# 1 Early Astronomy

The term **"astronomy"** is derived from the Greek words *astron*, meaning "star," and *nomos*, meaning "law." This reflects the discovery by ancient Greek astronomers that the motions of stars in the sky are not arbitrary but follow fixed laws. In modern times, astronomy is usually defined as the study of objects beyond the Earth's atmosphere, including not only stars but also celestial objects as small as interstellar dust grains and as large as superclusters of galaxies. The field of **cosmology**, which deals with the structure and evolution of the universe as a whole, is also regarded as part of astronomy.

In the late nineteenth century, the term **"astrophysics"** was invented, to describe specifically the field that studies how the properties of celestial objects are related to the underlying laws of physics. Thus, astrophysics could be regarded as both a subfield of physics and as a subfield of astronomy. However, because a knowledge of physics is crucial for any type of astronomical study, the terms "astronomy" and "astrophysics" are often used nearly interchangeably.

It is customary to start learning astronomy from a historical perspective. This is a natural way to learn about the universe; it permits our personal growth in knowledge to echo humanity's growth in knowledge, starting with relatively nearby and familiar objects, and then moving outward. Furthermore, as we will see, some of the most fundamental things we learn about the universe are based on simple, straightforward observations that don't require telescopes or space probes. Let us begin, therefore, by throwing away our telescopes and considering what we can see of the universe with our unaided eyes.

## 1.1 ▪ THE CELESTIAL SPHERE

When you look up at a cloudless night sky, you have little sense of depth. In Color Figure 1, for instance, it is not immediately obvious that the fuzzy streak in the upper part of the picture is a comet a few light-minutes away and that the fuzzy blob in the lower part is a galaxy two million light-years away. You can pick up a few clues about depth with your naked eyes (for instance, the Moon passes in front of stars, so it must

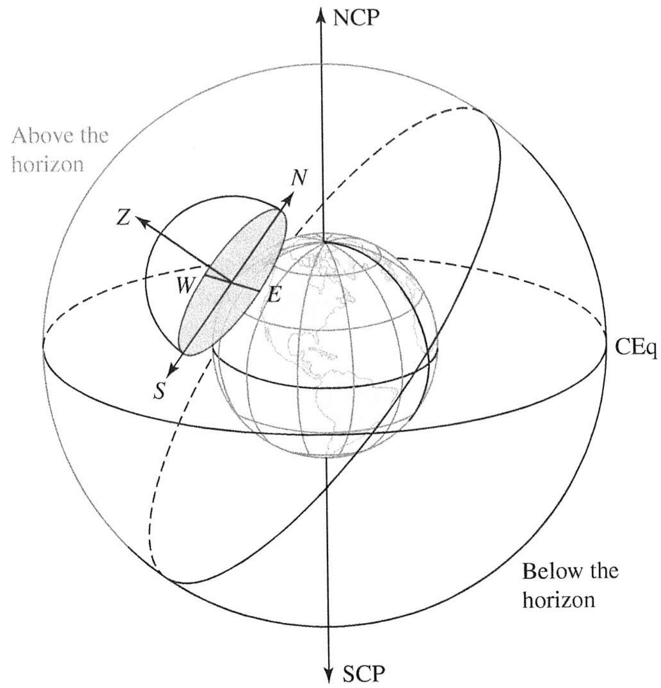


**FIGURE 1.1** The celestial sphere surrounding the Earth. The Earth's north pole, south pole, and equator project onto the north celestial pole (NCP), south celestial pole (SCP), and celestial equator (CEq), respectively. For any observer, the horizon plane is tangent to the observer's location, and the zenith (Z) is directly overhead.

be closer to us than the stars are) but for the most part, determining distances to celestial objects requires sophisticated indirect methods.[1]

Although it is difficult to determine the distance to celestial objects, it is much easier to determine their position projected onto the **celestial sphere**. The celestial sphere is an imaginary spherical surface, centered on the Earth's center, with a radius immensely larger than the Earth's radius. (In Figure 1.1, the spherical Earth is exaggerated in size relative to the outer celestial sphere, for easy visibility.) Given the Earth's inconvenient opacity, an observer on the Earth's surface can see the sky only above the **horizon**, defined as a plane tangent to the idealized, perfectly spherical Earth at the observer's location (that is, it touches the Earth at the observer's feet and at no other place). The horizon is always defined locally, meaning that it moves with the observer. The horizon intersects the celestial sphere in a great circle called the **horizon circle**.[2] The horizon circle divides the celestial sphere into two hemispheres; only the hemisphere above the

[1] Some of these distance-measuring techniques will be discussed in Chapter 13.

[2] A "great circle" is a circle on the surface of a sphere whose center coincides with the sphere's center.

horizon is visible to the observer. The point directly above the observer's head, in the middle of the visible hemisphere of the celestial sphere, is called the **zenith** (point Z in Figure 1.1). The point directly below the observer's feet, opposite the zenith, is the **nadir**.

Since the celestial sphere is indeterminately large, distances between points on the celestial sphere are measured in angular units, as seen by an Earthly observer, rather than in physical units such as kilometers. Astronomers most frequently measure angles in degrees, arcminutes, and arcseconds, with 360 degrees ($360^\circ$) in a circle, 60 arcminutes ($60'$) in a degree, and 60 arcseconds ($60''$) in an arcminute. When they measure angles smaller than an arcsecond, they revert to the decimal system and use milliarcseconds and microarcseconds.

When the Sun is above the horizon, it appears as a bright disk on the celestial sphere, 30 arcminutes across. The Moon, coincidentally, is also roughly 30 arcminutes in diameter as seen from Earth, but appears to change in shape as it waxes and wanes from new Moon to full and back again. When the Sun is below your horizon, you can see as many as 3000 stars with your unaided eyes.[3] The stars in the night sky appear as tiny lights, blurred by our imperfect human vision into blobs about an arcminute across. Starting in prehistoric times, astronomers have identified apparent groupings of stars called **constellations**. The stars in a constellation are not necessarily physically related, since they may be at very different distances from the Earth.

## 1.2 ▪ COORDINATE SYSTEMS ON A SPHERE

If we want to describe the approximate position of a star on the celestial sphere, we can say what constellation it lies within. However, since there are only 88 constellations on the entire celestial sphere, some of them quite large, merely knowing the constellation doesn't give a very precise location. For a more precise description of positions on the celestial sphere, we need to set up a coordinate system. On the two-dimensional celestial sphere, two coordinates will be needed to describe any position. Geographers have already shown us how to set up a coordinate system on a sphere; the system of **latitude** and **longitude** provides a coordinate system on the surface of the (approximately) spherical Earth.

On the Earth, the north and south poles represent the points where the Earth's rotation axis passes through the Earth's surface. The **equator** is the great circle midway between the north and south pole, dividing the Earth's surface into a northern hemisphere and a southern hemisphere. The latitude of a point on the Earth's surface is its angular distance from the equator, measured along a great circle perpendicular to the Earth's equator (Figure 1.2). Latitude is measured in degrees, arcminutes, and arcseconds, as is longitude. Thus, the use of latitude and longitude doesn't require knowing the size of

[3] This number assumes that you are in a dark location, far from the bright lights of the big city. In a populated area, you'll be lucky to see a few hundred stars.

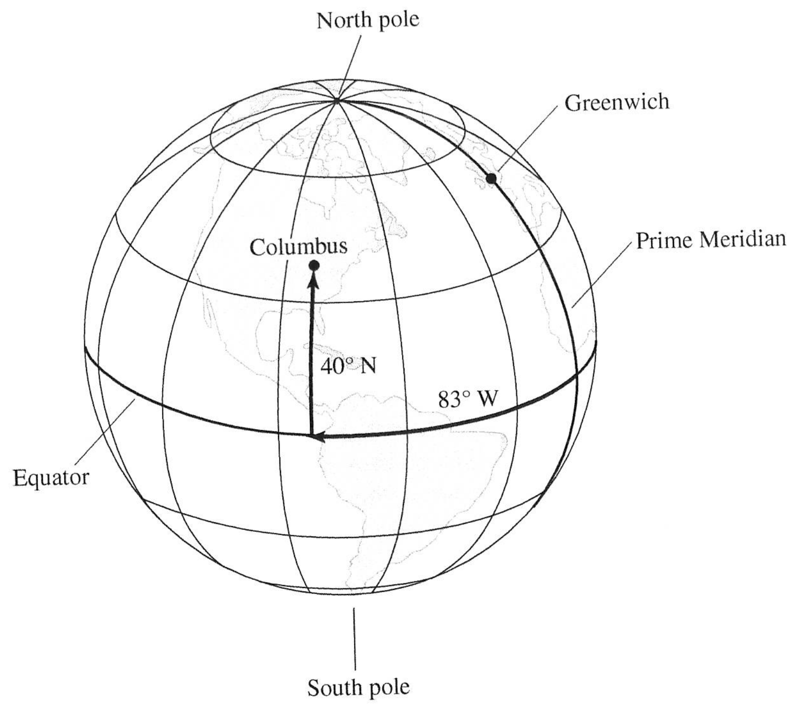


**FIGURE 1.2** Latitude and longitude of a point on the Earth's surface.

the Earth in kilometers or any other unit of length.[4] In the example shown in Figure 1.2, the city of Columbus, Ohio, has a latitude of 40° N; that is, it's located 40° north of the equator.

