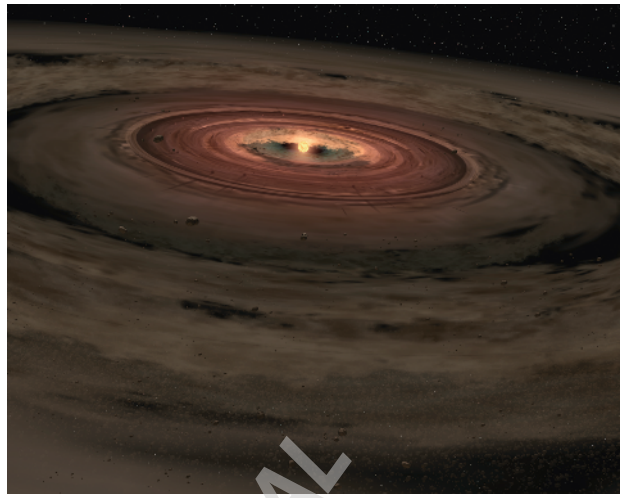
ONE
Beginnings



For millennia, people have studied the heavens and wondered about the nature and origins of the Sun, Moon and planets. Indeed, Solar System studies dominated the field of astronomy until the introduction of powerful telescopes and advanced instruments in the 19th century. In the last 50 years, spacecraft have flown past or orbited all of the major planets, landed on the Moon, Mars, Titan and an asteroid, and brought back samples of Moon rock, the solar wind, asteroid and comet dust. This era of robotic and human exploration has revolutionized scientists' knowledge of our corner of the Galaxy, and further astounding revelations are expected in the decades to come.

Wandering Stars

Since time immemorial, people have stared in wonderment at the night sky. In previous millennia, when the darkness of the sky was not degraded by artificial lighting, it was easy to recognize how the stellar patterns drifted from horizon to horizon as the night progressed, and how they changed as the seasons passed.

However, in addition to the familiar, twinkling stars, observers noted seven objects that moved with varying speeds against the background of "fixed" stars.[1] In order of greatest apparent brightness, they were the Sun, Moon, Venus, Jupiter, Mars, Mercury and Saturn. The ancient Greeks called them "planetes" ("wandering stars"), a designation we still use for all but the Sun and Moon.

For ancient astrologers and astronomers (the two disciplines were inextricably intertwined for many centuries) the most important of the wanderers were the Sun, which was responsible for daylight, and the Moon, which dominated the night. Both of these objects displayed visible disks and moved quite rapidly across the sky. Careful study of their regular motions and apparitions enabled people to devise calendars and introduce convenient ways of measuring time. Thus, a year was the period of time before the Sun returned to the same place in the sky, while the month was the period that elapsed between each new or full Moon.

The other five planets were rather less noticeable, though each had its own peculiar characteristics. For example, Mercury and Venus never strayed far from the Sun in the twilight skies of morning or evening. The other three moved more slowly from constellation to constellation, sometimes describing loops in the sky as they appeared to temporarily reverse direction.

It was also evident that the seven planets often came together in the sky or even passed behind the Moon during occultations. They always remained within a narrow band on the sky, known as the zodiac (after the Greek word for "animal"). The Sun's annual path across the sky, called the ecliptic, ran along the center of this celestial highway. Clearly, the planes of the planets' orbits were closely aligned with each other.

The Earth-Centered Universe

Until the mid-16th century, it was accepted as an established fact by most civilizations that Earth lay at the center of the Universe.[2] Like the axle of a wheel, everything else rotated around it.

[1] For a time, the ancient Greeks thought there were nine planets. Venus was named both as the Evening Star (Hesperus) and the Morning Star (Phosphorus). Similarly, Mercury was thought to be two different planets – Lucifer and Hermes.

[2] A Sun-centered, heliocentric model of the Universe was proposed by the Greek astronomer Aristarchus in the 3rd century BC, but it was not widely accepted.

Exploring the Solar System, First Edition. Peter Bond.

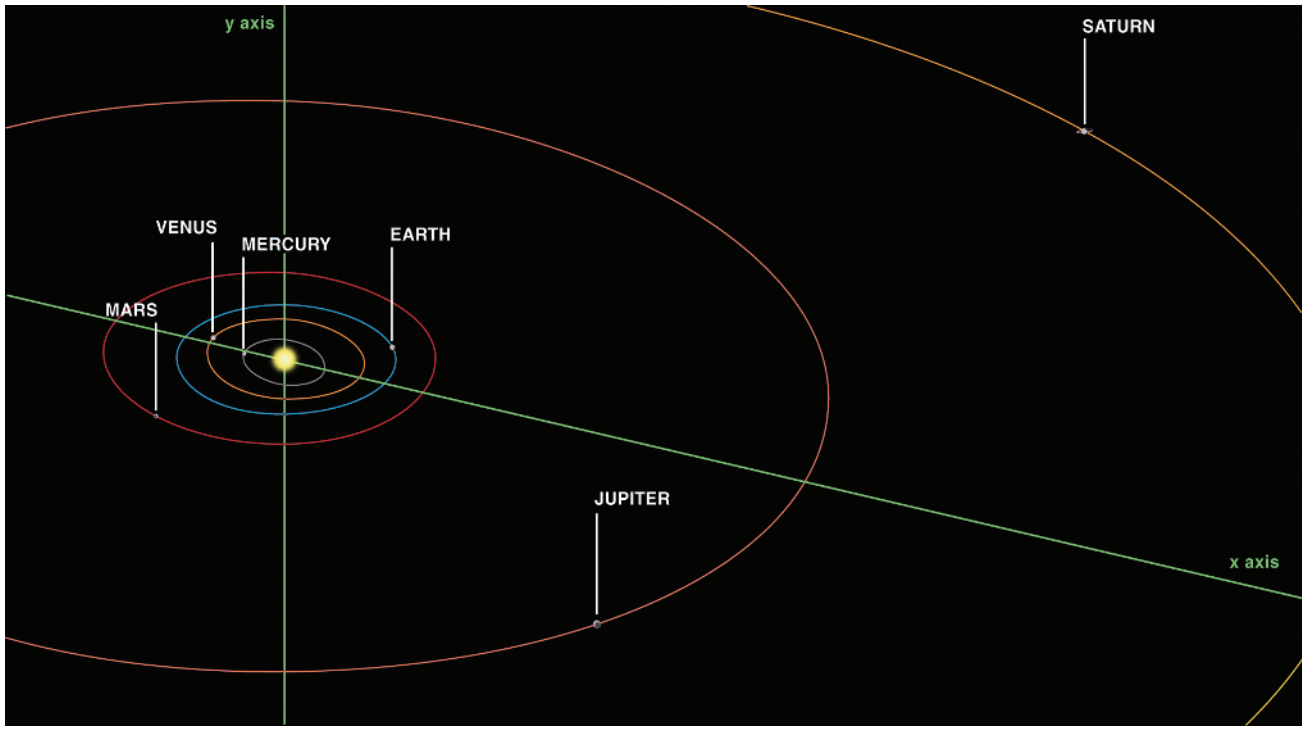


Figure 1.1 The relative sizes of the orbits of the seven "planets" (including the Sun and Moon) visible to the naked eye and recognized by ancient astronomers. All of the orbits are slightly elliptical and nearly in the same plane as Earth's orbit (the ecliptic). The diagram is from a view above the ecliptic plane and away from the perpendicular axis that goes through the Sun. (Lunar and Planetary Institute)

The reasons for this thinking seemed self evident. All of the celestial objects, including the Sun, moved across the sky from east to west (with the occasional exception of a comet or a shooting star). However, since no one experienced any of the sensations that would be expected if Earth was continually spinning, it seemed logical to believe that it was the heavens which were in motion around Earth.

According to this geocentric theory, the Sun, Moon and planets were carried by invisible, crystalline spheres which were centered on the Earth. A much larger celestial sphere carried the fixed stars around the central Earth once every day.

Although early civilizations accepted the visual evidence that Earth is (more or less) flat, this idea was contradicted by several lines of evidence (see Chapter 3). For example, different star patterns or constellations are visible from different places. However, if Earth is flat, then the same constellations should be visible everywhere at a certain time.

One key piece of evidence was the curved outline of Earth's shadow as it drifted across the face of the full Moon during a total lunar eclipse. This was the case no matter where the observation was made or at what time it took place. Since only a spherical body can cast a round shadow in all orientations, it seemed clear that Earth was round.

Similarly, observations of a sailing ship disappearing over the horizon showed that, instead of simply becoming smaller and smaller, its hull disappeared from view before the sails and mast. This could only be explained on a curved ocean.

Measuring Distances and Sizes

One of the most fundamental problems facing early astronomers was the scale of the Universe. How big were the Earth, Sun and Moon, and how far away were they? It seemed evident that Earth was huge compared with every other object, and since it was the home of humanity, it was assumed that Earth was pre-eminent.

The question of the size of the spherical Earth was solved in the 3rd century BC by Eratosthenes, who compared the length of shadows made at different locations at the time of the spring equinox (see Chapter 3). Some facts were also known about the relative sizes and distances of other objects.

Since its shadow easily covered the entire Moon during lunar eclipses, Earth had to be substantially larger than its satellite. During a solar eclipse, the Moon passed in front of the Sun, so the latter had to be further away. However, since their apparent sizes were identical, the Sun must be considerably larger than the Moon. Similarly, the Moon sometimes occulted or passed in front of stars and planets, so these, too, had to be much more remote.

Calculations by the Greek astronomers Aristarchus (c.310–c.230 BC) and Hipparchus (c.190–120 BC), based on the size of Earth's shadow, suggested that the Moon's diameter is about one third that of Earth and that its distance is nearly 59 times Earth's radius. This established the scale of the Earth-Moon system with a fair degree of accuracy. However, their simple geometric methods grossly underestimated the Sun's distance.

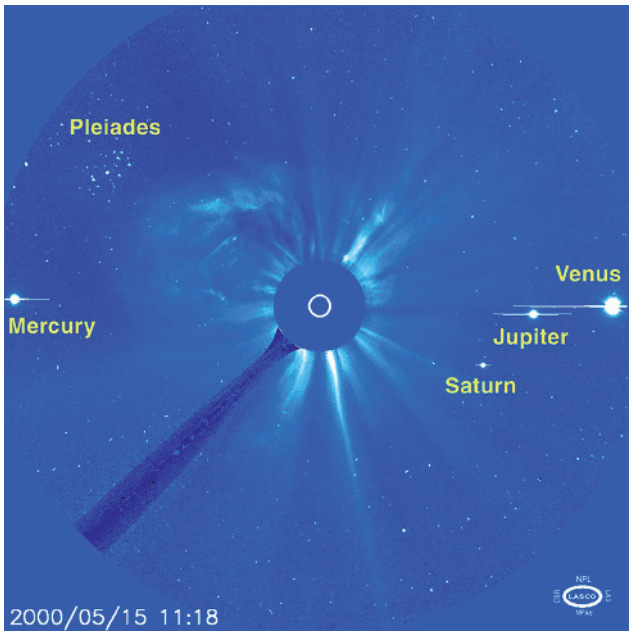


Figure 1.2 All of the major planets follow orbits that lie within 8° of the Sun's path across the sky – the ecliptic. This narrow celestial belt is known as the zodiac. In this image from the SOHO spacecraft, four planets appear close to the Sun (whose light is blocked by an occulting disk). Also in view are some background "fixed" stars, including the Pleiades cluster. (NASA-ESA)

Determination of the planetary distances remained problematical for a long time. It soon became clear to observers in the classical world that some planets move more slowly through the constellations of the night sky. Since a slow-moving planet such as Saturn was also fainter than the faster moving objects, Mars and Jupiter, it seemed logical that Saturn is further away from Earth.

It was also clear that the Sun, Moon and planets did not move at uniform speeds or follow simple curved paths across the sky. One of the most difficult observations to explain was an occasional "loop" in the motions of the more distant planets. This occurred when Mars, Jupiter and Saturn were shining brightly around midnight. At such times, the planet's nightly eastward ("prograde") motion would gradually come to a stop. It would then reverse direction toward the west, becoming "retrograde", before resuming its general movement toward the east.

