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

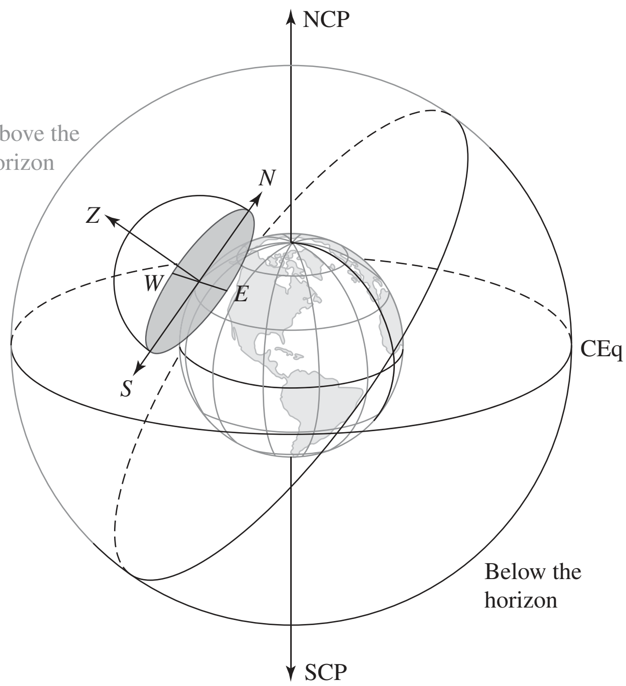

FIGURE 1.1 The celestial sphere surrounding the Earth. The Earth's north pole, south pole, and equator project onto the north celestial pole (NCP), south celestial pole (SCP), and celestial equator (CEq), respectively. For any observer, the horizon plane is tangent to the observer's location, and the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

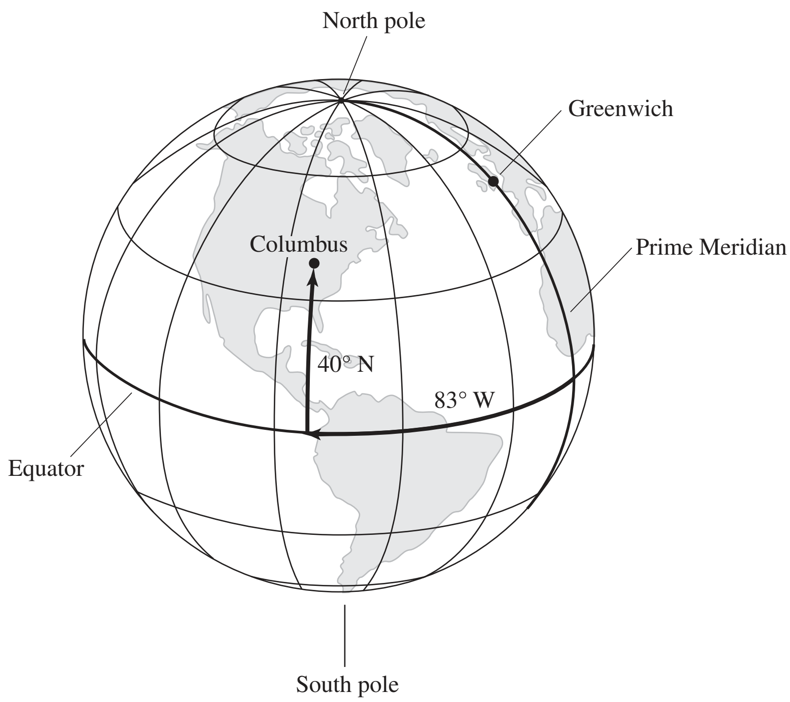

FIGURE 1.2 Latitude and longitude of a point on the Earth's surface.

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.)