Latitude alone doesn't uniquely specify a point on the Earth's surface. If you invited a friend to lunch at 40° N, he wouldn't know whether to expect hamburgers in Columbus, Peking duck in Beijing, or shish kebab in Ankara. The required second coordinate on the Earth's surface is the longitude. For each point on the Earth's surface, half a great circle can be drawn starting from the north pole, running through the point in question, and ending at the south pole. This half-circle, which intersects the equator at right angles, is called the point's **meridian of longitude**, or just "meridian" for short. The longitude of the point is the angle between the point's meridian and some other reference meridian. By international agreement, the reference meridian for the Earth, called the **Prime Meridian**, is the meridian that runs through the Royal Observatory at

[4] The use of latitude and longitude was successfully pioneered by the Greek astronomer Ptolemy in the second century AD, despite the fact that Ptolemy severely underestimated the size of the Earth. (Ptolemy's underestimate helped to encourage Christopher Columbus in his crazy plan to sail nonstop from the Canary Islands to Japan.)

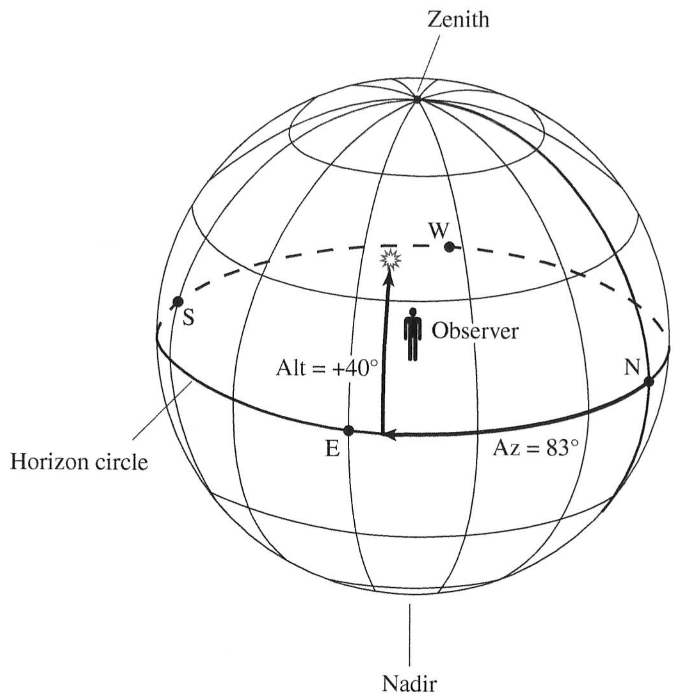


**FIGURE 1.3** Altitude (Alt) and azimuth (Az) of a point on the celestial sphere, as seen by an observer on Earth.

Greenwich, England.[5] In Figure 1.2, the city of Columbus has a longitude of 83° W; that is, the meridian of Columbus is 83° west of the Prime Meridian.

The latitude–longitude coordinate system can be applied to other planets (and to spherical satellites as well). The rotation axis of the planet defines the poles and equator; the Prime Meridian is generally chosen to go through a readily identifiable landmark. The Martian Prime Meridian, for instance, runs through the center of a particular small crater called Airy-0. A coordinate system using latitude-like and longitude-like coordinates can also be applied to the celestial sphere. We just need to specify a great circle that can play the role of the equator on Earth, and a perpendicular meridian that can play the role of the prime meridian.

One such coordinate system on the celestial sphere is based on an observer's horizon, and hence is called the **horizon coordinate system**. In this system, illustrated in Figure 1.3, the latitude-like coordinate is the **altitude**, defined as the angle of a celestial object above the horizon circle. The zenith (the point directly overhead) is at an altitude of 90°. Points on the horizon circle are at an altitude of 0°. The nadir is at an altitude of

[5] Before the International Meridian Conference of 1884 agreed to adopt the Greenwich meridian as the Prime Meridian, different nations used different reference meridians.

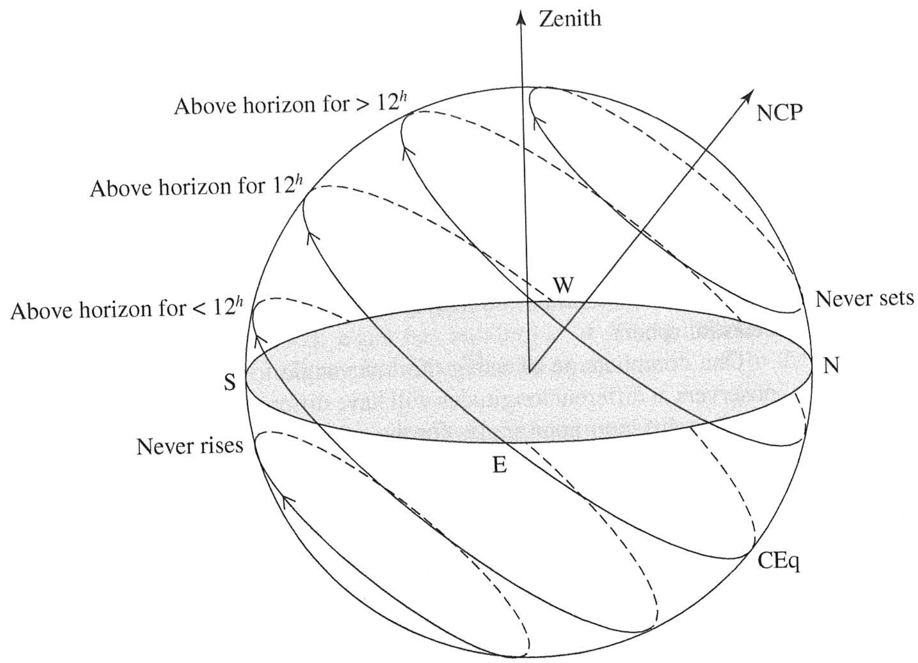


**FIGURE 1.4** Diurnal circles of stars as seen by an observer in the northern hemisphere. Circumpolar stars near the north celestial pole never set; similarly, stars near the south celestial pole never rise. Stars on the celestial equator are above the horizon for 12 hours and below the horizon for 12 hours.

any point on the celestial sphere would work equally well, just as any point on the Earth would work just as well as Greenwich.)

Half a great circle drawn on the celestial sphere, from the north celestial pole, through the vernal equinox, to the south celestial pole, is the celestial equivalent of the Prime Meridian on Earth (Figure 1.5). The longitude-like coordinate measured *eastward* from this "Prime Meridian" is called the **right ascension** ($\alpha$). The right ascension and declination of a star change slowly with time (just as the latitude and longitude of a city on Earth may change slowly thanks to plate tectonics), but they can be treated as constant over the course of a single night, unlike the inexorably changing hour angle. The right ascension of a celestial object, like its hour angle, is characteristically measured in hours, minutes, and seconds. The coordinate system using right ascension and declination is called the **equatorial coordinate system** and is widely used in astronomy; catalogs of stars, for instance, generally give their positions in terms of right ascension and declination. For the example shown in Figure 1.5, the star in question is at a right ascension $\alpha = 277° = 18^h28^m$ and a declination $\delta = +40°$. This is within the constellation Lyra, not far from the bright star Vega.