The explanation for this motion had to wait until astronomers realized that the Sun was at the center of the planetary system, and that Earth orbited the Sun (see "The Central Sun" below). The loops could then be accounted for by Earth traveling along a smaller orbit so that it would catch up with, then overtake, the outer planets (see Fig. 1.3) – like an athlete on an inside track.

Accurate calculations of planetary distances also had to wait until the 17th century, when observers were able to measure angular distances with reasonable accuracy. The basic geometrical method they used was called parallax.

This involved measurement of the apparent shift in position of an object when viewed from two different locations. To illustrate this, hold one finger upright in front of your nose and close first one eye and then the other. The finger seems to shift position against the background, although it is, of course, stationary. When the finger is moved closer, the shift appears larger, and vice versa.

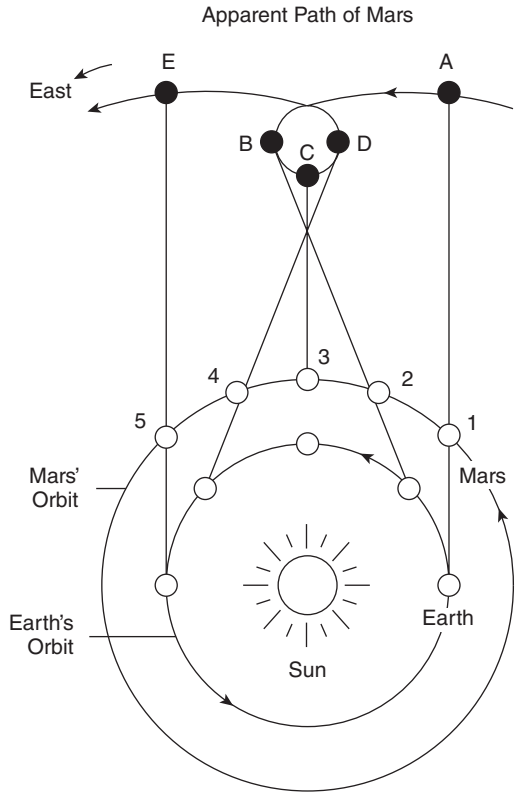


Figure 1.3 The apparent retrograde loops in the motions of Mars, Jupiter and Saturn are now known to be caused by the relative orbital movement of the planets and Earth. Since Earth moves faster along its orbit than the more distant planets, it overtakes them on the inside track. As Earth approaches and passes Mars, the slower moving outer planet (points 2 to 4) appears to move backward (points B to D) for a few months against the backcloth of "fixed" stars. (Kenneth R. Lang / Tufts University, NASA)

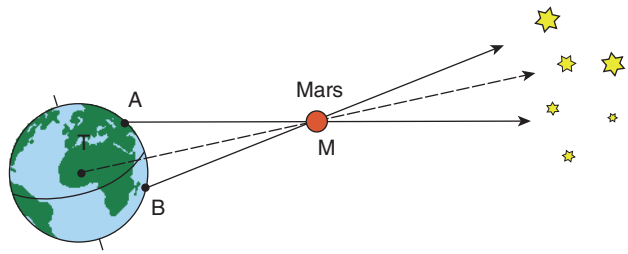


Figure 1.4 The distance of a planet such as Mars can be calculated by measuring its angle of sight – its location against the background of fixed stars – from two or more places on Earth. If the length of the baseline (e.g. the distance between two viewing sites, A–B) is known, the distance can be found by using simple trigonometry. (ESO)

Astronomers realized that, if a parallax shift in a planet's position could be measured from two widely separated locations, then its distance could be calculated. This method was first used by a French astronomer, Jean Richer, working in Cayenne (French Guiana) together with Giovanni Domenico Cassini and Jean Picard in Paris. They made simultaneous parallax observations of Mars during its closest approach in 1671, using the recently invented pendulum clocks to ensure that the measurements were made at precisely the same moment.[3]

Cassini's calculations led to a value of about 140 million km for the astronomical unit (AU) – the mean Sun-Earth distance. Now that this distance was known with reasonable accuracy, Kepler's third law (see Box 1.1) could be used to calculate the distances of the Sun and planets for the first time.

During the 18th century a great deal of time, money and effort was spent in attempting to refine these figures. One method was to observe rare transits of Venus across the face of the Sun from many different locations. The most famous transit observations took place in 1761 and 1769 when the British explorer, Captain James Cook, sailed to the Pacific as part of an army of 150 observers scattered across the globe, but these gave very inaccurate results (see Chapter 6).

More successful was the world-wide effort to determine the parallax of the asteroid Eros when it passed close to Earth in 1931. Highly accurate measurements were possible since Eros has no atmosphere and appears as a mere point of light in even the largest telescopes. The value of the astronomical unit turned out to be 149.6 million km.

Since then, more sophisticated techniques have been introduced to refine the scale of the Solar System. One of the most successful is radar, when radio signals are reflected from the surfaces of distant objects (see Chapters 5, 6 and 13). Since the velocity of these microwaves is known and the time taken between emission and reception can be measured to a fraction of a second, the distance can be readily calculated. (Radar has also revealed the sizes and shapes of hundreds of asteroids.) A similar technique used to calculate changes in the Earth-Moon distance involves the use of laser pulses bounced off special reflectors left on the lunar surface.

Once an object's distance is accurately known, the diameter can be determined from its apparent angular size, as seen in a telescope. Unfortunately, this is very difficult for the smaller or more distant members of the Solar System, particularly if their albedo, or surface reflectivity, is uncertain.

In general, the larger an object, the more light its surface reflects. However, some objects are much better mirrors than others. A small, reflective object can have the same apparent brightness as a large, dark object. New observations of some Kuiper Belt objects, beyond the orbit of Pluto, indicate that their albedos are greater than previously believed. Since they are more reflective than anticipated, astronomers have revised their diameters downwards.

Another method, involving the occultation of a star by a planet or other object, is especially valuable in relation to bodies which are normally difficult to observe. The object's diameter is calculated from the length of time during which it hides the star from view. Unfortunately, if it possesses a dense, cloudy atmosphere, the occultation only gives the diameter at the cloud tops.

The Central Sun

The difficult task of breaking with tradition and accepting the Sun as the center of the Universe began with a Polish priest and astronomer named Nicolaus Copernicus (1473–1543). He decided that the only way to make sense of the planetary orbits was to relegate Earth to the status of a planet that orbited the Sun. The movement of the stars across the sky was then explained by the rotation of the spherical Earth, while the calendar of seasons and changing constellations in the heavens were accounted for by its year-long journey around the Sun.

Copernicus' most significant work, called *De Revolutionibus Orbium Celestium* (Concerning the Revolutions of the Celestial Spheres), was published shortly before his death. Curiously, this did not provoke a violent reaction by the establishment of the day, nor did it immediately lead to any major upheaval in scientific thought. Lacking enough evidence to swing the argument one way or the other, the great minds of the day were faced with an impasse.

Half a century passed before the interventions of two great scholars swung the argument in favor of Copernicus' heliocentric theory. The first breakthrough was made in 1609 by a young German named Johannes Kepler. By one of those strange twists of irony, Kepler was a pupil of Tycho Brahe, one of the leading opponents of the Copernican order. Given the unenviable task of finding an explanation for the retrograde motion of Mars (see Fig. 1.3), Kepler was able to draw upon the excellent observational data recorded by his employer.

Brahe died in 1601, but Kepler continued to laboriously examine the problem before finally arriving at his eureka moment. The planetary orbits, he declared, were not circles but ellipses.[4] Within a short time, Kepler was able to draw up the first two laws of planetary motion (see Box 1.2). His third, and probably most important law, followed in 1619.

As a result, the relative distance of each planet from the Sun could be calculated accurately. Saturn, the most remote planet known at the time, turned out to be nearly ten times further from the Sun than Earth. Since the actual distances remained unknown, the standard unit of measurement became the astronomical unit.

In the same year that Kepler discovered elliptical orbits, an Italian scientist named Galileo Galilei made a simple refracting telescope, comprising two lenses at either end of a narrow tube, and began to study the heavens. Within a short time, he had obtained visual evidence to support the theories of Copernicus and Kepler. Galileo became the first person in history to see the phases of Venus caused by its movement around the Sun. He also observed mountains and craters on the Moon, and saw the planets as disks, rather than points of light.

Most significant of all was his discovery of four star-like objects close to Jupiter. By watching their daily motions, he was able to

[3] A by-product of this experiment was the discovery that a pendulum swung more slowly at Cayenne than at Paris, showing that gravity is slightly weaker at the equator. Newton later used this result to show that Earth's diameter is greatest at the equator.

[4] Kepler's task was made slightly easier by the fact that, of the five known planets, only Mercury followed a more elliptical path than Mars.

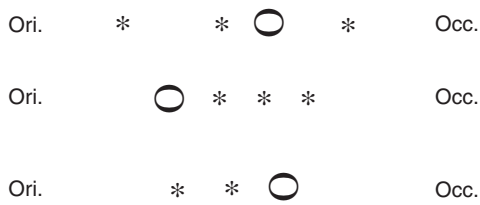


Figure 1.5 In January 1610, Galileo Galilei used his simple refracting telescope to discover three "stars" aligned on either side of Jupiter. Over a period of several weeks, a fourth "star" appeared. As they shifted positions, Galileo correctly deduced that these were satellites. *Occ.* is the Latin abbreviation for "west" and *Ori.* Stands for "east."

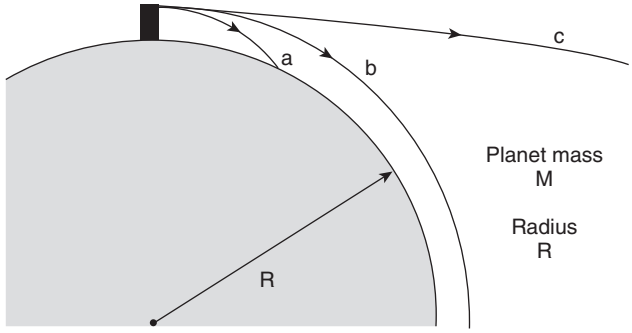


Figure 1.6 **(a)** If a spacecraft does not accelerate to orbital velocity, it will fall back to the planet's surface. **(b)** If it reaches orbital velocity, it will remain in a closed path (orbit) around the planet under free fall conditions. **(c)** If the spacecraft reaches escape velocity, it will be able to break free from the planet's gravitational pull and travel to another planet. The same rules apply to planets and spacecraft in orbit around the Sun.

calculate their orbital periods and show that they were Jovian moons (see Chapter 7). The discovery of the first planetary satellites (other than the Moon) supported theories that Earth was not at the center of the Universe and that everything did not revolve around our world.

Galileo's discoveries caused a sensation, although the leaders of the Roman Catholic Church obstinately continued to support a geocentric Universe. In 1633, Galileo was brought before the Inquisition and forced to recant under threat of torture.

Newton and Gravity

The next challenge was to find an explanation for Kepler's laws. Although Galileo conducted numerous experiments into the effects of gravity, he did not realize the full significance of his discoveries. This was left to an Englishman, Isaac Newton, who was born in 1642, the year that Galileo died.

One anecdote attributes Newton's discovery of universal gravitation to him observing an apple falling from a tree. Whatever the truth, by 1684 Newton was able to explain planetary motions. His law of gravitation stated that all objects attract each other, and that the strength of this gravitational attraction is proportional to their mass (see Chapter 8).

Clearly, since the Sun has nearly all of the mass in the Solar System, it should pull all of the other bodies into it. Newton explained that this did not happen because their orbital velocities are just sufficient to counteract the Sun's gravity. The result is that the planets fall towards the Sun in such a way that the curve of their fall takes them completely around it. This is sometimes known as free fall. (The same explanation, of course, applies to artificial satellites.)

