are low-mass stars, less massive and less luminous than the Sun. The nearest star to the solar system is Proxima Centauri, which is a low-mass (M ≈ 0.1M☉), distant companion to Alpha Centauri, which itself is a double-star system of two close-orbiting solar-type stars. Proxima Centauri is currently about 1.3 pc from the Sun and about 0.06 pc (1.3 × 10⁶ AU) from the Alpha Centauri pair it is orbiting. The second nearest star is Barnard’s star, a fast-moving red dwarf at a distance of 1.83 pc. The brightest star within 5 pc of the Sun is Sirius, an A1 star (M ≈ 2.0M☉) about 2.6 pc away. Sirius also is a double star, with a faint, white dwarf companion. The stars in the solar neighborhood are shown in Fig. 16.

The Sun’s motion relative to the LSR, as well as the random velocities of the stars in the solar neighborhood, will occasionally result in close encounters between the Sun and other stars. Using the foregoing value for the density of stars in the solar neighborhood, one can predict that about 12 star systems (single or multiple stars) will pass within 1 pc of the Sun per million years. The total number of stellar encounters scales as the square of the encounter distance. This rate
has been confirmed in part by data from the *Hipparcos* astrometry satellite, which measured the distances and proper motions of \( \sim 118,000 \) stars and which was used to reconstruct the trajectories of stars in the solar neighborhood.

Based on this rate, the closest stellar approach over the lifetime of the solar system would be expected to be at \( \sim 900 \) AU. Such an encounter would result in a major perturbation of the Oort cloud and would eject many comets to interstellar space. It would also send a shower of comets into the planetary region, raising the impact rate on the planets for a period of about 2–3 million years, and having other effects that may be detectable in the stratigraphic record on the Earth or on other planets. A stellar encounter at 900 AU could also have a substantial perturbative effect on the orbits of comets in the Kuiper belt and would likely disrupt the outer regions of that ecliptic comet disk.

Obviously, the effect that any such stellar passage will have is a strong function of the mass and velocity of the passing star.

The advent of space-based astronomy, primarily through Earth-orbiting ultraviolet and X-ray telescopes, has made it possible to study the local interstellar medium surrounding the solar system. The structure of the local interstellar medium has turned out to be quite complex. The solar system appears to be on the edge of an expanding bubble of hot plasma about 120 pc in radius, which appears to have originated from multiple supernovae explosions in the Scorpius–Centaurus OB association. The Sco–Cen association is a nearby star-forming region that contains many young, high-mass O- and B-type stars. Such stars have relatively short lifetimes and end their lives in massive supernova explosions, before collapsing into black holes. The expanding shells of hot gas blown off the
stars in the supernova explosions are able to “sweep” material before them, leaving a low-density “bubble” of hot plasma.

Within this bubble, known as the Local Bubble, the solar system is at this time within a small interstellar cloud, perhaps 2–5 pc across, known as the Local Interstellar Cloud. That cloud is apparently a fragment of the expanding shells of gas from the supernova explosions, and there appear to be a number of such clouds within the local solar neighborhood.

V. THE FATE OF THE SOLAR SYSTEM

Stars like the Sun are expected to have lifetimes on the main sequence of about 10 billion years. The main sequence lifetime refers to the time period during which the star produces energy through hydrogen fusion in its core. As the hydrogen fuel in the core is slowly depleted over time, the core contracts to maintain the internal pressure. This raises the central temperature and, as a result, the rate of nuclear fusion also increases and the star slowly brightens. Thus, temperatures throughout the solar system will slowly increase over time. Presumably, this slow brightening has already been going on since the formation of the Sun and solar system.

A $1M_\odot$ star like the Sun is expected to run out of hydrogen at its core in about 10 billion years. As the production of energy declines, the core again contracts. The rising internal temperature and pressure are then able to ignite hydrogen burning in a shell surrounding the depleted core. The hydrogen burning in the shell heats the surrounding mass of the star and causes it to expand. The radius of the star increases and the surface temperature drops. The luminosity of the star increases dramatically and it becomes a red giant. Eventually the star reaches a brightness about $10^4$ times more luminous than the present-day Sun, a surface temperature of 3000 K, and a radius of 100–200 solar radii. A distance of one hundred solar radii is equal to 0.46 AU, larger than the orbit of Mercury. Two hundred radii is just within the orbit of the Earth. Thus, Mercury and likely Venus will be incorporated into the outer shell of the red giant Sun and will be vaporized.

The increased solar luminosity during the red giant phase will result in a fivefold rise in temperatures throughout the solar system. At the Earth’s orbit, this temperature increase will vaporize the oceans and roast the planet at a temperature on the order of ~1400 K or more. At Jupiter’s orbit it will melt the icy Galilean satellites and cook them at a more modest temperature of about 600 K, about the same as current noontime temperatures on the surface of Mercury. Typical temperatures at the orbit of Neptune will be about the same as they are today at the orbit of the Earth. Comets in the inner portion of the Kuiper belt will be warmed sufficiently to produce visible comae.

The lowered gravity at the surface of the greatly expanded Sun will result in a substantially increased
solar wind, and the Sun will slowly lose mass from its outer envelope. Meanwhile, the core of the Sun will continue to contract until the central temperature and pressure are great enough to ignite helium burning in the core. During this time, hydrogen burning continues in a shell around the core. Helium burning continues during the red giant phase until the helium in the core is also exhausted. The core again contracts and this permits helium burning to ignite in a shell around the core. This is an unstable situation and the star can undergo successive contractions and reignition pulses, during which it will blow off part or all of its outer envelope into space. These huge mass ejections produce an expanding nebula around the star, known as a planetary nebula (because it looks somewhat like the disk of a Jovian planet through a telescope). For a star with the mass of the Sun, the entire red giant phase lasts about 700 million years.

As the Sun loses mass in this fashion, the orbits of the surviving planets will slowly spiral outward. This will also be true for comets in the Kuiper belt and Oort cloud. Since the gravitational sphere of influence of the Sun will shrink as a result of the Sun’s decreasing mass, comets will be lost to interstellar space at a greater rate from the outer edges of the Oort cloud.

As a red-giant star loses mass, its core continues to contract. However, for an initially $1M_\odot$ star like the Sun, the central pressure and temperature cannot rise sufficiently to ignite carbon burning in the core, the next phase in nuclear fusion. With no way of producing additional energy other than gravitational contraction, the luminosity of the star plunges. The star continues to contract and cool, until the contraction is halted by degenerate electron pressure in the superdense core. At this point, the mass of the star has been reduced to about 70% of its original mass and the diameter is about the same as that of the present-day Earth. Such a star is known as a white dwarf. The remnants of the previously roasted planets will be plunged into a deep freeze as the luminosity of the white dwarf slowly declines.

The white dwarf star will continue to cool over a period of about 1 billion years, to the point where its luminosity drops below detectable levels. Such a star is referred to as a black dwarf. A nonluminous star is obviously very difficult to detect. There is some suggestion that they may have been found through an observing technique known as micro-lensing events. Dark stars provide one of the possible explanations of the dark matter in the galaxy.

VI. CONCLUDING REMARKS

This chapter has introduced the solar system and its varied members, viewing them as components of a large and complex system. Each of them—the Sun, the planets, their satellites, the comets and asteroids—is also a fascinating world in its own right. The ensuing chapters provide more detailed descriptions of each of these members of the solar system.

BIBLIOGRAPHY