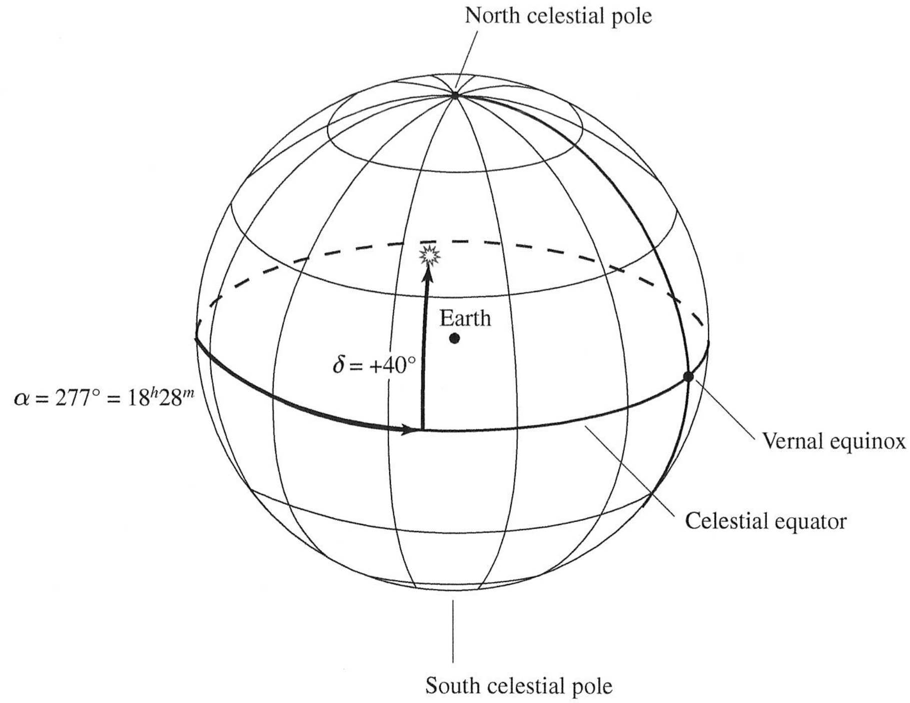


**FIGURE 1.5** The right ascension ($\alpha$) and declination ($\delta$) of a point on the celestial sphere.

## 1.3 ▪ CELESTIAL MOTIONS

As mentioned above, and illustrated in Figure 1.4, an observer on the rotating Earth sees stars move in diurnal circles, just as if the Earth were stationary and the stars were glued to a rigid, rotating celestial sphere. The horizon plane of an observer bisects the celestial sphere, and thus also bisects the celestial equator (labeled "CEq" in Figure 1.4). Thus, stars on the celestial equator are above the horizon for 12 hours a day and below the horizon for 12 hours a day. The diurnal circles of stars not on the celestial equator are not bisected by the horizon (except in the special case when the observer is on the equator, when all diurnal circles are bisected). Consider an observer somewhere in the Earth's northern hemisphere, as shown in Figure 1.4.[10] For stars north of the celestial equator, more than half of their diurnal circles are above the horizon, so they spend more time above the horizon than below. For an observer at latitude $\ell$, all stars within an angular distance $\ell$ of the north celestial pole (that is, with declination $\delta > 90° - \ell$)

[10] In our examples, we will practice blatant northern hemisphere chauvinism, rationalized by the fact that ~90% of the human population lives in the northern hemisphere. Description of apparent motions for a southern hemisphere observer is left as an exercise for the reader.

**FIGURE 1.6** Star trails over Mauna Kea, Hawaii, showing circumpolar stars around the north celestial pole.

will have diurnal circles that don't intersect the horizon plane at all. These stars, called **circumpolar stars**, never fall below the observer's horizon but can be seen to move in counterclockwise circles about the north celestial pole.

Figure 1.6 shows a long exposure of the night sky over Mauna Kea, Hawaii, at a latitude $\ell = 20°$; the star trails cover about 1/12 of a full circle, indicating the photographic exposure was $\sim 2$ hours long. By contrast with circumpolar stars, stars within an angular distance $\ell$ of the *south* celestial pole never rise above the horizon; again, the horizon plane never intersects their diurnal circles. For stars south of the celestial equator but farther than $\ell$ from the south celestial pole, less than half of their diurnal circles are above the horizon; these stars spend less than 12 hours per day above the northern observer's horizon, rising in the southeast and soon setting in the southwest.

As well as the stars, the Sun, Moon, and planets are seen to move in diurnal circles. However, if the Sun, Moon, and planets are observed for times much longer than a single night, additional motions are also seen. The most important motions are the following:

- The relative positions of **stars** can be approximated as constant, over human time scales. Although stars are in motion relative to each other and to the Sun, on time scales shorter than decades the motion cannot be detected without a telescope.
- The **Sun** moves eastward relative to the stars by about 1° per day. This is because the Earth is orbiting the Sun, and we see the Sun in projection against different background stars as we orbit around it. Because of the relative motion of the Sun and stars, the stars rise 4 minutes earlier each day relative to the Sun.

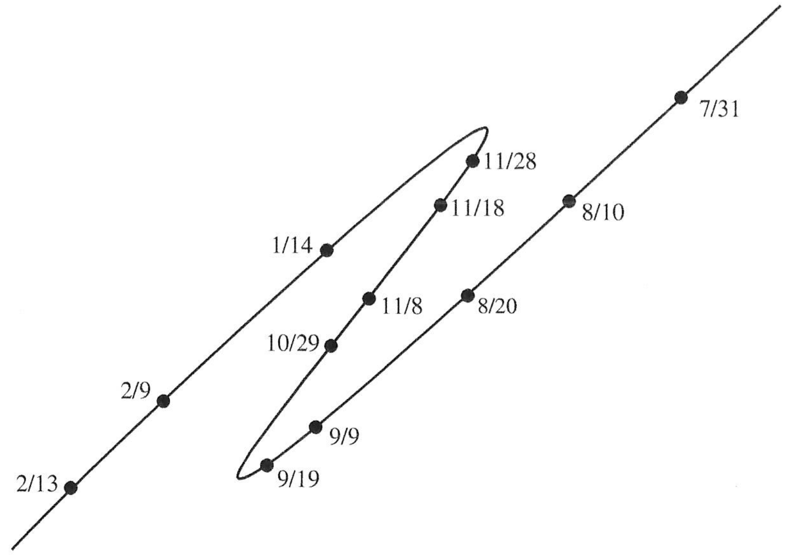


**FIGURE 1.7** The apparent motion of Mars relative to the stars during late 2005 and early 2006. Mars was in retrograde motion from 2005 October 1 to 2005 December 9.

- The **Moon** also moves eastward relative to the stars, by about 13° per day. This is because the Moon orbits around the Earth in an eastward direction, taking 27.3 days for a complete orbit. The Moon's motion around the sky ($360°/27.3$ days $\approx$ $13° \text{ day}^{-1}$) is slow compared to the Earth's eastward rotation ($360° \text{ day}^{-1}$), so we still see the Moon rise in the east and set in the west, just like the Sun. Relative to the Sun, the Moon moves eastward by about 12° per day, so it takes $360°/12° \text{ day}^{-1} \approx 30$ days for the Moon to "lap" the Sun. Because of the relative motion of Sun and Moon, the Moon rises about 50 minutes later each day.
- The **planets** known prior to the invention of the telescope were Mercury, Venus, Mars, Jupiter, and Saturn (in addition to the Earth, of course). Without a telescope, the planets look like unresolved stars. Early astronomers distinguished them from stars by the fact that planets move relative to the stars.[11] Ordinarily, planets move slowly eastward relative to the stars. On occasion, however, they reverse their motion and move westward relative to the stars for a short period. This reversed motion is called **retrograde motion**. Figure 1.7, illustrates, an example of retrograde motion for the planet Mars.

The great circle along which the Sun moves on the celestial sphere is called the **ecliptic**. The ecliptic represents the plane of the Earth's orbit around the Sun, projected onto the

[11] The word "planet" comes from the Greek word meaning "wanderer." The word "plankton" derives from the same root; plankton are tiny aquatic creatures condemned to wander where the ocean currents take them.

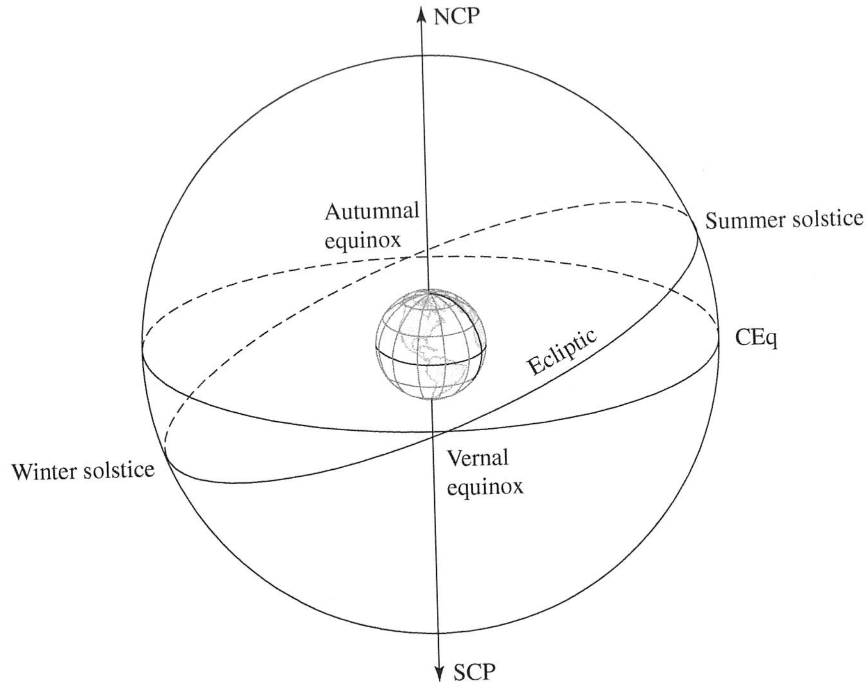


**FIGURE 1.8** The relative positions of the ecliptic and the celestial equator on the celestial sphere. The equinoxes and solstices are indicated.

celestial sphere. The ecliptic, as shown in Figure 1.8, is inclined by 23.5° relative to the celestial equator. The tilt of 23.5° between the ecliptic and celestial equator is called the **obliquity of the ecliptic**. The obliquity is nonzero because the Earth's rotation axis is not exactly perpendicular to the orbit of the Earth around the Sun; instead, the axis is tilted by 23.5° from the perpendicular.