Newton's law also stated that the strength of gravitational attraction decreases with distance. For example, if planet A is twice as far from the Sun as planet B, then the gravitational force exerted by the Sun on planet A is one quarter that exerted on planet B. In practical terms, this means that a satellite in low Earth orbit must travel at 8 km/s, whereas the Moon only has to circle the Earth at 1 km/s in order to avoid crashing into our planet. Similarly, planets further from the Sun are able to move more slowly around their orbits than those in the inner Solar System. Newton's law also explained why a planet's orbital speed increased as it approached perihelion (closest point to the Sun) and slowed near aphelion (furthest point from the Sun).

From this time on the orbital mechanics of the Solar System were very well understood. With the exception of Mercury, whose orbital motion refused to obey Newton's law (see Chapter 5), the only significant problems involved minor variations in orbits caused by gravitational interactions between the planets, particularly those involving massive Jupiter. Careful study of unexpected changes in the orbital velocity of Uranus even enabled the position of an unknown planet, Neptune, to be successfully calculated (see Chapter 11) – although there are those who consider the discovery to be pure chance.

What is a Planet?

In the ancient world, astronomers counted eight planets. When the Sun, Earth and Moon are removed from their list, the number of planets visible to the naked eye is reduced to five: Mercury, Venus, Mars, Jupiter and Saturn.

With the invention of the telescope, the possibility arose of finding fainter, more remote planets. The first newcomer, Uranus, was discovered far beyond the orbit of Saturn by William Herschel in 1781. The list was further increased in 1801, when Giuseppe Piazzi found Ceres in the gap between the orbits of Jupiter and Mars. Pallas, Juno and Vesta – objects in similar orbits to Ceres – were discovered between 1802 and 1807. Since they were clearly much smaller and less substantial than the other planets, they were soon downgraded to "minor planets" or "asteroids" (star-like objects).

Almost 40 years passed before the eighth planet, Neptune, was discovered by Johann Galle and Heinrich D'Arrest. However, neither Uranus nor Neptune seemed to be following its expected path, suggesting that an even more distant planet might be

Box 1.1 Orbits

The direction a spacecraft or other body travels in orbit can be prograde, when a satellite moves in the same direction as the planet (or star) rotates, or retrograde, when it goes in a direction opposite to the planet's (or star's) rotation. All of the planets orbit the Sun in a prograde direction – west to east or counterclockwise as observed from above the Sun's north pole. However, many comets move in a retrograde (clockwise) direction.

Various technical terms are used to describe the characteristics of these orbits. The time an object takes to complete one orbit is known as the **orbital period**. The closest point of an orbit has the prefix "peri" – hence perigee for a satellite of the Earth and perihelion for an object orbiting the Sun. (Helios = Sun) The furthest point in an orbit has the prefix "ap" – as in apogee and aphelion.

The plane of Earth's orbit around the Sun is called the ecliptic.The orbits of the other planets, comets and asteroids are tilted to this plane. The angle of the tilt is the **orbital** inclination. The inclination of a satellite's orbit is measured with respect to the planet's equator. Hence, an orbit directly above the equator has an inclination of 0°, while one passing over a planet's poles has an inclination of 90°.

A planet, asteroid or comet crosses the ecliptic twice during each orbit of the Sun. The points where an orbit crosses a plane are known as nodes. When an orbiting body crosses the ecliptic plane going north, the node is referred to as the **ascending node**. Going south, it is the **descending node**. The line that joins the ascending node and the descending node of an orbit is called the **line of nodes**.

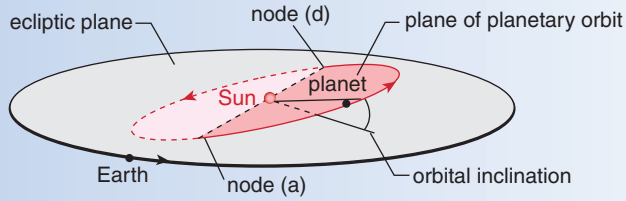


Figure. 1.7 Some important characteristics of a planet's orbit. Here the planet is inferior, that is, closer to the Sun than Earth. Its orbit is inclined to the ecliptic – the plane of Earth's orbit. The planet's orbit crosses the ecliptic at two nodes – the ascending node (a) and the descending node (d). (After Open University)

One of the most important orbital, or Keplerian, elements, is the semimajor axis, the average distance of an object from its primary (planet or Sun). The shape of the orbit is described by its eccentricity, measured as a number between zero and 1. An eccentricity of zero indicates a circular orbit. A parabola has an eccentricity of 1.

influencing the movements of its neighbors. The search for this world concluded in 1930 when Clyde Tombaugh recorded the tiny image of Pluto.

For many years, it was generally accepted that there were nine planets, despite growing concerns that Pluto seemed to be too small and lacking in mass to deserve this title. The crunch came in 2003, when Mike Brown discovered 2003 UB313 (now named Eris), an object that is comparable in size to Pluto. With the introduction of ever more sensitive detectors, it seemed likely that there would soon be dozens of Pluto-sized planets.

Aware that there was no generally accepted definition of the term "planet," and faced with a fierce debate over whether Pluto should be demoted, members of the International Astronomical Union gathered in Prague for the 2006 General Assembly. After a lengthy discussion, they agreed to define a planet as a celestial body that: (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighborhood around its orbit.

Based on these criteria, the Solar System now consists of eight planets: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune. A new distinct class of objects called "dwarf planets" was also introduced. To be classified as a dwarf planet, an object must orbit the Sun and have a nearly round shape. The first dwarf planets to be announced were Ceres (the largest asteroid), Pluto and Eris, although many more are expected to be discovered in the future.

This decision has not met with universal approval. One common criticism relates to what exactly is meant by a planet "clearing its neighborhood." For example, critics argue that Neptune is accepted as a planet, even though many Kuiper Belt objects (including Pluto) cross its orbit. Perhaps, they suggest, it would be more appropriate to use size as a criterion, particularly

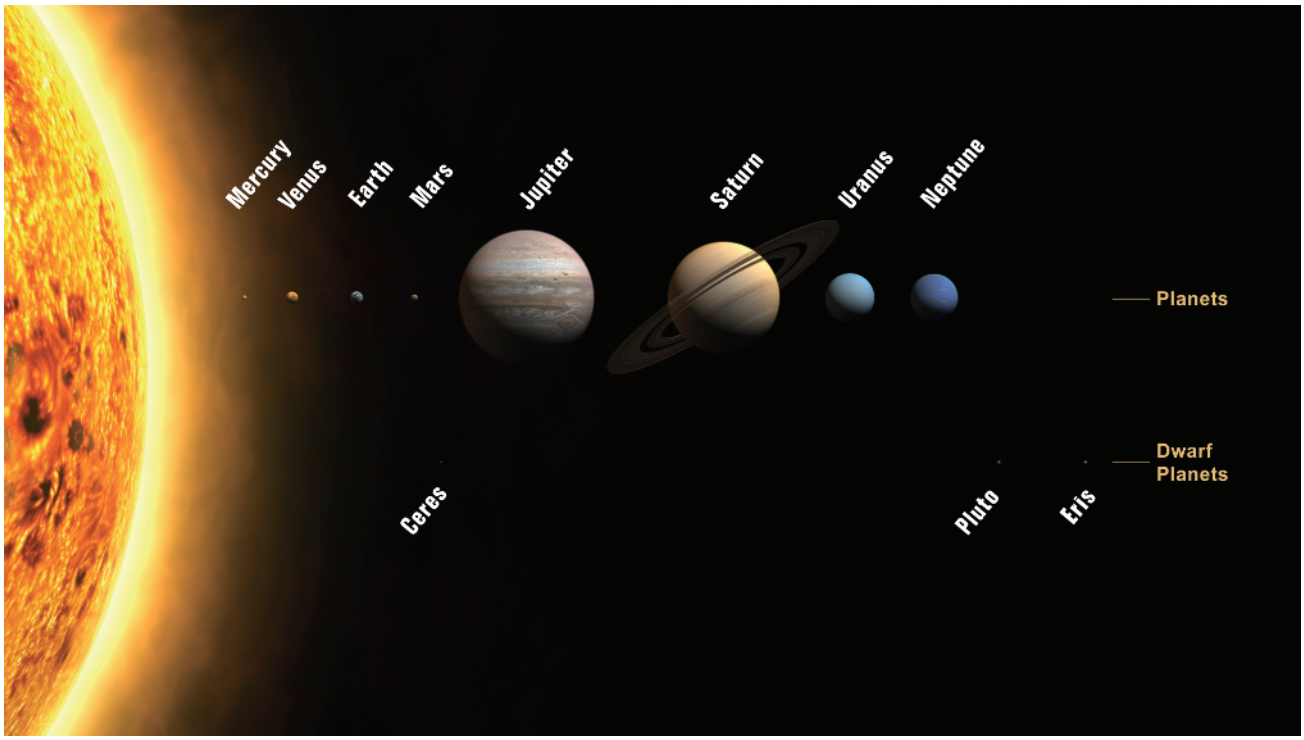


Figure 1.8 In the "new" Solar System, as defined by the International Astronomical Union in 2006, there are eight planets: Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus and Neptune (see here in order of their distance from the Sun). A new, distinct class of objects called "dwarf planets" includes the largest asteroid, Ceres, and the two largest known Kuiper Belt objects, Pluto and Eris. The relative sizes of the planets and the Sun are shown. Jupiter's diameter is about eleven times that of Earth, and the Sun's diameter is about ten times that of Jupiter. The distances of the planets are not shown to scale. (IAU)

bearing in mind the diameters of objects that are large enough for gravity to dominate structural strength. There is also some discomfiture with defining Ceres – the largest of the asteroids – as a dwarf planet.

Another complication arises when the current definition is extended to extrasolar planets, that is, planets orbiting other stars (see Chapter 14). Size is not a useful factor, since many of these planets are similar in size and mass to small, cool "failed stars" known as brown dwarfs. Instead, astronomers attempt to distinguish between a giant extrasolar planet and a brown dwarf by determining how they were born. A star is formed during the gravitational collapse of a gaseous nebula, whereas a planet is the product of collisions and accretion (snowball-like growth) between particles in a disk of gas and dust around a central star. Even so, this method of differentiation is difficult to apply, especially in the case of planet-sized objects that have been flung into interstellar space and no longer orbit any star.

The Solar System

Fifty years ago, the population of the Solar System included one central star, nine planets, 31 satellites and thousands of comets and asteroids. However, since the arrival of the Space Age and the development of ever more sensitive ground-based instruments, the inventory of objects has swollen remarkably. Today, the astronomical community recognizes eight planets and five dwarf planets, the tally of planetary satellites has passed 150 and the number of identified small objects is climbing rapidly as increasingly sensitive searches discover thousands of Sun-grazing comets and icy Kuiper Belt objects that orbit beyond Neptune.

In terms of numbers, the Solar System is dominated by debris, in the form of comets, asteroids, meteorites and dust. These are the left-overs from the formation of the planets, four and a half billion years ago. The main asteroid belt, between Mars and Jupiter, is populated by millions of rocky objects that are shepherded by the powerful gravity of the nearby gas giant. They are thought to represent planetesimals – small planetary building blocks – that were unable to accrete due to the gravitational interference of Jupiter.

Beyond the orbit of Neptune are two more swarms of small objects, this time largely made of ice. The inner region, known as the Kuiper Belt, is where short-period comets originate. Pluto and Eris are the largest known inhabitants. The orbital periods of Kuiper Belt objects range from 200–400 years for objects such as Pluto to 1 000 years or longer for those that follow very elliptical orbits that take them far from the Sun.

The Kuiper Belt poses a serious challenge for theories of planet formation, since it contains less than one percent of the mass expected from the protosolar nebula theory. If the Kuiper

Table 1.1
The Planets: Relationship between solar distance and mean density.

Planet	Distance from Sun (AU)	Mean Density (g/cm^3)
Mercury	0.3871	5.43
Venus	0.7233	5.24
Earth	1.0	5.52
Mars	1.5237	3.91
Jupiter	5.2028	1.33
Saturn	9.5388	0.69
Uranus	19.1914	1.29
Neptune	30.0611	1.64

Belt objects formed like the terrestrial planets, growing by accumulating smaller objects as they orbit the Sun, it would take longer than the age of the Solar System to make one KBO!