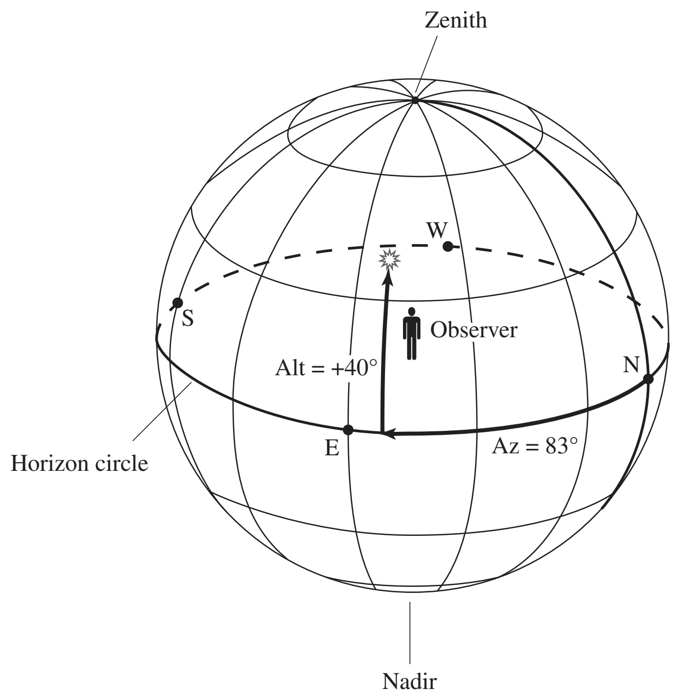

FIGURE 1.3 Altitude (Alt) and azimuth (Az) of a point on the celestial sphere, as seen by an observer on Earth.

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.

$-90°$, but in practice, negative altitudes are seldom used, since they represent objects that are hidden by the Earth. The longitude-like coordinate in the horizon coordinate system is called the **azimuth**.[6]

For any point on the celestial sphere, half a great circle can be drawn from the zenith, through the point in question, to the nadir. The half-circle that runs through the north point on the horizon circle acts as the "prime meridian" in the horizon coordinate system. The azimuth is measured in degrees running from north through east. An object due north of an observer has an azimuth of 0°, an object due east has an azimuth of 90°, and so forth. If you know the altitude and azimuth of any object in your horizon coordinate system, you know where to point your telescope in order to see it. In the example shown in Figure 1.3, a star has an altitude of 40° and an azimuth of 83°; in other words, it's nearly halfway from the horizon to the zenith, off to the east of the observer.

One shortcoming of the horizon coordinate system is that every observer on Earth has a different, unique horizon and hence has a different, unique horizon coordinate system. A star that is near the zenith (altitude $\approx 90°$) for an observer in Buenos Aires will be near the nadir (altitude $\approx -90°$) for an observer in the antipodal city of Shanghai. To describe positions of objects on the celestial sphere, it is useful to have a coordinate system that all astronomers, regardless of location, can agree on, just as geographers all agree to use latitude and longitude to describe positions on the Earth.

To build a coordinate system useful for all Earthlings, we start by projecting the Earth's poles and equator outward onto the celestial sphere. The Earth's rotation axis, which passes through the north and south poles of the Earth, intersects the celestial sphere at the **north celestial pole** (labeled as NCP in Figure 1.1) and the **south celestial pole** (labeled as SCP). The north celestial pole is at the zenith for an observer at the Earth's north pole; more generally, for an observer at a latitude ℓ north of the equator, it will be at an altitude of ℓ and an azimuth of 0°.[7] The projection of the Earth's equator onto the celestial sphere is called the **celestial equator** (labeled as CEq in the figure). The celestial equator passes through the zenith for an observer on the Earth's equator.

On the Earth's surface, a point's latitude is its angular distance north or south of the equator. Similarly, on the celestial sphere, a point's **declination** (δ) is its angular distance north or south of the celestial equator. For points north of the celestial equator, the declination is positive ($0° < \delta \leq 90°$), and for points south of the celestial equator, the declination is negative ($-90° \leq \delta < 0°$).[8] However, the declination alone is insufficient

[6] The words "azimuth," "zenith," and "nadir," like many terms in astronomy, are derived from Arabic. ("Altitude" is from the Latin *altus*, meaning "high.")