Since the ecliptic and celestial equator are two different great circles on a sphere, they intersect at two points, 180° apart. The two points of intersection are called the **equinoxes**. The point where the Sun moves from the northern celestial hemisphere to the southern is called the **autumnal equinox**; the Sun is at the autumnal equinox around September 21. The point where the Sun moves from the southern celestial hemisphere to the northern is called the **vernal equinox**; the Sun is at the vernal equinox around March 21. (Recall from Section 1.2 that the vernal equinox was chosen as the origin for the measurement of right ascension).

The point on the ecliptic that is farthest north of the celestial equator (it has declination $\delta = +23.5°$) is called the **summer solstice**; the Sun is at the summer solstice around June 21. The point on the ecliptic that is farthest south of the celestial equator ($\delta = -23.5°$) is called the **winter solstice**; the Sun is at the winter solstice around December 21. Astronomers usually use the terms "equinox" and "solstice" to refer to points on the celestial sphere; however, the terms can also refer to the *time* at which the Sun reaches those points.

The Sun's diurnal path varies during the year because its declination changes as it moves along the ecliptic. The time per day that the Sun is above the horizon depends on where it is relative to the celestial equator. At the equinoxes, the Sun is exactly on the celestial equator, and thus spends 12 hours above the horizon and 12 hours below the horizon.[12] When the Sun is north of the celestial equator, it is above the horizon for more than 12 hours for a northern hemisphere observer. When it's south of the celestial equator, it is above the horizon for less than 12 hours for a northern hemisphere observer. In the northern hemisphere, the shortest night of the year occurs when the Sun is at the summer solstice, its point farthest north of the celestial equator. Similarly, the longest night of the year in the northern hemisphere occurs when the Sun is at the winter solstice.[13]

As mentioned on page 10, stars with a declination $\delta > 90° - \ell$ are circumpolar stars for an observer at latitude $\ell$ north of the equator. This implies that a star in the northern celestial hemisphere, with declination $\delta > 0°$, will be a circumpolar star for all observers with latitude $\ell > 90° - \delta$. When the Sun is at the summer solstice, it has a declination $\delta = +23.5°$, and hence is a circumpolar star for observers north of latitude 66.5° N. Within this region, bounded by the **Arctic Circle**, observers experience the phenomenon of the midnight Sun around June 21; the Sun never sets but makes a complete circle in azimuth over 24 hours. At the same time, observers within the **Antarctic Circle**, at latitude 66.5° S, never see the Sun rise over the horizon during the course of 24 hours (see Figure 1.9). At the time of the winter solstice, around December 21, the situation is reversed; observers within the Arctic Circle have 24 hours of darkness while observers within the Antarctic Circle have 24 hours of sunlight.

Globes of the Earth usually have the Arctic and Antarctic Circles drawn on them (see Figure 1.9). They also display the Tropic of Cancer at 23.5° N and the Tropic of Capricorn at 23.5° S. At a latitude $\ell$ north of the equator, the zenith has a declination $+\ell$; thus, the Tropic of Cancer represents the latitude at which the Sun passes directly overhead when it's at the summer solstice. At a latitude $\ell$ *south* of the equator, the zenith has a declination $-\ell$; thus, the Tropic of Capricorn represents the latitude at which the Sun passes directly overhead when it's at the winter solstice. The region on Earth between the Tropic of Cancer and the Tropic of Capricorn is known as "the tropics."[14]

The Sun's annual motion along the ecliptic carries it through a group of constellations that comprise the **zodiac**. The 12 traditional members of the zodiac are Pisces (within which the vernal equinox is located), Aries, Taurus (where the summer solstice is located), Gemini, Cancer, Leo, Virgo (where the autumnal equinox is located), Libra,

[12] This equality accounts for the name "equinox," which comes from the Latin *equus* (equal) + *nox* (night). In other words, it's where the Sun is located when night is equal in length to day.

[13] The term "solstice" comes from the Latin *sol* (Sun) + *sistere* (to stand still). The solstices are the points where the Sun's declination reaches an extremum. Thus, although the Sun doesn't literally stand still relative to the background stars (its right ascension is continuously increasing), its declination is momentarily constant at a solstice.

[14] The words "tropic" and "tropical" derive from the Greek word *trope* (meaning "a turning"—as when the Sun, which has been moving away from the celestial equator, turns around and moves back toward the celestial equator).

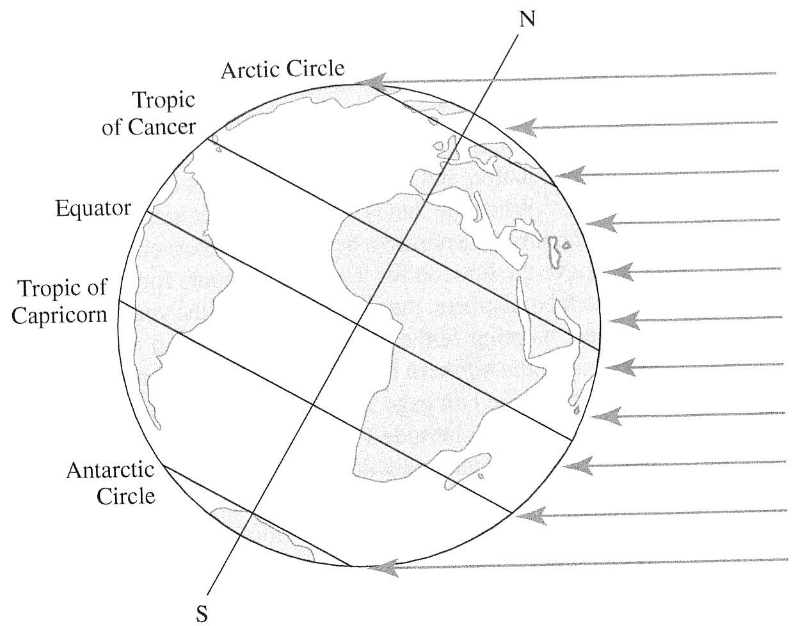


**FIGURE 1.9** Sun's rays striking the Earth, around the time of the summer solstice. The Arctic and Antarctic Circles, as well as the Tropics of Capricorn and Cancer, are indicated.

Scorpius, Sagittarius (where the winter solstice is located), Capricornus, and Aquarius. However, using the constellation boundaries defined by the International Astronomical Union, the Sun also passes through the constellation Ophiuchus (from December 1 to December 18).

The vernal equinox has not always been in Pisces. In the second century BC, the Greek astronomer Hipparchus discovered that the equinoxes and solstices move westward along the ecliptic with respect to the fixed stars of the zodiac. This motion, called the **precession of the equinoxes**, occurs at a rate of $50.3''$ per year, completing a full circuit in 25,800 years. In the time of Hipparchus, the vernal equinox was located in the constellation Aries.[15] We know today, as Hipparchus did not, that the precession of the equinoxes is due to the precession, or "wobble," of the Earth's rotation axis, which moves in a cone of opening angle 47°, with a precession period of 25,800 years (Figure 1.10).

In Section 4.1, we examine the physical causes that make the Earth precess like a dying top. For the moment, however, we will focus on the practical implications of the precession. As the Earth's rotation axis precesses, the north and south celestial poles, which are the projections of that axis onto the celestial sphere, move in a circle of diameter

[15] Thus, the vernal equinox is sometimes referred to, anachronistically, as the "first point of Aries." Similarly, the Tropic of Cancer and Tropic of Capricorn gained their names when the summer and winter solstices were in the constellations Cancer and Capricornus, respectively.