Even further out – indeed, so far that none of the objects have ever been observed – is the postulated Oort Cloud, the home of most long-period comets.

The basic characteristics of the Solar System are straightforward to describe. Close to the Sun, where temperatures are higher, there are four quite small, but dense, "terrestrial" planets that are composed of rock. Beyond Mars, where temperatures are always well below zero, is the realm of the gas giants, Jupiter and Saturn, and the ice giants, Uranus and Neptune.

The orbits of the major planets are approximately circular, and close to the ecliptic plane. All of the planets and main belt asteroids circle the Sun in the same direction – counter-clockwise as seen from above the Sun's north pole. This is also the direction of the Sun's rotation. However, the beautiful symmetry breaks down when it comes to the smaller, more icy members of the Solar System. Comets can arrive from any direction, and the orbits of the Kuiper Belt objects have no particular orientation, suggesting that there is a spherical swarm of these objects surrounding the Sun and major planets.

Of the four inner planets, Venus and Earth both possess dense atmospheres – though they are very different in nature – while Mercury is too lightweight to have retained a substantial gaseous envelope. Whereas the most common gas on both Venus and Mars is carbon dioxide, Earth is something of an oddball, with an atmosphere dominated by nitrogen and oxygen. This latter gas can be accounted for by the fact that Earth is – as far as we know – the only abode of life in our Solar System. Satellites are rare: Earth is orbited by the Moon, while Mars has two small companions that are generally considered to be captured asteroids.

As their name suggests, the giants are characterized by their large size – tens to thousands of times bigger than Earth – and low bulk densities which can be accounted for by the dominance of hydrogen and helium in their interiors. All four of the giants have ring systems composed of dust, ice and rocky debris, and their gravitational influence is such that they retain dozens of satellites – most of them captured billions of years ago.

Since they are relatively close to the Sun, all of the terrestrial planets have high orbital velocities with periods of less than two Earth years (see Box 1.2). In contrast, their axial rotations are slow and their axial inclinations are very different.

Mercury's axis is almost at right angles to its orbit. It takes 58 days to rotate once, or about two thirds of the time it takes to orbit the Sun. Venus resembles a top that has been knocked completely upside down. As a result, it rotates in a retrograde direction that takes 243 Earth days, longer than its orbital period. Earth and Mars have very similar days and seasons – at least in the present epoch – since their sidereal periods of axial rotation are both around 24 hours and both axes are inclined about 24–25° to their orbits.

The motions of the outer planets are very different. Their large distances from the Sun require modest velocities to maintain their orbits. Orbital periods range from almost 12 years for Jupiter to about 165 years for Neptune. However, despite their swollen spheres, they all spin much faster on their axes than their terrestrial siblings, with sidereal periods in the range 10–20 hours.[5] However, there is considerable variation in their axial tilts. Jupiter is almost upright, Saturn and Neptune are inclined more than Earth and Mars, while Uranus spins on its side so that the polar regions alternately point toward or away from the Sun.

The Birth of the Solar System – Early Ideas

The Sun, which contains over 99% of the system's mass, completes one rotation in about 24 days. In contrast, the largest planets, Jupiter and Saturn, rotate once in about ten hours. When combined with their orbital motion, it turns out that Jupiter accounts for some 60% of the Solar System's angular momentum, with another 25% accounted for by Saturn. This compares with about 2% for the sluggardly Sun.

Any theory of cosmogony that attempts to account for the formation of the Solar System must take into account the angular momentum of the Solar System objects, as well as the facts that all of the planets travel in the same direction and more or less in the same plane. The obvious conclusion is that they all formed in the same manner and at about the same time.

There seem to be two main possibilities: the planets were either created by material derived from the Sun or a nearby companion star, or they formed from a cloud of diffuse matter that surrounded the Sun. However, theorists have struggled for centuries to match the hypotheses to the known facts, in order to choose between them.

One of the earliest, and most successful, attempts to explain how the Solar System came about was the nebular hypothesis – the idea that the Sun and planets formed from a vast, slowly rotating disk of gas and dust. A modified version of this hypothesis is the generally accepted explanation today.

[5] **The sidereal period is the time a planet takes to orbit the Sun, with respect to a particular background star.**

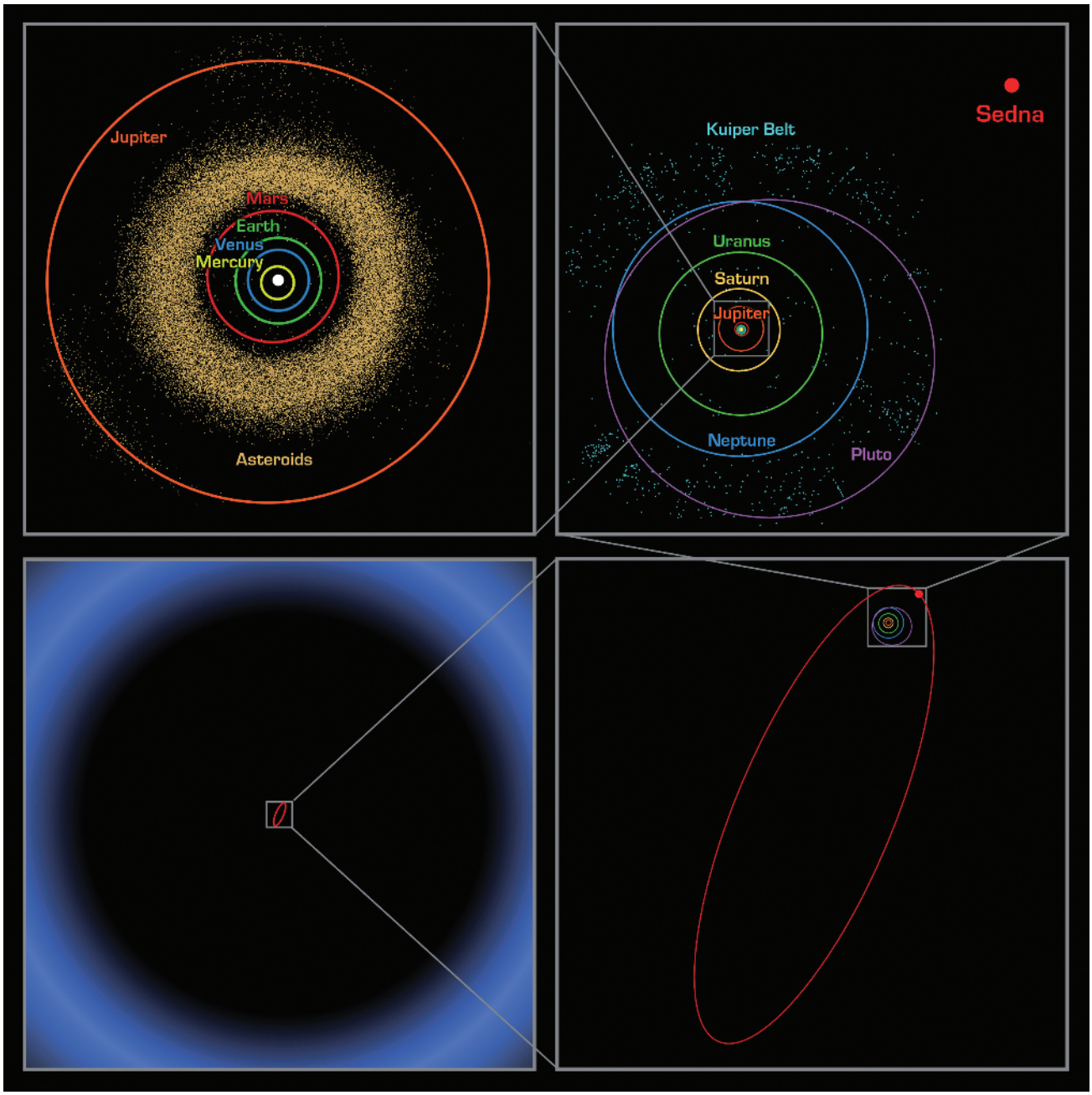


Figure 1.9 These four panels show the scale of the Solar System as we know it today. At top left are the orbits of the inner planets and the main asteroid belt. Top right shows the orbits of the outer planets and the Kuiper Belt. Lower right shows the orbit and current location of Sedna, which travels further from the Sun than any other known object in the Solar System. Lower left shows that even Sedna's highly elliptical orbit lies well inside the inner edge of the proposed Oort Cloud (shown in blue). This spherical cloud contains millions of icy bodies orbiting at the limits of the Sun's gravitational pull. (NASA/JPL/R. Hurt, SSC-Caltech)

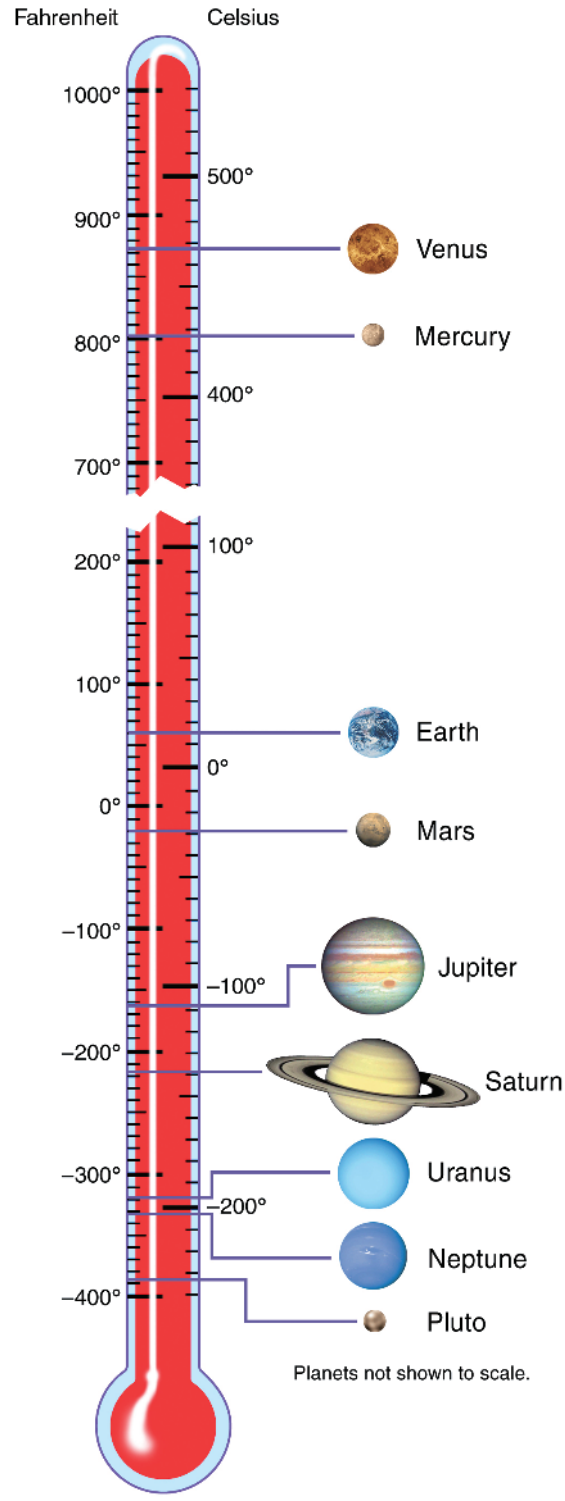


Figure 1.10 In general, a planet's surface temperature decreases with its distance from the Sun. Venus is the exception, since its dense carbon dioxide atmosphere traps infrared radiation. The runaway greenhouse effect raises its surface temperature to 467°C. Mercury's slow rotation and thin atmosphere result in the night-side temperature being more than 500°C colder than the day-side temperature shown above. Temperatures for Jupiter, Saturn, Uranus, and Neptune are shown for an altitude in the atmosphere where pressure is equal to that at sea level on Earth. Earth lies in the center of the "habitable zone," where water can exist as a liquid and conditions are favorable to life. (Lunar and Planetary Institute)

Some of the key evidence comes from modern observations of distant star systems. Today, spaceborne telescopes can peer into the hearts of giant molecular clouds, such as the Orion Nebula, and search for young, Sun-like stars that replicate the conditions that prevailed in our Solar System some 4.6 billion years ago.