[7] Similarly, the south celestial pole is at the zenith for an observer at the Earth's south pole; more generally, for an observer at a latitude ℓ south of the equator, it will be at an altitude ℓ and an azimuth of 180°.

[8] By analogy with the celestial poles and the celestial equator, a more logical term for declination might be "celestial latitude." However, the term "declination" has been in use for over six centuries; in Chaucer's *A Treatise on the Astrolabe* (ca. AD 1391), the poet included what he called "tables of the declinacions of the sonne."

to uniquely locate a point on the celestial sphere, just as latitude alone is insufficient to uniquely locate a point on the Earth. To determine the equivalent of longitude on the celestial sphere, it is necessary to choose a celestial "prime meridian" running from the north celestial pole to the south celestial pole. If we let the observer's zenith act as the celestial "Greenwich," then the **zenith meridian**, defined as the arc running from the north celestial pole through the zenith to the south celestial pole, will act as a celestial "prime meridian."[9] The longitude-like coordinate, measured westward from the zenith meridian, is called the **hour angle** (H). For a given observer at a given time, the declination (angular distance from the celestial equator) and hour angle (angular distance from the zenith meridian) uniquely specify the location of a star, or other object, on the celestial sphere.

One complication of using the hour angle to specify the location of a star is that observers at different longitudes will have different observer's meridians, and hence will measure different hour angles for the same star. If a star is on your zenith meridian, it will be 1° east of the zenith meridian for an observer 1° of longitude west of you. Another complication results from the fact that the Earth is rotating about the axis between its north and south poles, completing one rotation in about 24 hours. Although we know perfectly well, at an intellectual level, that the Earth is rotating from west to east, an observer pinned by gravity to the Earth's surface experiences a strong illusion that the Earth is stationary and the celestial sphere is rotating from east to west. Stars thus appear to follow circular paths called **diurnal circles** that are parallel to the celestial equator; that is, they stay a fixed angular distance from the celestial equator, and their declination remains constant. This situation is illustrated in Figure 1.4. Over the course of 24 hours, the hour angle of a star changes by 360° as it travels in its diurnal circle. Because of the constant rate of change of the hour angle (15° per hour), the hour angle is often measured in units of hours (h), minutes (m), and seconds (s) instead of degrees, arcminutes, and arcseconds, with $1^h = 15°$, $1^m = 15'$, and $1^s = 15''$. A star that is on the zenith meridian right now has hour angle $H = 0^h$; 6 hours from now it will be at $H = +6^h$, off to the observer's west; 12 hours from now it will be at $H = +12^h$, on the nadir meridian. Thus, the hour angle of a star can be thought of as the time that has elapsed since it was last on the zenith meridian.

The hour angle of a star is constantly changing because it is measured relative to an observer's meridian that is tied to the rotating Earth. If we want a longitude-like coordinate that is constant for a given star over the course of 24 hours, we need to measure it relative to a new meridian, one that is tied to the celestial sphere rather than to the Earth. In short, we need a point on the celestial sphere that acts as the astronomical equivalent of Greenwich, England. Instead of selecting one particular star to serve as a "Greenwich," astronomers have chosen a point on the celestial equator termed the "vernal equinox." (In Section 1.3, we give the technical definition of the vernal equinox; but remember,

[9] We can also define a complementary **nadir meridian** running from the north celestial pole through the nadir to the south celestial pole. The zenith meridian and the nadir meridian constitute the two halves of a great circle called the **observer's meridian**.