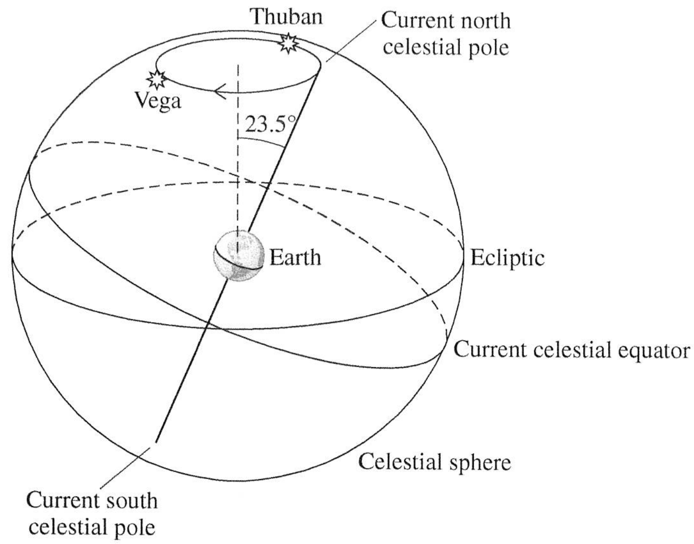


**FIGURE 1.10** Precession of the Earth's rotation axis, with the resulting motion of the north celestial pole on the celestial sphere.

47°, taking 25,800 years for a complete circuit. The north celestial pole (at declination $\delta = 90°$) is currently near the bright star called Polaris, in the constellation Ursa Minor. In the year 2700 BC, the star Thuban, in the constellation Draco, was very close to the north celestial pole.[16] In the year AD 14,000, the bright star Vega, in the constellation Lyra, will be close to the north celestial pole (see Figure 1.10).

As the celestial poles and equator continuously move relative to the background stars, the declination of those stars must also continuously change. Also, since the vernal equinox continuously moves through the zodiac, the right ascension of stars (which is measured relative to the vernal equinox) must continuously change. Since the vernal equinox is moving westward across the celestial sphere, and right ascension is measured eastward from the vernal equinox, the right ascension of a fixed star will increase with time. Due to the time dependence of the coordinates, when a right ascension and declination are given, the epoch at which they are measured must also be specified. The most common standard used today is "equinox 2000.0," indicating right ascension and declination at the beginning of the year 2000. (Some older star charts and catalogs use 1950.0 or 1900.0 as their right ascension and declination standards.)

[16] When William Shakespeare put the words "I am as constant as the northern star" into the mouth of Julius Caesar, he was making an astronomical blunder. In the year 44 BC, when Caesar was assassinated, the closest bright star to the north celestial pole was Kochab, in the constellation Ursa Major; at a distance of 9° from the pole, its diurnal circle would have been blatantly obvious, and calling it "constant" would have been a real stretch.

## 1.4 ▪ BASIC TIMEKEEPING

Astronomy was initially developed largely for its practical applications, such as celestial navigation and timekeeping. Calendars are particularly important for agriculture; planting a crop at the correct time of year is vital. Thus, virtually all agrarian cultures developed astronomy to varying levels of sophistication. Some archeological sites have been shown to have a connection with astronomy. Stonehenge, on Salisbury Plain in the south of England, is a spectacular example of a prehistoric observatory, built in stages during the time span 2500–1700 BC. Various alignments of stones mark key events of the calendar; for instance, the direction in which the Sun rises at the time of the winter solstice and of the summer solstice. Other prehistoric structures throughout the world show similar alignments, giving concrete evidence for humanity's long-standing interest in astronomy.

With the invention of writing, astronomers began leaving systematic records of their observations of the sky. Chinese, Egyptian, and Mayan astronomers all made meticulous records, as did the ancient Babylonians. The Babylonian Empire was the dominant power in southern Mesopotamia (modern Iraq) from the reign of Hammurabi in the eighteenth century BC until it was absorbed into the Persian Empire in the sixth century BC. During that interval, careful observations of the Sun and Moon by Babylonian astronomers enabled an accurate determination of the length of the year and month. The Babylonians used a sexagesimal number system (base 60) rather than a decimal system (base 10); it is thanks to the Babylonians that there are 360 ($6 \times 60$) degrees in a full circle, 60 arcminutes in a degree, and 60 arcseconds in an arcminute.

Astrology, which tracks the positions of planets in the belief that they influence human events, was also a major motivation for the development of astronomy in Babylonia and elsewhere. In fact, until relatively modern times, astronomy was not clearly distinguished from astrology at all. As late as the seventeenth century, for instance, the astronomer Johannes Kepler augmented his inadequate salary by casting horoscopes. "God provides for every animal his means of sustenance—for an astronomer he has provided astrology," Kepler wrote.

All common units of time are ultimately astronomical in origin. The **day** is based on (but is not identical to) the rotation period of the Earth. The **hour** is defined as a fraction of the day. Ancient cultures divided the day into 12 hours of daylight and 12 hours of darkness; thus, the daylight hours were longest near the time of the summer solstice and shortest near the time of the winter solstice. The division of the day into 24 hours of equal length didn't occur until the mechanical clock was invented in the thirteenth century. By the end of the Middle Ages, clocks were accurate enough to allow the subdivision of each hour into 60 **minutes**.[17] The measurement of **seconds**, defined as 1/60 of a minute, wasn't feasible until the invention of pendulum clocks in the seventeenth century.

The **month** and the **year** are based on (but are not identical to) the orbital period of the Moon around the Earth, and the Earth around the Sun, respectively. Even the **week** is tied, albeit loosely, to astronomy. The seven-day week currently in use is the merger

[17] The division of hours into 60 minutes was modeled on the much earlier division of degrees into 60 (arc)minutes. Thus, the 60 tick marks around the edge of a clock face ultimately trace back to the Babylonians.

**TABLE 1.1** Days of the Week

| Latin Name | Italian Name | English Name | Notes |
|---|---|---|---|
| dies Solis | domenica | Sunday | domenica = "Lord" |
| dies Lunae | lunedi | Monday | Moon = Luna |
| dies Martis | martedi | Tuesday | Mars $\approx$ Tiw |
| dies Mercurii | mercoledi | Wednesday | Mercury $\approx$ Woden |
| dies Iovis | giovedi | Thursday | Jupiter = Jove $\approx$ Thor |
| dies Veneris | venerdi | Friday | Venus $\approx$ Frigg |
| dies Saturni | sabato | Saturday | sabato = "Sabbath" |

of two different cycles: the Jewish week, containing six work days plus the Sabbath, and the planetary week, in which each day is presided over by one of the seven wandering objects (or planets) known to ancient astronomers. In the planetary week, which may have originated among Egyptian astrologers, the days of the week are named, in order, after the Sun, the Moon, Mars, Mercury, Jupiter, Venus, and Saturn.[18] The Latin names for the days of the week, shown in Table 1.1, preserve this order. In Romance languages (Italian is shown as an example in the Table), the planetary names are retained for the workweek; however, Saturday is given a name derived from the Sabbath of the Jewish calendar, and Sunday is named the "Lord's Day." In the English names for the days of the week, the links to Saturn, the Sun, and the Moon are obvious in Saturday, Sunday, and Monday. The planetary associations are obscured for the other four days of the week, however, since the names of Roman deities have been replaced with their approximate Teutonic equivalents (see Table 1.1).

## 1.5 ■ SOLAR AND SIDEREAL TIME

In Section 1.4, we noted that the length of the day, as it is most commonly defined, is not exactly equal to the rotation period of the Earth. Let's see why this is true. By convention, we define the "day" to be the interval between successive **upper transits** of a celestial object. Because of the Earth's rotation, a celestial object will cross, or **transit**, the observer's meridian twice a day. The upper transit occurs when the object crosses the zenith meridian, and the lower transit occurs half a day later, when it crosses the nadir meridian.[19] The time between two upper transits of a star is a **sidereal day**;

[18] To modern astronomers, the Sun is classified as a star, and the Moon is classified as a satellite. However, ancient astronomers lumped together the Sun and the Moon with the other "wanderers" they could see in the sky.

[19] For circumpolar objects, both transits are visible above the horizon, so it is particularly important to distinguish between them. The upper transit for a circumpolar object occurs when the object crosses the observer's meridian at a higher altitude traveling westward; the lower transit occurs when the object crosses the meridian at a lower altitude traveling eastward.

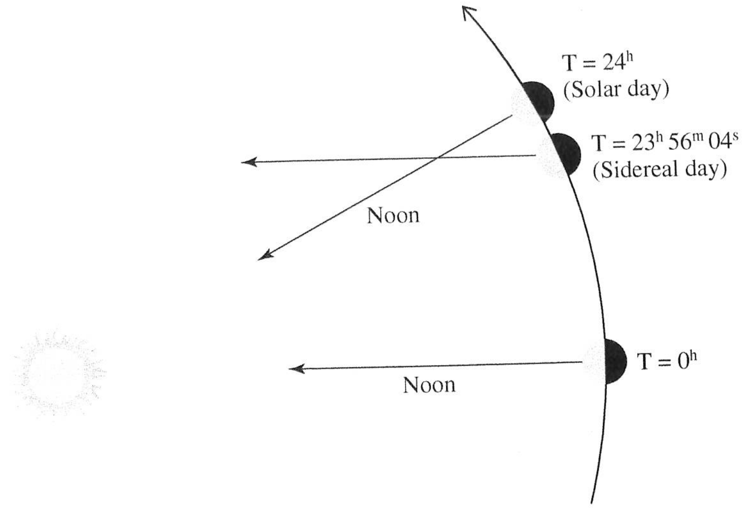


**FIGURE 1.11** The relation between the solar and sidereal day; the solar day is slightly longer than the sidereal day because of the Earth's orbital motion around the Sun.

this represents the Earth's rotation period relative to the distant fixed stars.[20] The time between two upper transits of the *Sun* is a **solar day**, which is slightly longer than the sidereal day, as seen in Figure 1.11. The fundamental measure of time used by humans is solar time, since people find it more convenient to schedule their lives around how the Sun moves in the sky rather than how the inconspicuous nighttime stars move.