These observations show that so-called protoplanetary disks, or proplyds, exist around most very young stars – those less than ten million years old. Many of the disks are larger than our Solar System. Observations of slightly older stars show how these disks evolve as time goes by, with the formation of swarms of rocky and icy debris and gaps in the clouds created by fledgling planets.

As currently envisaged, the Solar System began with the collapse of a cloud of interstellar gas. The trigger for this collapse may have been the passage of an externally generated shock wave – perhaps from a supernova explosion, density waves passing through the galaxy, or a major reduction in the cloud's magnetic field or temperature. The first of these explanations is a prime candidate, since many stars form in clusters within clouds containing thousands of solar masses of material. When the giant stars of the cluster run through their short life spans, they are likely to produce a series of supernovas, preceded by powerful stellar winds.

Over millions of years, the original cloud may be broken up into smaller fragments, each mixed with heavier elements from the dying stars, as well as the ubiquitous hydrogen and helium gas. Once a fragment reaches a critical density, it is able to overcome the forces associated with gas pressure and begins to collapse under its own gravity.

The contracting cloud begins to rotate, slowly at first, then faster and faster – rather like when an ice skater pulls in his arms. Since material falling from above and below the plane of rotation collides at the mid-plane of the collapsing cloud, its motion is cancelled out. The cloud begins to flatten into a disk, with a bulge at the center where the protostar is forming. The disk was probably thicker at a greater distance from the protostar, where gas pressure was lower.

Such a nebula would almost certainly rotate slowly in the early stages, but as it contracts, conservation of angular momentum causes the cloud to spin faster. If this process continues, the core forming at the center of the nebula will spin up so fast that it flies apart before it has a chance to form a star. Somehow, that angular momentum must be removed before a star can form.

Studies of other young stars and their surrounding disks provide evidence that, as the interstellar gas collapses, it also winds up the magnetic field which permeates the nebula. Gas which is rotating too fast to collapse is expelled and dispersed along the magnetic field.

This process naturally forms a spiral-shaped magnetic field that helps to generate polar jets and outflows associated with very young stars. At the same time, the jets remove angular momentum, allowing other material to accrete and collapse. Gravitational instability, turbulence and tidal forces within the "lumpy" disk may also play a part, helping to transfer much of the angular momentum to the outer regions of the forming disk.

The protoplanetary disk is heated by the infall of material. The inner regions, where the cloud is most massive, become hot enough to vaporize dust and ionize gas. As contraction continues and the cloud becomes increasingly dense, the temperature at its core reaches the point where nuclear fusion commences. The merging protostar begins to emit copious amounts of ultraviolet radiation. Radiation pressure drives away much of the nearby dust, causing the star to decouple from its nebula.

The youngster may remain in this T Tauri stage for perhaps 10 million years, after which most of the residual nebula has evaporated or been driven into interstellar space. All that remains of the original cloud is a rarefied disk of dust particles, mainly silicates and ice crystals.

Meanwhile, the seeds of the planets have begun to appear. More refractory elements condense in the warm, inner regions of the nebula, while icy grains condense in the cold outer regions. Individual grains collide and stick together, growing into centimeter-sized particles. These swirl around at different rates within the flared disk, partly due to turbulence and partly as the result of differences in the drag exerted by the gas. After a few million years, these dusty or icy golf balls accrete into kilometer-sized planetesimals and gravity becomes the dominant force.

The Solar System now resembles a shooting gallery, with objects moving at high speed in chaotic fashion and enduring frequent collisions with each other. Some of these impacts are destructive, causing the objects to shatter and generate large amounts of dust or debris. Other collisions are constructive, resulting in a snowballing process. Over time, the energy loss resulting from collisions means that construction eventually dominates.

Eventually, the system contains a relatively small number of large bodies or protoplanets. Millions of years pass as they continue to mop up material from the remnants of the solar nebula and to collide with each other, finally resulting in a population of widely separated worlds occupying stable orbits and traveling in the same direction around the young central star.

It is worth noting here that computer simulations of the early Solar System show that even the slightest differences in initial conditions can produce different planetary systems. Depending on exactly where each embryo started out, the orbital positions of new planets vary randomly from simulation to simulation. The total number of planets – and hence, their final masses – may also vary greatly. It seems that planet formation is a very chaotic process.

Rocky Planets

Modeling suggests that collisions between planetesimals initially occur at low velocities, allowing them to merge and grow. At the Earth's distance from the Sun, it takes only about 1 000 years for 1 km sized objects to grow into 100 km objects. Another 10,000 years produces 1 000 km diameter protoplanets, which double in diameter over the next 10,000 years. Such models indicate that Moon-sized objects can form in a little over 20,000 years.

As planetesimals within the protosolar disk grow larger and more massive, their gravity increases, and once a few of the objects reach a size of 1 000 km, they begin to stir up the remaining smaller objects. Near encounters accelerate the smaller, asteroid-sized chunks of rock to higher and higher speeds. Eventually, they

Box 1.2 Kepler's Laws of Planetary Motion

Johannes Kepler (1571–1630) was one of the most important characters in the story of unraveling how the Solar System works. The German-born mathematician was appointed assistant to Tycho Brahe (1546–1601), the most famous observer of the day. Granted access to Brahe's catalog of positional data, Kepler was given the task of explaining the orbit of Mars. After four years of calculations, Kepler finally realized in 1605 that the orbits of the planets were not perfect circles, but elongated circles known as ellipses.

Whereas a circle has one central point, an ellipse has two key interior points called foci (singular: focus). **The sum of the distances from the foci to any point on the ellipse is a constant**. For Solar System objects, the Sun always lies at one focus (see below).

In order to draw an ellipse, place two drawing pins some distance apart and loop a piece of string around them. Place a pencil inside the string, draw the string tight and move the pencil around the pins. Now move one of the pins and repeat the process. Note how the shape of the ellipse has changed.

The amount of "stretching" or "flattening" of the ellipse is termed its eccentricity. All ellipses have eccentricities lying between zero and one. A circle may be regarded as an ellipse with zero eccentricity. As the ellipse becomes more stretched, its eccentricity approaches one.

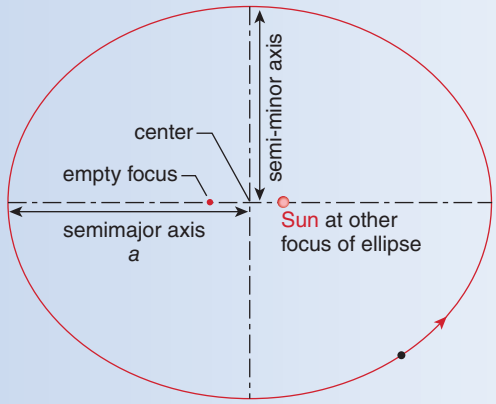


Figure 1.11 A circle has an eccentricity of zero. As the ellipse becomes more stretched (i.e. the foci move further apart) the eccentricity approaches one. Half of the major axis is termed the semimajor axis. The average distance of a planet from the Sun as it follows its elliptical orbit is equal to the length of the semimajor axis. The eccentricity is calculated by dividing the distance between the two foci by the length of the major axis. (Open University)

In reality, most of the planets follow orbits that are only slightly elliptical. Their eccentricities are so small that they look circular at first glance. Pluto and Mercury are the main exceptions, with eccentricities exceeding 0.2.

Another key characteristic of an ellipse is its maximum width, known as the **major axis**. Half of the major axis is termed the semimajor axis. The average distance of a planet from the Sun as it goes around its elliptical orbit is equal to the length of the semimajor axis.

After intensive work on the implications of his discovery, Kepler eventually formulated his Three Laws of Planetary Motion.

- Kepler's First Law: The orbits of the planets are ellipses, with the Sun at one focus of the ellipse. (Generally, there is nothing at the other focus).
- Kepler's Second Law: The line joining the planet to the Sun sweeps out equal areas in equal times as the planet travels around the ellipse. In order to do so, a planet must move faster along its orbit near the Sun and more slowly when it is far away. A planet's point of nearest approach to the Sun is termed perihelion; the furthest point from the Sun on its orbit is termed aphelion. Hence, a planet moves fastest when it is near perihelion and slowest when it is near aphelion.
- Kepler's Third Law: The square of a planet's sidereal (orbital) period is proportional to the cube of its mean distance (semi-major axis) from the Sun. This means that the period, or length of time a planet takes to complete one orbit around the Sun, increases rapidly with its distance from the Sun. Thus, Mercury, the innermost planet, takes only 88 days to orbit the Sun, whereas remote Pluto takes 248 years to do the same.

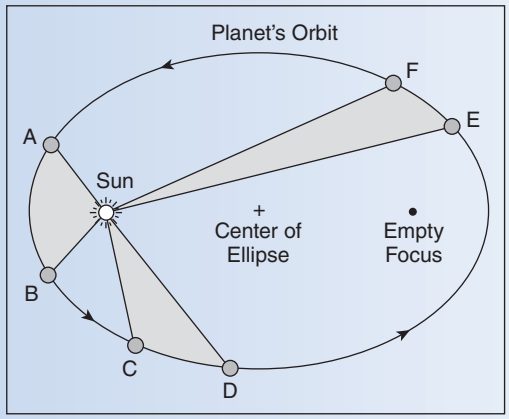


Figure 1.12 Kepler's first law states that the orbit of a planet about the Sun is an ellipse with the Sun at one focus. The other focus of the ellipse is empty. According to Kepler's second law, the line joining a planet to the Sun sweeps out equal areas in equal times. In this diagram, the three shaded areas ABS, CDS and EFS all have equal areas. A planet takes as long to travel from A to B as from C to D and E to F. It moves most rapidly when it is nearest the Sun (at perihelion) and slowest when it is farthest from the Sun (at aphelion). (Kenneth R. Lang, Tufts University)

This law can be used to make some useful, but fairly simple, calculations. For example, if the period is measured in Earth years and the distance is measured in astronomical units (AU), the law may be written in the simple form: $P(\text{years})^2 = R(\text{AU})^3$.

This equation may also be written as: $P(\text{years}) = R(\text{AU})^{3/2}$. Thus, if we know that Pluto's average distance from the Sun (semi-major axis) is 39.44 AU, we can calculate that the orbital period is: $P = (39.44)^{3/2} = 247.69$ years. Similarly, if we know that Mars takes 1.88 Earth years to orbit the Sun, we can calculate that its semimajor axis is: $R = (1.88)^{2/3} = 1.52$ AU.