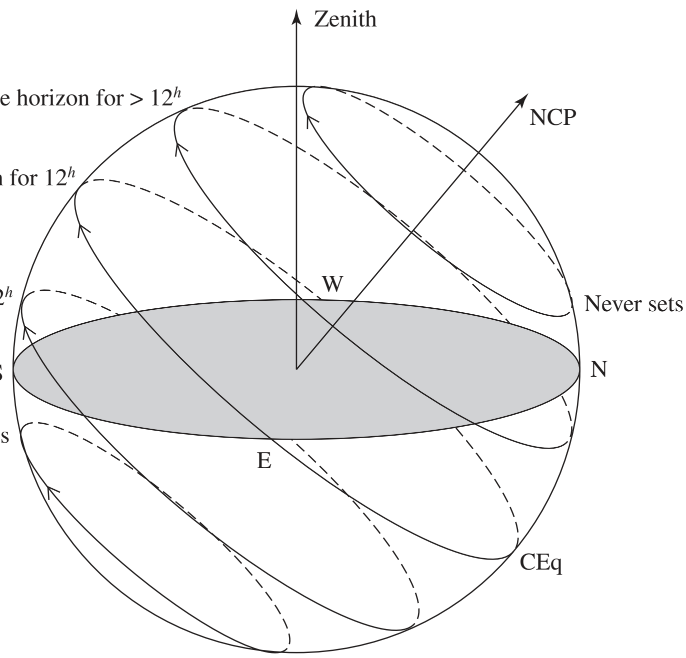

FIGURE 1.4 Diurnal circles of stars as seen by an observer in the northern hemisphere. Circumpolar stars near the north celestial pole never set; similarly, stars near the south celestial pole never rise. Stars on the celestial equator are above the horizon for 12 hours and below the horizon for 12 hours.

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** (α). 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 $\alpha = 277° = 18^h 28^m$ and a declination $\delta = +40°$. This is within the constellation Lyra, not far from the bright star Vega.

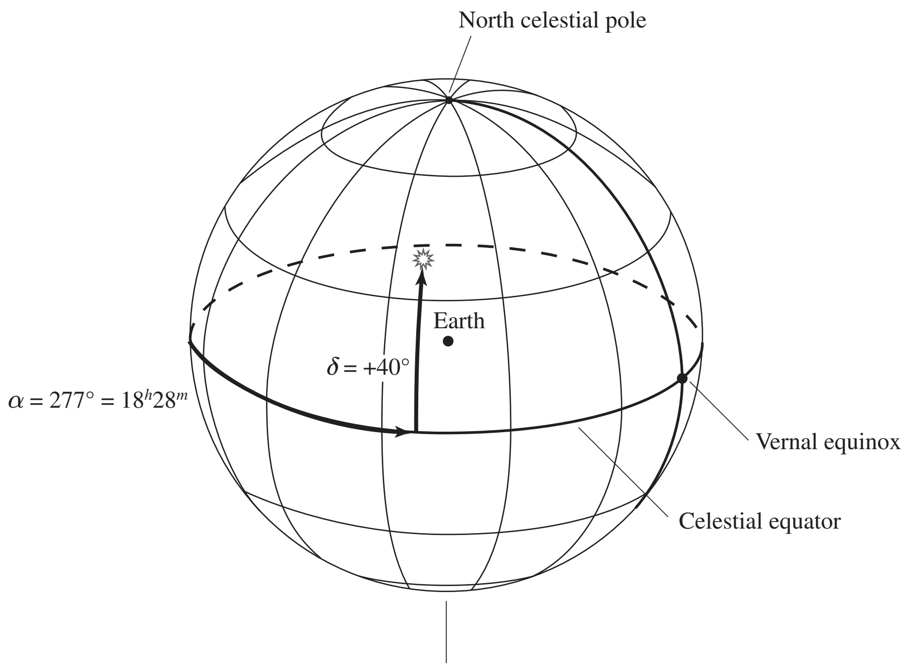

FIGURE 1.5 The right ascension (α) and declination (δ) of a point on the celestial sphere.

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 ℓ, all stars within an angular distance ℓ 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 $\sim$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 ~ 2 hours long. By contrast with circumpolar stars, stars within an angular distance ℓ 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 ℓ 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.