The difference in length between the sidereal and solar day is the result of a change in the observer's frame of reference. The sidereal day is the Earth's rotation period measured in the nonrotating frame of reference of the fixed stars, also known as the sidereal frame. The solar day is the Earth's rotation period measured in a reference frame that co-rotates with a line drawn from the Earth to the Sun. To examine the mathematical relation between the sidereal day and the solar day, let $\vec{\boldsymbol{\omega}}_{\mathbf{sid}}$ be the angular velocity of the Earth's rotation in the sidereal frame and let $\vec{\boldsymbol{\omega}}_{\mathbf{E}}$ be the angular velocity of the Earth's orbital motion in the same frame of reference. The difference between these is the angular velocity of the Earth's rotation in a reference frame that co-rotates with the Earth–Sun line; let's call this $\vec{\boldsymbol{\omega}}_{\mathbf{sol}}$. Specifically, we see that

$$\vec{\boldsymbol{\omega}}_{\mathbf{sid}}(t) = \vec{\boldsymbol{\omega}}_{\mathbf{sol}}(t) + \vec{\boldsymbol{\omega}}_{\mathbf{E}}(t). \tag{1.1}$$

If the angular velocity vectors are parallel, this can be rewritten as a scalar equation,

$$\omega_{\mathrm{sid}}(t) = \omega_{\mathrm{sol}}(t) + \omega_{\mathrm{E}}(t). \tag{1.2}$$

[20] The word "sidereal" is derived from the Latin word *sidereus*, meaning "starry."

For the Earth–Sun system, $\omega_{\text{sid}}$ and $\omega_{\text{E}}$ aren't exactly parallel, since they are tilted by 23.5° relative to each other. However, the parallel assumption gives a reasonable first approximation.

If, in addition, $\omega_{\text{sid}}$ and $\omega_{\text{E}}$ are constant, then equation (1.2) can be rewritten in terms of time rather than angular velocity. In that case, $|\omega| = 2\pi/P$, where $P$ is the period of the circular motion in question. Thus, if $P_{\text{sid}}$ is the length of the sidereal day, $P_{\text{sol}}$ is the length of the solar day, and $P_{\text{E}}$ is the Earth's orbital period around the Sun,

$$\begin{aligned} \frac{2\pi}{P_{\text{sid}}} &= \frac{2\pi}{P_{\text{sol}}} + \frac{2\pi}{P_{\text{E}}} \\ \frac{1}{P_{\text{sid}}} &= \frac{1}{P_{\text{sol}}} + \frac{1}{P_{\text{E}}}. \end{aligned} \tag{1.3}$$

If we define the solar day to be $P_{\text{sol}} \equiv 1$ day, then we note that $P_{\text{E}} \approx 365$ days $\gg P_{\text{sol}}$. Thus, we may write

$$\begin{aligned} P_{\text{sid}} &= \left(\frac{1}{P_{\text{sol}}} + \frac{1}{P_{\text{E}}}\right)^{-1} = P_{\text{sol}}\left(1 + \frac{P_{\text{sol}}}{P_{\text{E}}}\right)^{-1} \\ &\approx P_{\text{sol}}\left(1 - \frac{P_{\text{sol}}}{P_{\text{E}}}\right). \end{aligned} \tag{1.4}$$

The difference between the solar day and the sidereal day is then

$$\begin{aligned} P_{\text{sol}} - P_{\text{sid}} &\approx P_{\text{sol}}\left(\frac{P_{\text{sol}}}{P_{\text{E}}}\right) \\ &\approx 1\text{ day}\left(\frac{1}{365}\right)\left(\frac{24\text{ hr}}{1\text{ day}}\right)\left(\frac{60\text{ min}}{1\text{ hr}}\right) \\ &\approx 4\text{ min}. \end{aligned} \tag{1.5}$$

Thus, the length of the sidereal day is $23^{\text{h}}56^{\text{m}}$. This means that, relative to the Sun, the stars rise 4 minutes earlier each day as the Sun moves slowly eastward along the ecliptic.

Although the Sun makes a convenient clock for terrestrial observers, and one that never needs winding, defining time in terms of the solar day has one major problem. The length of the apparent solar day, defined as the time between one upper transit of the Sun and the next, varies over the course of a year. The variations in the apparent solar day are not huge: the shortest apparent solar days, which occur in March and September, are less than a minute shorter than the longest apparent solar days, which occur in June and December. Nevertheless, the differences in the length of the apparent solar day were known to ancient Babylonian astronomers, thanks to their careful observations. From a purely empirical standpoint, astronomers circumvent the problem of the variable length of the apparent solar day by defining two types of time measurement:

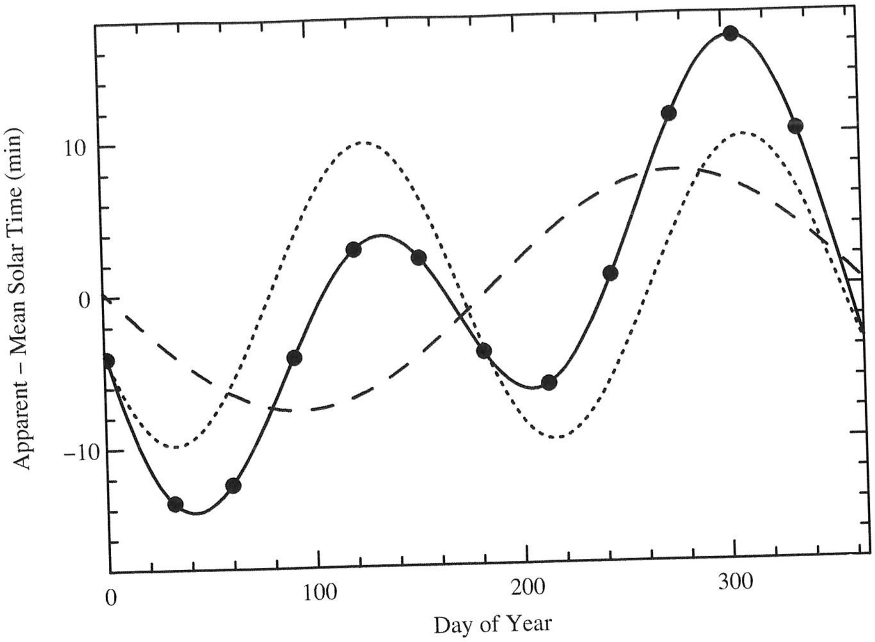


**FIGURE 1.12** The solid line is the empirically determined equation of time; dots represent the first day of each calendar month. The dotted line is the contribution to the equation of time from the obliquity of the ecliptic; the dashed line is the contribution from the Earth's changing orbital speed.

- **Apparent solar time** is measured by the Sun's position relative to the local observer's meridian. Apparent solar time is the time measured by a sundial.[21]
- **Mean solar time** is the time kept by a fictitious "mean Sun" that travels eastward along the celestial equator at a constant rate, completing one circuit in one year. The mean solar day is thus equal to the average length of an apparent solar day. The mean solar day, which is constant over time, is the basis for the time kept by mechanical and electronic clocks.

These two measures of time are related by the **equation of time**. Specifically,

$$\text{Equation of Time} = \text{Apparent Solar Time} - \text{Mean Solar Time}. \tag{1.6}$$

The equation of time, as calculated from observations of the Sun, is shown in Figure 1.12. In mid-February, the accumulation of long apparent solar days causes apparent solar time to fall as much as 14 minutes behind mean solar time. Conversely, during early November, apparent solar time runs more than 16 minutes ahead of mean solar time.

[21] Brief reflection on how a sundial works, combined with the knowledge that mechanical clocks were invented in the northern hemisphere, should lead the reader to an understanding of why clocks run "clockwise."