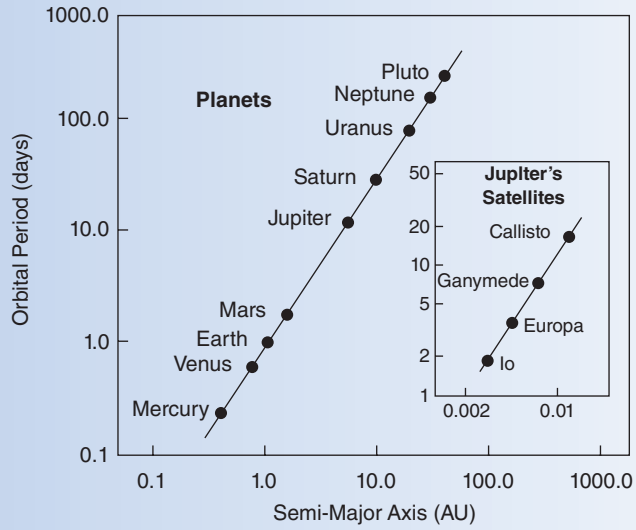


Figure 1.13 A graph showing the orbital periods of the planets plotted against their semimajor axes, using a logarithmic scale. The straight line that connects the planets has a slope of 3/2, verifying Kepler's third law which states that the squares of the orbital periods increase with the cubes of the planetary distances. This law applies to any bodies in elliptical orbits, including Jupiter's four largest satellites (inset). (Kenneth R. Lang, Tufts University)

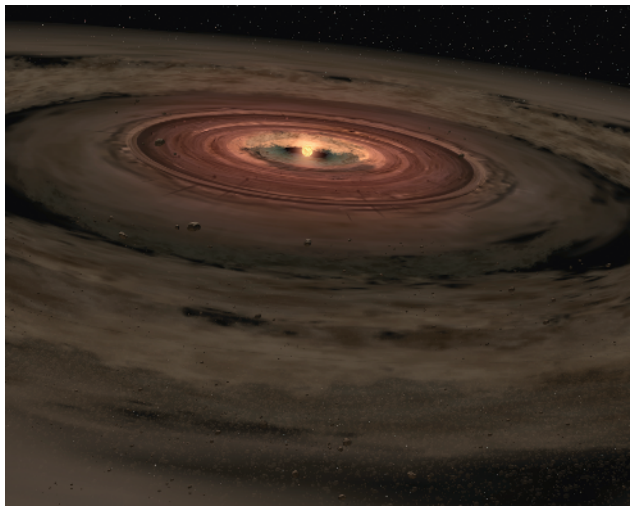


Figure 1.14 The planets formed about 4.5 billion years ago from a huge nebula – cloud of gas and dust – that surrounded the young Sun. Within a few million years, colliding particles in the nebula accreted until golf-ball-sized pebbles appeared. Further collisions caused these to increase in mass, eventually growing into the planets we see today. Some of these, further from the Sun, were able to pull in huge atmospheres of hydrogen and helium. Those in the warmer, inner regions were made of rock rather than ices and light gases. The remnants formed comets and asteroids. (NASA/JPL-Caltech/T. Pyle, SSC)

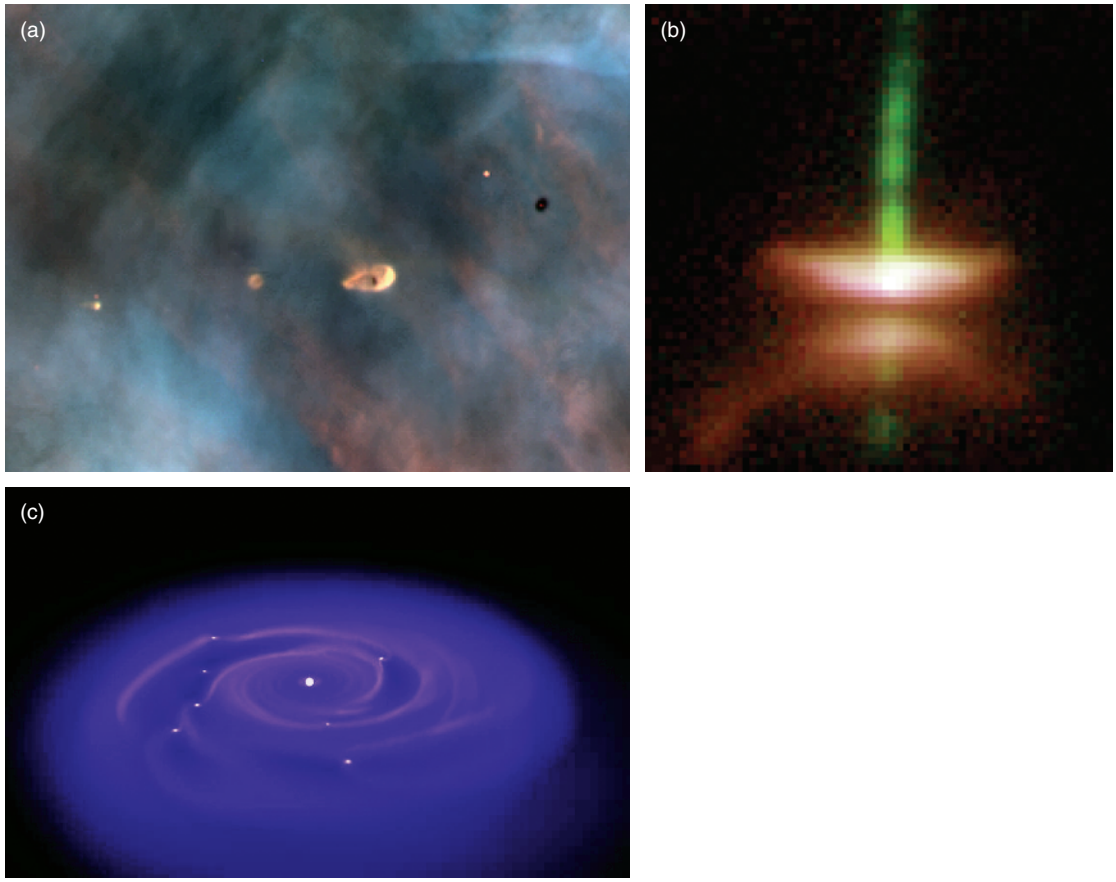


Figure 1.15 The early stages of star and planet formation: **(a)** A Hubble Space Telescope (HST) view of five young stars in the Orion Nebula. Four are surrounded by gas and dust trapped in orbit as the stars formed. These are possibly protoplanetary disks, or "proplyds," that might eventually produce planets. The bright proplyds are closest to the hottest stars of the parent star cluster, while the object farthest from the hottest stars appears dark. (C.R. O'Dell/Rice University; NASA) **(b)** This HST image shows Herbig-Haro 30, a young star surrounded by a thin, dark disk. The disk extends 64 billion km, dividing the nebula in two. The central star is hidden from direct view, but its light reflects off the upper and lower surfaces of the disk to produce the pair of reddish nebulas. Gaseous jets (green) remove material from above and below the disk and transfer angular momentum outwards. **(c)** A computer simulation showing how a protoplanetary disk surrounding a young star begins to fragment and form gas giant planets with stable orbits. (Mayer, Quinn, Wadsley, Stadel).

Box 1.3 Mass and Density

Two of the basic properties of Solar System objects are mass and density. Mass is a measure of the amount of matter in a particle or object. The standard unit of mass in the International System (SI) is the kilogram (kg). This is usually determined by measuring the object's gravitational influence on other objects, for example, natural or artificial satellites.

Once the volume of an object is known, its bulk density can be calculated. In this book, density is usually expressed in grams per cubic centimeter (g/cm^3). As a guide, the density of water is $1.0\,g/cm^3$. Objects which have a density lower than water are able to float (assuming enough water is available!).

If a planet has a high density, it means that it is largely made of dense, rocky or metallic materials. Objects often have low densities because they contain a lot of gases or ices, but few rocky materials. This is why all of the giant planets in the Solar System have low densities, despite their huge size.

The planet with the lowest density ($0.7\,g/cm^3$) is Saturn. The reason that Saturn has such a low density is that it is mainly composed of gas, particularly hydrogen and helium. There is only a small rocky core at its center.

Other objects, including many small satellites and asteroids, have low densities because they are piles of loosely consolidated rubble or highly porous, that is, they contain numerous empty spaces.

The densities of planets are also a reflection of their size and the layering of their interiors. Earth has the highest density of all the planets in the Solar System because it is made of dense, rocky materials. At the surface, crustal rocks have densities between 2.5 and $3.5\,g/cm^3$. However, Earth's average density is much higher ($5.5\,g/cm^3$).

This is partly because the denser elements, such as iron and nickel, have sunk to the center of the planet, while the less dense materials have risen to the surface. Many planets were internally differentiated in this way early in their lives.

The centers of the planets are also more compressed by the weight of the overlying material. In the case of Earth, for example, the normal, uncompressed density of its rocks is about $4.4\,g/cm^3$, but the central core is compressed to greater than normal density by the overlying layers.

More massive planets should experience greater compression at their centers, and hence higher average densities, if they are made of the same rocky and metallic materials as Earth. The opposite should apply to smaller planets. However, the smallest of the rocky planets, Mercury, actually has an average density of $5.4\,g/cm^3$, only slightly lower than Earth's.

Mercury's density rises to a remarkable $5.3\,g/cm^3$ after it has been corrected for the effects of internal compression – much higher than Earth's. The only way to explain this is to assume that the little planet has a huge core of iron and nickel that takes up almost half of its interior (see Chapter 5).

are traveling so quickly that when they collide, they pulverize each other instead of merging.

While the largest protoplanets continue to grow, the remaining rocky planetesimals grind each other into dust. Some of this dust is gathered in by the surviving planets, while much of the remainder is swept out of the Solar System when the Sun evolves into a hydrogen-burning star. (A cloud of micron-sized dust particles still exists in the ecliptic plane of the Solar System. Known as the zodiacal cloud, it is composed of silicate particles that are largely derived from collisions between main belt asteroids.)

One of the problems that has to be solved by Solar System theorists is an explanation for the silicate and metal–rich nature of the terrestrial planets and the dominance of hydrogen and helium in the outer planets. Clearly, the marked difference in composition between the inner and outer planets must be related to the materials that made up different regions of the disk.

The dense, rocky nature of the Earth and its neighbors suggests that they simply formed through the accretion of dust grains in the solar nebula. However, studies of primitive chondritic meteorites show the presence of millimeter-sized droplets (chondrules) that were once liquid.

It seems that, before they amalgamated to form the meteorites, these existed for a brief period as independent spheroids at temperatures above 1 500°C. Some chondrules seem to include other chondrules, indicative of being exposed to high temperatures on more than one occasion (see Chapter 13). The source of the heating is uncertain, although shock waves, solar heating and collisions between planetesimals have been suggested.

Laboratory experiments indicate that these molten globules were cooled very rapidly within ten million years of the collapse of the molecular cloud. The cause of such sudden cooling events remains unclear. What does seem certain is that the chondrules and dust began to stick together and grow in size, creating chunks of chondritic material. Drag from gas in the nebula encouraged the pebble-sized objects to creep inward, all the time gathering in more material.

Once a population of large planetesimals evolved, their destiny was determined largely by chance. A fast, head-on collision caused the objects to break apart. A slow, gentle encounter enabled the participants to merge into an even larger object. In this way, the terrestrial planets grew to more or less their current size over a period of some ten million years.

The huge amounts of kinetic energy dumped in the planets by frequent, massive impacts caused partial or total melting and the creation of magma oceans. This led to internal differentiation, with the denser elements, such as iron, sinking to the core and the lighter ones rising to the surface to create silicate crusts. Early atmospheres were generated by outgassing of volatile molecules such as water, methane, ammonia, hydrogen, nitrogen and carbon

Box 1.4 Key Steps in the formation of rocky planets (after Kenyon and Bromley)

1. A molecular cloud made up of gas and dust begins to collapse.
2. A protostar begins to form at the core of the collapsing nebula.
3. A disk-shaped nebula of orbiting dust and gas develops in the protostar's equatorial plane.
4. Dust grains in the disk collide and merge.
5. Large (1 mm) dust grains fall into a thin, dusty sheet.
6. Collisions produce planetesimals 1 meters to 1 km across.
7. More collisions between planetesimals produce planetary embryos.
8. Planetary embryos stir up the leftover planetesimals.
9. Planetesimals then collide and fragment.
10. A cascade of collisions reduces fragments to dust.
11. Planets sweep up some of the dust.
12. Radiation and a "wind" of charged particles from the central star remove the remaining gas and dust.

dioxide. A final heavy bombardment, which ended about 3.8 billion years ago, is clearly marked in the crater record of the Moon, and this has been applied to other planets and satellites.

Occasionally a satellite was created as the by-product of a major impact. Such is thought to be the case with Earth and its Moon. Debris from an ancient collision between the young Earth and a Mars-sized planetesimal created a ring of debris that eventually came together to form the Moon. A similar explanation has been put forward for the Pluto-Charon system (see Chapters 3 and 12).

Gas Giants and Ice Giants

In the outer reaches of the solar nebula, temperatures were low enough for ices to form. Indeed, it seems that ice particles were much more abundant than silicate dust particles. This being the case, any planetesimals born in the frigid outer zone would have resembled icy dirt balls, much like the comets we see today. Unfortunately, the main constituents of Jupiter and Saturn are hydrogen and helium, rather than water. Since temperatures in the nebula would have been too warm for these gases to condense, accretion of hydrogen and helium snowflakes cannot have occurred. Another explanation must be found.