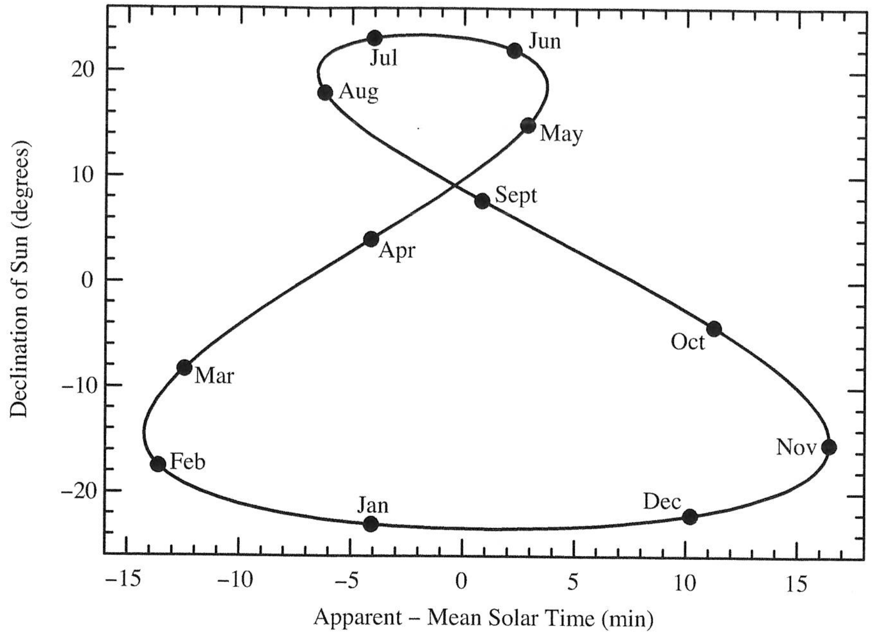


**FIGURE 1.13** The analemma; that is, a plot of the Sun's declination as a function of the equation of time. The dots represent the Sun's position on the first day of each calendar month.

If the equation of time is plotted as a function of the Sun's declination rather than as a function of date, the result is a figure known as the **analemma** (Figure 1.13). The lopsided "figure eight" shape of the analemma is sometimes found printed on globes. Perhaps more striking, if you take a multiple exposure photograph of the Sun, taking an exposure at the same time each day (as measured by a clock) throughout the year, the resulting Sun images trace out the shape of an analemma. Such a photograph, taken from Arizona, is shown in Figure 1.14. Analemma photographs provide graphic evidence that the length of the apparent solar day is variable; if its length were constant, then the analemma would be a straight line segment, not a warped figure eight. The obvious next question is Why does the apparent solar day vary in duration?

The variation in length of the apparent solar day has two causes: the obliquity of the ecliptic (that is, the fact that $\vec{\boldsymbol{\omega}}_{\mathbf{sid}}$ and $\vec{\boldsymbol{\omega}}_{\mathbf{E}}$ are not parallel) and the nonuniform orbital speed of the Earth (that is, the fact that $\omega_E(t)$ varies with time).[22] Even if the Sun moved at a perfectly constant rate along the ecliptic, the obliquity of the ecliptic would create a variable eastward motion of the Sun. The Sun's eastward motion (that is, the rate of increase of its right ascension) is its projected motion onto the celestial equator. The

[22] The angular velocity of the Earth's rotation, $\omega_{\rm sid}$, varies at a much, much slower rate than the angular velocity of the Earth's orbital motion.

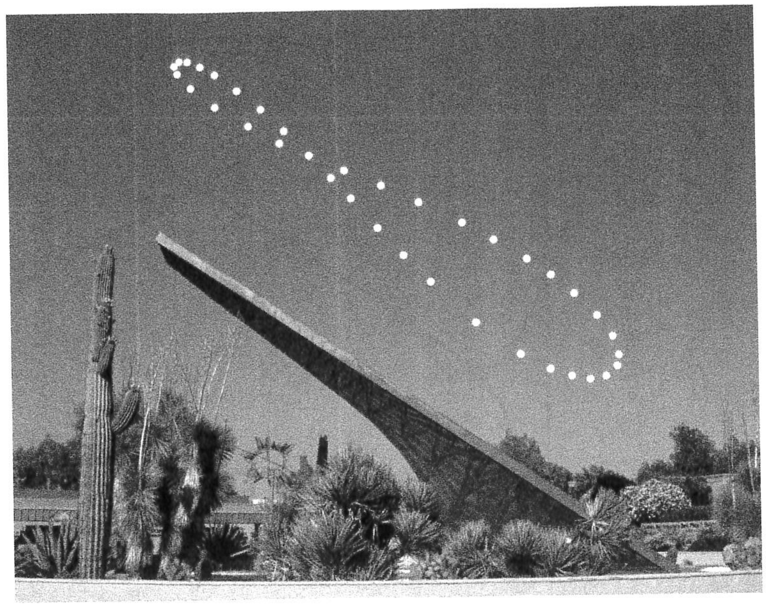

**FIGURE 1.14** Analemma photographed over Carefree, Arizona, from 1990 September to 1991 August, at 8:00 am local standard time.

projected motion is greatest at the solstices, when the Sun's motion is parallel to the equator, and smallest at the equinoxes, when the eastward motion is reduced by a factor $\cos(23.5°) \approx 0.917$. As a result, the apparent solar day tends to be longer at the solstices (late June and late December) than at the equinoxes (late March and late September).

The contribution of the obliquity of the ecliptic to the equation of time is shown as the dashed line in Figure 1.12. The other contribution to the equation of time is due to the fact that the angular speed of the Earth on its orbit is not constant. As discussed in detail in Section 2.5, the Earth's orbit is not perfectly circular but is a mildly eccentric ellipse. The angular speed of the Earth is greatest when the Earth is at its closest approach to the Sun; this occurs near the beginning of January. The motion of the Sun, as seen from Earth, will thus be largest in January and smallest in July, six months later, when the Earth is at its greatest distance from the Sun. The contribution of the changing angular speed to the equation of time is shown as the dotted line in Figure 1.12. When the contributions of the obliquity of the ecliptic and the variable angular speed are added together, they produce the observed equation of time (the solid line in the figure).

Even after switching from apparent to mean solar time, a remaining difficulty is that time is defined locally, not globally. **Local noon**, defined as the instant when the center of the Sun makes an upper transit, is different for observers at different longitudes. For every degree of longitude that you travel westward, local noon occurs 4 minutes later. Prior to the nineteenth century, when it took far longer than 4 minutes to travel one degree in longitude, this was not a problem. However, the advent of high-speed communication (the telegraph) and high-speed transportation (the railways) raised a problem. To prevent railway conductors from having to reset their watch at each station, every railway company adopted a standard time, usually that of the company headquarters or the largest city serviced by the company. This resulted in unpleasant chaos at railroad stations served by more than one railway company.

In 1883, the major railway companies of the United States and Canada simplified matters by adopting **time zones**, within which all clocks would strike noon simultaneously. The adoption of time zones has since spread throughout the world. Each time zone is nominally 15° wide, but adjusted locally along political boundaries; the time within each time zone is called the "civil time" and can vary significantly from the local mean solar time, especially in broader time zones. Since the civil time increases by one hour for each time zone you travel to the east, there must be a boundary drawn from the north pole to the south at which the civil time jumps backward by 24 hours as you travel to the east. Otherwise, the civil time would be multiple-valued as it wound around and around and around the globe. This boundary is called the **International Date Line**. The International Date Line is based on the meridian opposite the Prime Meridian, but jogs back and forth to ensure that the division line doesn't pass through any nations.

Astronomers, and other scientists, frequently want to use a time measure that is independent of the observer's position on Earth. During the ninteenth century, when the meridian through Greenwich was adopted as the Prime Meridian, it was natural to use Greenwich Mean Time (GMT) as the universal standard, where GMT is defined as the mean solar time as measured at the Prime Meridian. Locally, mean solar time is then

$$\text{Mean Solar Time} = \text{GMT} + \ell_{\text{east}}, \tag{1.7}$$

where $\ell_{\text{east}}$ is the east longitude of the observer. Civil time, however, is

$$\text{Local Civil Time} = \text{GMT} + N_{\text{zone}} \times 1\,\text{hr}, \tag{1.8}$$

where $N_{\text{zone}}$ is the integral number of time zones the observer is displaced eastward from the prime meridian.[23]

Until the twentieth century, the rotating Earth provided the ultimate basis for human measurements of time. However, the rotation rate of the Earth is not perfectly constant. The Moon's tidal effect (discussed in more detail in Section 4.2) slows the Earth's rotation rate by approximately $0.0016\,\text{s century}^{-1}$. In addition, the Earth's rotation rate varies seasonally because of changes in atmospheric and oceanic temperature and has irregular changes due to earthquakes, which make tiny changes to the Earth's moment of inertia. The Earth, in other words, is a clock that is "winding down," and is doing so at an irregular rate. In the twentieth century, atomic clocks were devised that measured time more accurately than the rotating Earth. The fundamental measure of time used today is therefore International Atomic Time (abbreviated TAI, from the French Temps Atomique International). In the SI system of units, the **second** is defined as 9,192,631,770 times the period of the radiation emitted by the hyperfine transition of the cesium-133 atom at absolute zero temperature. This definition was chosen so that the second was equal in length to 1/60 of 1/60 of 1/24 of a mean solar day, measured around the year AD 1900.

[23] The actual situation is complicated by the fact that some regions, such as India, Newfoundland, and central Australia, have time zones that are not offset by an integral number of hours.

Because of the slowing of the Earth's rotation, a mean solar day in AD 2000 was not $60 \times 60 \times 24 = 86{,}400$ s. Instead, it was 86,400.0016 s.

We thus have an inherent tension between time as measured by highly accurate atomic clocks and time as measured by the not-quite-as-accurate clock provided by the rotating Earth. To resolve this tension, scientists have adopted a reference time for Earth, called Coordinated Universal Time, or **UTC**.[24] In UTC, seconds are defined in accordance with the SI definition. UTC is synchronized with the (gradually slowing) mean Sun by occasionally interpolating a **leap second** when necessary to keep UTC within 0.9 seconds of the time measured by the mean Sun. Through the year 2008, a total of 34 leap seconds were required to align UTC with mean solar time.

One might ask how a spin-down rate of only 0.0016 s century$^{-1}$ can lead to 34 leap seconds over a period of just over a century. This is because the effect of slowing is cumulative. Think of the Earth as being a clock that slows at a rate $\epsilon = 0.0016$ s century$^{-1}$ = $4.4 \times 10^{-8}$ s day$^{-1}$. If the length of a day is $P_0$, then during the first day we use the Earth as a clock, it loses a time $\epsilon P_0 = 4.4 \times 10^{-8}$ s. During the second day, however, it loses a time $2\epsilon P_0 = 8.8 \times 10^{-8}$ s, and in general, on the $N$th day, it loses $N\epsilon P_0$. The total time lag after $N$ days will be

$$\Delta t = \epsilon P_0 + 2\epsilon P_0 + \ldots + N\epsilon P_0 = \epsilon P_0 \sum_{i=1}^{N} i = \epsilon P_0 \frac{N(N+1)}{2}. \quad (1.9)$$

After a time $t$ equal to many days has passed ($N \gg 1$), the time lag that must be filled in with leap seconds is

$$\Delta t \approx \epsilon P_0 \frac{N^2}{2} \approx \frac{\epsilon P_0}{2}\left(\frac{t}{1\text{ day}}\right)^2. \quad (1.10)$$

Thus, the time lag due to the Earth's spin-down is *quadratic* in $t$, not linear. Since there are 36,525 days in a century, the time lag expected, in seconds, is

$$\Delta t \approx \frac{(4.4 \times 10^{-8}\text{ s day}^{-1})(1\text{ day})}{2}\left(\frac{36{,}525\text{ days}}{1\text{ century}}\right)^2\left(\frac{t}{1\text{ century}}\right)^2$$
$$\approx 30\text{ s}\left(\frac{t}{1\text{ century}}\right)^2. \quad (1.11)$$

This is only an approximate formula, because of the occasional small glitches in the Earth's rotation rate due to earthquakes. However, it gives the correct long-term trend: each century will require a greater number of leap seconds to keep the gradually lengthening mean solar day in synch with atomic clocks.

In addition to solar time, astronomers frequently find it useful to use an alternative time system, **sidereal time**. Because a sidereal day is the time between upper transits of a star other than the Sun, it represents the rotation period of the Earth relative to the

[24] English speakers wanted the abbreviation CUT; French speakers wanted TUC, for *temps universel coordiné*. UTC was chosen as the compromise.

distant fixed stars. A clock measuring sidereal time runs faster than a clock measuring mean solar time, by about 4 minutes per day.

Technically speaking, the **local sidereal time** (LST) is defined as the hour angle of the vernal equinox, which by definition has a right ascension $\alpha = 0$. Thus, when the vernal equinox makes an upper transit, the local sidereal time is $0^h$. Local sidereal time is based on a 24-hour clock, running from $0^h$ to $24^h$. If the vernal equinox is not above the horizon, the local sidereal time can be computed after measuring the hour angle ($H$) of a star with known right ascension ($\alpha$):

$$\text{LST} = H + \alpha. \tag{1.12}$$

In practice, astronomers use this equation to compute the hour angle of a star with known right ascension at a particular local sidereal time.

## 1.6 ▪ CALENDARS

As mentioned in Section 1.4, having an accurate calendar is useful for an agrarian society. For a calendar to remain useful for agricultural purposes, it must remain in phase with the seasons of the year. That is, the Sun should return to the vernal equinox on the same calendar date each year. (In the calendar currently in use, that date happens to be March 21.) The interval of time that elapses between successive passes of the Sun through the vernal equinox is called the **tropical year**.[25] The length of the tropical year is 365.24219 mean solar days. Because of the precession of the equinoxes, the tropical year is slightly different in length from the **sidereal year**, which is the time it takes the Sun to make a complete circle of the ecliptic relative to the fixed background stars. The sidereal year, which is the orbital period of the Earth around the Sun, is 365.25636 days, or about 20 minutes longer than the tropical year.

The fact that the number of mean solar days in a tropical year, 365.24219, is not an integer led to a certain amount of difficulty when ancient cultures set up calendars. During the time of the Roman Republic, for instance, the Roman calendar contained 12 months adding up to only 355 days. It was the job of the board of pontifices (Roman priests) to interpolate an extra month when the calendar fell out of synchronization with the seasons. However, the priests were largely driven by nonastronomical considerations; they added the extra month when politicians friendly to them were in office, effectively extending their elected term, but omitted the month when their enemies were in power. By the time Julius Caesar became effective dictator of Rome, the Roman calendar was badly out of alignment with the seasons. In the year 46 BC, Caesar interpolated not one but three extra months to return the time of the vernal equinox to its traditional date in late March.[26] After consulting with an Alexandrian astronomer named Sosigenes, who

[25] It's called the "tropical" year because it is the time required for the Sun to go from being overhead at the Tropic of Cancer to being overhead at the Tropic of Capricorn and back again.

[26] Caesar called the year 46 BC, with its unusual length of 445 days, the *ultimus annus confusionis*, or "last year of confusion." Humorists in Rome emphasized the alternate meaning of the phrase: "the year of ultimate confusion."

was familiar with the 365-day calendar used by the Egyptians, Julius Caesar proclaimed a new calendar. In this **Julian calendar**, years ordinarily had 365 days; however, every fourth year, an extra day, called a **leap day**, was added. The Julian year thus has 365.25 days, on average; this is a fairly close approximation to the tropical year of 365.24219 days.

The initial small difference between the Julian year and the tropical year accumulated with time, amounting to one day every 128 years. By the sixteenth century, the vernal equinox fell on the date March 11, according to the Julian calendar. This caused problems for the Church, which computed the date of Easter using a formula devised in the fourth century that assumed that the vernal equinox occurred on March 21. Thus, the average date of Easter was gradually drifting later and later, relative to the true date of the equinox. Pope Gregory XIII foresaw, with displeasure, a future in which Easter fell during the summer, then during the fall, and eventually the winter. In the year 1582, therefore, Gregory issued a papal bull reforming the calendar. In October of that year, the calendar skipped 10 days, going straight from October 4 to October 15, and thus returning the date of the vernal equinox to March 21.

In addition, the papal bull announced a new algorithm for computing leap days; years evenly divisible by 4 would contain a leap day *unless* the year number was evenly divisible by 100 and not by 400. This means that the years 1600 and 2000 in the new **Gregorian calendar** were leap years, but that 1700, 1800, and 1900 were not. In 400 Gregorian years, there are 97 leap years and 303 regular years, totaling

$$N_{\rm Greg} = 97 \times 366 + 303 \times 365 = 146{,}097 \text{ days.} \tag{1.13}$$

For comparison, 400 tropical years will contain

$$N_{\rm trop} = 400 \times 365.24219 = 146{,}096.88 \text{ days,} \tag{1.14}$$

a difference that amounts to only 1 day in 3225 years (about 4% as large as the error in the Julian calendar). The accuracy of the Gregorian calendar eventually caused it to be adopted by members of all religions. Today all nations, even those that use other calendars for religious purposes, use the Gregorian calendar for business.

## PROBLEMS

**1.1** The Polynesian inhabitants of the Pacific reportedly held festivals whenever the Sun was at the zenith at local noon. How many times per year was such a festival held? At what time(s) of year was the festival held on Tahiti? At what time(s) of year was it held on Oahu? (Hint: any reputable world atlas will give you the latitude of Tahiti and Oahu. You may also find the information in Figure 1.13 useful.)