There seem to be two possibilities. Studies of gas giant interiors suggest that Jupiter and Saturn may possess rocky cores at least as large as the Earth. It may be, therefore, that the early stages of growth of these planets resembled the accretion taking place in the inner Solar System, with the growth of massive nuclei of ice and dust. Once these became sufficiently large, about 5 to 15 times the mass of Earth, they were able to attract and hold onto even the lightest gases in the surrounding solar nebula. As their mass and gravitational grasp grew, their spheres became ever more bloated.

Alternatively, they could simply have developed as the result of large-scale gravitational instabilities in the solar nebula. Since the disk in the outer reaches contained both dust and condensed ices, there was plenty of raw material for large planets to develop and grow.

Any theory must also account for the fact that Jupiter and Saturn are huge hydrogen-helium planets, whereas Uranus and Neptune are notably smaller and composed mostly of elements that form ices: oxygen, carbon and nitrogen. If the latter pair began as icy nuclei, they must have grown quite slowly in the more rarefied conditions of the presolar nebula beyond about 15 AU. By the time they were massive enough to draw in large amounts of gas, the nebula had dissipated and the supply was cut off.

Migrating Planets

Our picture of the early Solar System is complicated by the likelihood that the giant planets migrated considerable distances before they ended up in their present positions. Such large-scale movement is supported by the discovery of numerous large, extrasolar planets that orbit within a fraction of an astronomical unit of a star.

In the case of our Solar System, this migration can be explained by the exchange of orbital momentum between giant planets and innumerable planetesimals. One current model (the Nice model) envisages a chaotic early Solar System occupied by the major planets out to a distance of about 15 AU (closer than the present orbit of Uranus). Jupiter may have been born a little farther out in the Solar System than it is today, whereas the other giants were closer to the infant Sun than at present. Beyond the planets was a region swarming with leftover planetesimals.

Whereas Jupiter was massive enough to eject large numbers of planetesimals to the outer reaches of the Solar System or out of the system altogether, the three, smaller, giants were unable to do this. Instead, they flung similar numbers of planetesimals toward the Sun and away from it. Whenever Uranus or Neptune decelerated a nearby planetesimal, causing the object to move closer to the Sun, the planet gained a tiny amount of momentum. The resultant acceleration caused it to drift away from the Sun.

Over time, after billions of such gravitational interactions, Jupiter spiraled inward a modest distance, while Saturn drifted outward. When Jupiter reached a distance of 5.3 AU and Saturn arrived at 8.3 AU, the two planets were in a 2:1 orbital resonance, so that one orbit of Saturn lasted precisely two Jupiter orbits. The repeated gravitational pull of Jupiter caused Saturn's orbit to become much more elongated.

Saturn began to create havoc with the orbits of Uranus and Neptune, causing them to become more elliptical. They began to plow through the outer swarm of icy planetesimals, scattering billions of them in all directions. By the time the planets had cleared most of the intruders from their vicinities and the system had settled down again, Saturn had migrated out to about 9.5 AU. The effect on the outer planetary pair was even more extreme. Uranus had moved from about 13 to 19 AU, while Neptune had been catapulted from 15 to 30 AU.

Another consequence of this 500 million year long planetary reshuffle was that the remaining planetesimals, perhaps 0.1% of

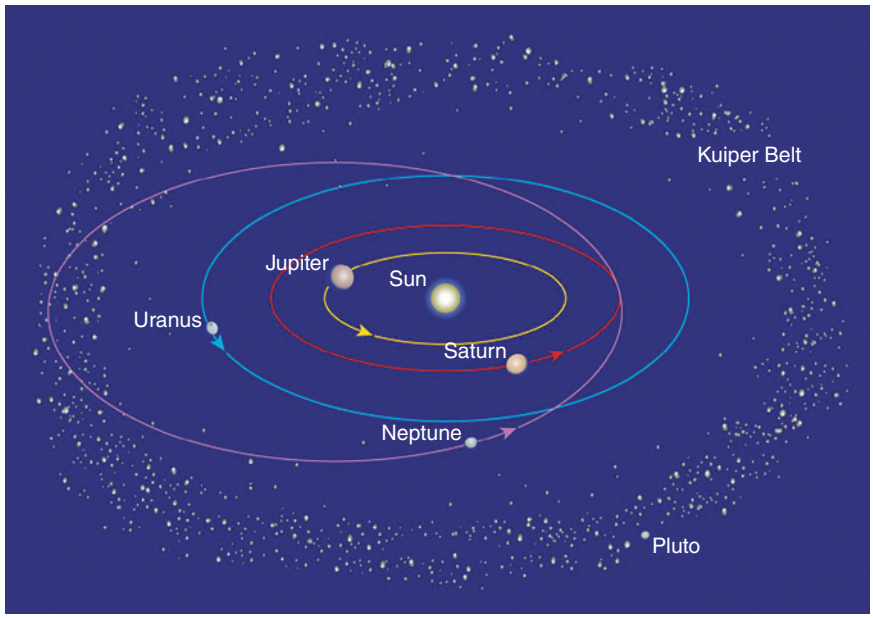


Figure 1.16 According to a recent model, Saturn scattered Neptune outwards beyond Uranus into the Kuiper Belt during the epoch of planet migration. The gravitational interaction of Neptune with icy planetesimals sent billions of Kuiper Belt objects inward, towards the Sun. As a result, Neptune and Uranus migrated outward to their present orbits. Some KBOs, such as Pluto, were locked into orbital resonances with Neptune. (Nature).

the original planet-building population, were relocated beyond 30 AU, where they now reside as Kuiper Belt objects.

Furthermore, the asteroid belt was also strongly perturbed during this burst of migration, adding to the sudden, massive delivery of planetesimals to the inner Solar System. As their pock-marked surfaces show, the Moon and terrestrial planets suffered heavily during this Late Heavy Bombardment, around four billion years ago.

Planetary Satellites

The Solar System contains well over 100 planetary satellites, but, as might be expected from the wide range of sizes and compositions, these seem to have arisen in several different ways.

As mentioned above, Earth's Moon is thought to have been born during a massive, grazing collision between the young Earth and a Mars-sized planetary embryo. The mixture of debris from both objects formed a ring around the scarred Earth, eventually accreting into a large satellite.

Other satellites may also have been created by sizeable impacts early in the Solar System's history. For example, the Pluto-Charon system may have originated during a collision between two large, icy planetesimals over four billion years ago. Simulations show that some of the debris from the collision would be blasted into orbit around the surviving protoplanet, eventually coalescing to form Charon and several smaller satellites.

Most of the major satellites seem to have followed a less traumatic path, gradually accreting from a protoplanetary disk, much like the planets. The most obvious example is the Jovian system, with its family of four Galilean moons. The inner pair, Io and Europa, are smaller but more dense (with a higher proportion of rock) than the outer pair, ice-rich Ganymede and Callisto. All of them orbit Jupiter in the same direction and in more or less the same plane.

These characteristics can be explained if the moons were born from a spherical cloud of dust and gas being drawn inward from the solar nebula by a fledgling planet. As time went by, the cloud flattened into a disk around the protoplanetary core. This disk was hotter and denser near the center, allowing condensation and accretion of the less volatile materials. Further out, the icy volatiles could also condense and accrete to form Ganymede and Callisto.

Although the Saturnian family of satellites is dominated by planet-sized Titan, none of the members are particularly rocky, with many only slightly denser than water. Titan itself is similar in density to Ganymede and Callisto. If the proto-Saturn was surrounded by a collapsing cloud, it seems to have been only about one quarter the mass of Jupiter's. This suggests that the cloud contained less silicate (rocky) material and more ice than its counterpart in the warmer environs of Jupiter.

Certainly, there is a general increase in size and mass moving outward from Saturn toward Titan, with a marked decrease in both properties beyond Titan. This has led theorists to suggest that Titan grew sufficiently quickly to collect much of the solid material in the disk around Saturn, leaving only a modest amount for the medium-sized satellites to accumulate.

A modified version of the accretion scenario has recently been proposed by Robin Canup and William Ward of the Southwest Research Institute. They suggested that a growing satellite's gravity induces spiral waves in a surrounding disk of gas, primarily

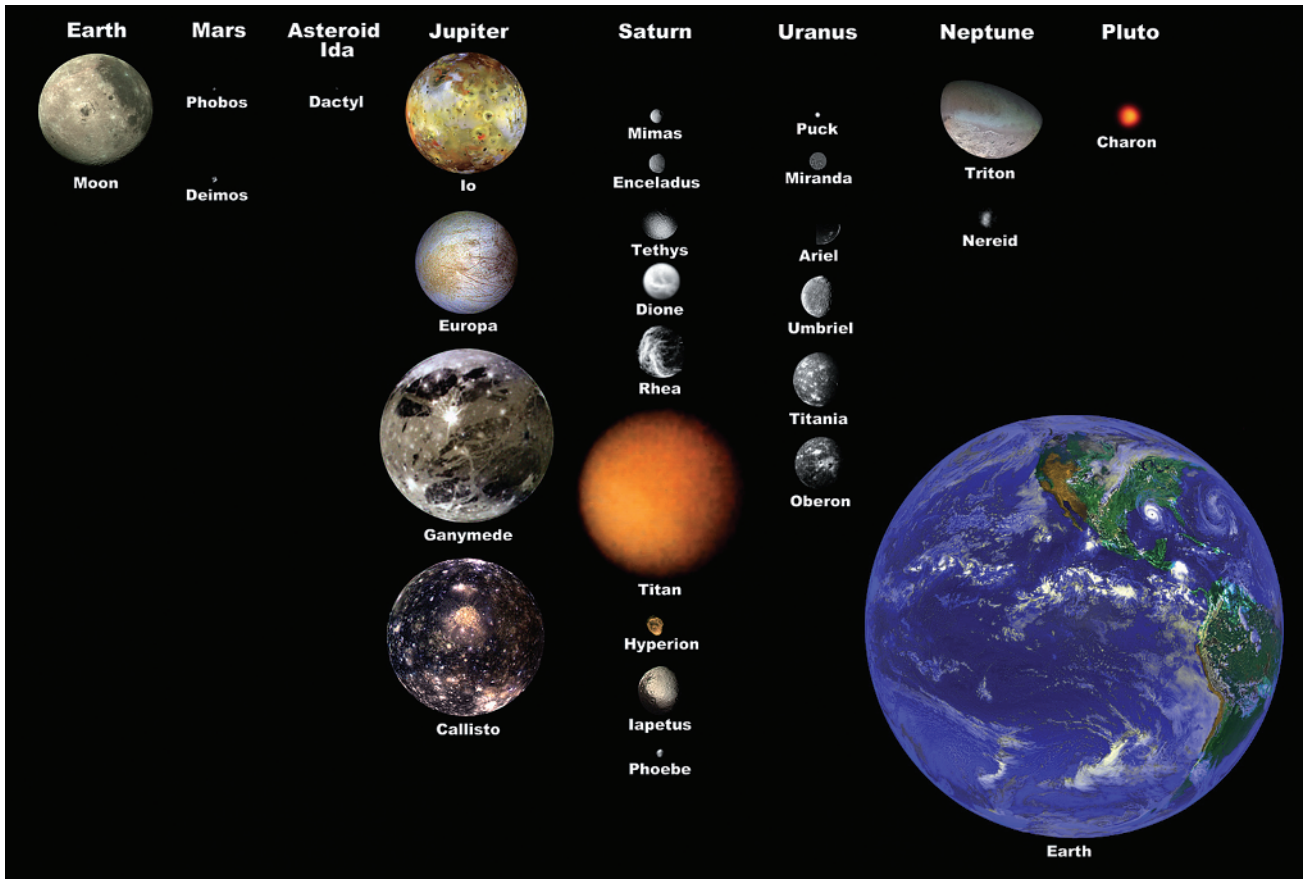


Figure 1.17 The most significant satellites in our Solar System are shown beside the Earth, with their correct relative sizes and colors. Ganymede and Titan are larger than Mercury and eight satellites are larger than Pluto. Earth's Moon is the fifth largest, with a diameter of 3 476 km. Most of them are thought to have formed from a disk of gas and dust in orbit around their home planet. However Triton and many of the smallest satellites (including the moons of Mars) are thought to be captured asteroids or Kuiper Belt objects that formed elsewhere in the Solar System. Earth's Moon and Charon (and possibly the moons of Uranus) are thought to have formed as the result of major impacts. (NASA)

hydrogen. Gravitational interactions between these waves and the satellite cause the moon's orbit to contract. This effect becomes stronger as a satellite grows, so that the bigger a satellite gets, the faster its orbit spirals inward toward the planet. They proposed that the balance between the inflow of material to the satellites and the loss of satellites through collision with the planet implies a maximum size for a satellite of a gas giant.

Numerical simulations and analytical estimates of the growth and loss of satellites showed that multiple generations of satellites were likely, with today's satellites being the last surviving generation to form as the planet's growth ceased and the gas disk dissipated.

The origin of the Uranian (and Neptunian) satellites is open to debate. Models suggest that the ice giants grew more slowly than their larger cousins. By the time they were large enough to gather a disk of material, most of the gas and dust had been dispersed, probably after the young Sun entered its active T Tauri phase. If the regular satellites of Uranus and Neptune could not have formed through large-scale accretion from a circumplanetary disk, how did they come about?

One idea is that the planets were larger and hotter during their accretion phase. As they subsequently cooled and contracted, they left behind a "spinout disk" from which small satellites could grow by accretion.

One complication is the fact that the Uranian satellites orbit in circles close to the planet's equator, even though it spins on its side. Neptune's rotation axis is also tilted quite markedly, aligned at about 30° to its orbital plane, while the orbits of its small satellites are circular and near-equatorial. This suggests that the planets were involved in major impacts early in their histories, and that the satellites were born during or after these collisions.

It may be that impacts with planet-sized objects blasted out clouds of hot material that formed orbiting disks around the ice giants. When the material cooled and condensed, the ice-rock ingredients were available for medium-sized satellites to form.

The major exception is Triton, the largest satellite of Neptune. One clue to its origin is that most of its bulk properties are very similar to those of Pluto, one of the largest known members of the Kuiper Belt. Furthermore, unlike the other Neptunian moons, it follows a retrograde path which is quite steeply inclined to the

planet's equator. This unusual orbit has led to speculation that Triton was a Kuiper Belt object that ventured too close to Neptune and was somehow captured.

The Heliosphere

The motion of superhot plasma inside the Sun generates a powerful magnetic field. The Sun's atmosphere extends into interplanetary space through the motion of the electrically charged particles (mainly electrons and protons) of the solar wind, which streams outward in all directions at typical speeds of between 400 and 7 500 km/s. As the particles spiral around the Sun, they carve out an invisible bubble which extends outward for many billions of kilometers. Although electrically neutral atoms, cosmic rays and dust particles from interstellar space can penetrate this bubble, virtually all of the atomic particles in the heliosphere originate in the Sun itself.

The region of space in which the Sun's magnetic field and the wind of charged particles (solar wind) dominate the interstellar medium is known as the heliosphere. The shape of the heliosphere and the distance of the heliopause are determined by three main factors: the motion of the Sun as it plows through the interstellar medium, the density of the interstellar plasma and the pressure exerted on its surroundings by the solar wind.

From theoretical studies and spacecraft observations of planetary magnetospheres and the solar wind, it is known that the density of the solar wind decreases as the inverse square of its distance from the Sun. In other words, solar wind density at 4 AU is only one quarter its density at 2 AU. The strength of the Sun's magnetic field also weakens with distance, although at a slower rate. Eventually, the density and magnetic influence of the solar wind decrease so much that its outward motion is impeded by the sparse plasma of the interstellar medium.

The heliosphere acts like an island in a stream, causing interstellar plasma to be diverted around it. At first it was thought that the heliosphere really was spherical, but the two Voyager spacecraft, which are currently heading out of the Solar System on different paths, observed what seemed to be a "squashed" heliosheath. In this new model, the heliosphere resembled a huge wind sock or tadpole – much like a comet's elongated tail – that is shaped by the motion of the Sun as it plows through a hot, tenuous cloud of interstellar gas and dust. Studies of the motion of nearby stars show that the Sun is traversing the cloud at a velocity of 25.5 km/s. The interstellar medium forces the solar wind to turn back and confines it within the heliosphere.

This picture had to be revised in 2009, when data from the IBEX spacecraft and the Cassini spacecraft in orbit around Saturn showed that the heliosphere is roughly spherical – perhaps like an elongated balloon – after all. Instruments on the spacecraft were used to map the intensity of the energetic neutral atoms ejected from the heliosheath as the solar wind interacts with the interstellar medium. The data showed a belt of hot, high-pressure particles where the interstellar wind flows by the heliosphere. Their distribution indicates that the heliosphere resembles a huge bubble which expands and contracts under the influence of the local interstellar magnetic field as it sweeps past.

The interaction of the heliosphere with the interstellar medium takes place in several stages. For a spacecraft traveling out of the Solar System, the first boundary to be reached is the termination shock. This is a standing shock wave where the supersonic solar wind slows dramatically from more than 100 km/s to about half that speed.

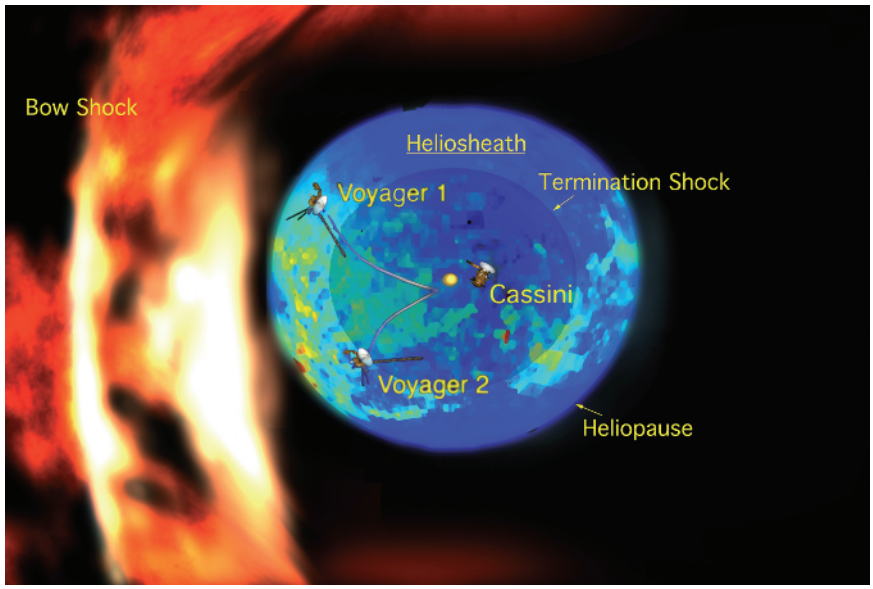


Figure 1.18 For many years the shape of the heliosphere was thought to resemble a comet, with a blunt "head" and an elongated tail. Recent spacecraft observations suggest that it is more like an elongated balloon or bubble. In this artwork the multicolored (blue and green) bubble represents Cassini measurements of the emission of particles known as energetic neutral atoms. The first crossing into interstellar space by Voyager 1 may occur around 2017. (NASA-JPL/JHU-APL).

Beyond the termination shock is a region known as the heliosheath, where particles of the solar wind and interstellar gas mix. In December 2004, Voyager 1 became the first spacecraft to cross into the heliosheath. NASA's two Voyager spacecraft are heading out of the Solar System in different directions. Voyager 1 crossed the termination shock on December 17, 2004, becoming the first spacecraft to enter the heliosheath. Voyager 2 crossed the termination shock on August 30, 2007 – 30 years after it was launched from Florida. The Voyager 2 crossing took place almost 1.6 billion km closer to the Sun than Voyager 1's, confirming that the outer boundary of the Solar System is curved.

Observations by the Magnetospheric Imaging Instrument (MIMI) on board Cassini show that the heliosheath is about 40 to 50 AU (6 billion to 7.5 billion km) thick. Further out is the heliopause, the boundary between the interstellar medium and the heliosphere. Beyond the heliopause, the interstellar ions flow around the heliosphere, modifying its size and shape.

Still further out, there is probably a bow shock, another shock surface where the supersonic flow of the interstellar medium is suddenly slowed as it approaches the heliosphere. All of these boundaries are thought to be moving back and forth at speeds of up to 100 km/s as the heliosphere is squeezed and released due to gusts in the solar wind and variations in the interstellar magnetic field.

The Future

The Solar System is continually evolving and changing. The collision of comet Shoemaker-Levy 9 with Jupiter in July 1994 illustrated that impacts and planetary evolution are continuing today. More significantly, the Sun is also evolving, as nuclear fusion continues to create helium from hydrogen in its core.

Since its birth, 4.54 billion years ago, the Sun has grown 30% brighter and this change will continue. Over the next 1.2 billion years, its surface temperature will increase by about 150°C and its luminosity will increase by another 10%. By this time, the oceans will have boiled away. Over the next 2 billion years, even the water vapor is lost, turning Earth into an arid planet comparable to Venus today.

Models suggest that, about 7 billion years into the future, the Sun will swell into a red giant with a diameter perhaps 200 times larger than today's value – large enough to reach almost to Earth's present orbit. However, an increase in the solar wind will cause up to 25% of the Sun's mass to be blown away. This drop in mass will cause the orbits of the planets to expand outwards, so that Venus may recede to Earth's current orbit, while Earth may lie near the present orbit of Mars. However, this outward retreat will probably be partially balanced by solar tidal drag, which will cause our planet to spiral slowly inward. The Earth's fate will hang in the balance.

Further out, Mars will briefly become warm enough to melt its icy volatiles, leading to a temporary spell of warmth with a dense atmosphere. However, the planet's gravity is not strong enough to maintain the situation for very long. Jupiter's ice-rich Galilean moons will also develop thick atmospheres of water vapor, but again, these wet greenhouse conditions will only represent a short-lived Eden. On Saturn's giant moon Titan, an ocean of liquid ammonia may survive for several hundred million years, perhaps providing a brief interlude when primitive life may evolve.

With its hydrogen now exhausted, the Sun will shrink and become 100 times less luminous as it switches to helium for its energy source. However, the process of helium to carbon fusion will only prolong its active life for a few hundred million years. As the helium becomes exhausted, the Sun will expand once more into a red giant. Riven by sudden pulsations in size, it may well consume Earth – if it still exists. One hundred million years after the second red giant phase, the Sun will eject its outer layers, forming a beautiful (from the outside!) planetary nebula. All that will be left is a tiny, extremely hot, superdense core known as a white dwarf.

The final layout of the Solar System is hard to predict, but it may be that the scorched remnants of Earth and Mars, along with the outer giants, will continue to orbit the fading dwarf star, largely undisturbed, for hundreds of billions of years.

Meanwhile, our galactic environment will also have changed dramatically. About five billion years from now, the Andromeda galaxy and our Milky Way will collide, combining to form a single, football-shaped elliptical galaxy. By then, the Sun will be an aging star nearing the red giant phase and the end of its life.

Models suggest that the Solar System likely will reside 100,000 light-years from the center of the new galaxy – four times further than the current distance. Any human descendants observing the future sky will experience a very different view. The band of the Milky Way will be gone, replaced by a huge bulge of billions of stars.

Questions

- What did the word "planet" originally mean? Why were objects given this name?
- How many planets were recognized before the invention of the telescope?
- How many planets are recognized today? What is the current definition of a planet?
- List six characteristics of the present Solar System that any theory of Solar System formation must explain.
- Explain the main features of the nebular hypothesis that is favored by most scientists today.
- Explain three possible origins of planetary satellites. Give likely examples of each type.
- Describe the main features of the heliosphere.