CHAPTER

1 Introduction

Cosmology is the study of the universe, or cosmos, regarded as a whole. Attempting to cover the study of the entire universe in a single volume may seem like a megalomaniac's dream. The universe, after all, is richly textured, with structures on a vast range of scales; planets orbit stars, stars are collected into galaxies, galaxies are gravitationally bound into clusters, and even clusters of galaxies are found within larger superclusters. Given the complexity of the universe, the only way to condense its history into a single book is by a process of ruthless simplification. For much of this book, therefore, we will be considering the properties of an idealized, perfectly smooth, model universe. Only near the end of the book will we consider how relatively small objects, such as galaxies, clusters, and superclusters, are formed as the universe evolves. It is amusing to note in this context that the words *cosmology* and *cosmetology* come from the same Greek root: the word *kosmos*, meaning harmony or order. Just as cosmetologists try to make a human face more harmonious by smoothing over small blemishes such as pimples and wrinkles, cosmologists sometimes must smooth over small "blemishes" such as galaxies.

A science that regards entire galaxies as being small objects might seem, at first glance, very remote from the concerns of humanity. Nevertheless, cosmology deals with questions that are fundamental to the human condition. The questions that vex humanity are given in the title of a painting by Paul Gauguin (Figure 1.1): "Where do we come from? What are we? Where are we going?" Cosmology grapples with these questions by describing the past, explaining the present, and predicting the future of the universe. Cosmologists ask questions such as "What is the universe made of? Is it finite or infinite in spatial extent? Did it have a beginning some time in the past? Will it come to an end some time in the future?"

Cosmology deals with distances that are very large, objects that are very big, and timescales that are very long. Cosmologists frequently find that the standard SI units are not convenient for their purposes: the meter (m) is awkwardly short, the kilogram (kg) is awkwardly tiny, and the second (s) is awkwardly brief. Fortunately, we can adopt the units that have been developed by astronomers for dealing with large distances, masses, and times.

One distance unit used by astronomers is the astronomical unit (AU), equal to the mean distance between the Earth and Sun; in metric units, $1\,\text{AU} = 1.5 \times 10^{11}\,\text{m}$. Although the astronomical unit is a useful length scale within the solar

FIGURE 1.1 *Where Do We Come From? What Are We? Where Are We Going?* Paul Gauguin, 1897. [Museum of Fine Arts, Boston]

system, it is small compared to the distances between stars. To measure interstellar distances, it is useful to use the parsec (pc), equal to the distance at which 1 AU subtends an angle of 1 arcsecond; in metric units, $1\,\text{pc} = 3.1 \times 10^{16}\,\text{m}$. For example, we are at a distance of 1.3 pc from Proxima Centauri (the Sun's nearest neighbor among the stars) and 8500 pc from the center of our galaxy. Although the parsec is a useful length scale within our galaxy, it is small compared to the distances between galaxies. To measure intergalactic distances, we use the megaparsec (Mpc), equal to 10^6 pc, or 3.1×10^{22} m. For example, we are at a distance of 0.7 Mpc from M31 (otherwise known as the Andromeda galaxy) and 15 Mpc from the Virgo cluster (the nearest big cluster of galaxies).

The standard unit of mass used by astronomers is the solar mass ($M_\odot$); in metric units, the Sun's mass is $1\,M_\odot = 2.0 \times 10^{30}\,\text{kg}$. The total mass of our galaxy is not known as accurately as the mass of the Sun; in round numbers, though, it is $M_{\text{gal}} \approx 10^{12}\,M_\odot$. The Sun, incidentally, also provides the standard unit of power used in astronomy. The Sun's luminosity (that is, the rate at which it radiates away energy in the form of light) is $1\,L_\odot = 3.8 \times 10^{26}$ watts. The total luminosity of our galaxy is $L_{\text{gal}} = 3.6 \times 10^{10}\,L_\odot$.

For times much longer than a second, astronomers use the year (yr), defined as the time it takes the Earth to go once around the Sun. One year is approximately equal to 3.2×10^7 s. In a cosmological context, a year is frequently an inconveniently short period of time, so cosmologists often use gigayears (Gyr), equal to 10^9 yr, or 3.2×10^{16} s. For example, the age of the Earth is more conveniently written as 4.6 Gyr than as 1.5×10^{17} s.

In addition to dealing with very large things, cosmology also deals with very small things. Early in its history, as we shall see, the universe was very hot and dense, and some interesting particle physics phenomena were occurring. Consequently, particle physicists have plunged into cosmology, introducing some terminology and units of their own. For instance, particle physicists tend to measure energy units in electron volts (eV) instead of joules (J). The conversion factor between electron volts and joules is $1\,\text{eV} = 1.6 \times 10^{-19}$ J. The rest energy of an electron, for instance, is $m_e c^2 = 511{,}000\,\text{eV} = 0.511\,\text{MeV}$, and the rest energy of a proton is $m_p c^2 = 938.3\,\text{MeV}$.

When you stop to think of it, you realize that the units of meters, megaparsecs, kilograms, solar masses, seconds, and gigayears could only be devised by ten-fingered Earthlings obsessed with the properties of water. An eighteen-tentacled silicon-based lifeform from a planet orbiting Betelgeuse would probably devise a different set of units. A more universal, less culturally biased system of units is the Planck system, based on the universal constants G, c, and $\hbar$. Combining the Newtonian gravitational constant, $G = 6.7 \times 10^{-11}\,\text{m}^3\,\text{kg}^{-1}\,\text{s}^{-2}$, the speed of light, $c = 3.0 \times 10^8\,\text{m}\,\text{s}^{-1}$, and the reduced Planck constant, $\hbar = h/(2\pi) = 1.1 \times 10^{-34}\,\text{J}\,\text{s} = 6.6 \times 10^{-16}\,\text{eV}\,\text{s}$, yields a unique length scale, known as the Planck length:

$$\ell_P \equiv \left(\frac{G\hbar}{c^3}\right)^{1/2} = 1.6 \times 10^{-35}\,\text{m}. \tag{1.1}$$

The same constants can be combined to yield the Planck mass,[1]

$$M_P \equiv \left(\frac{\hbar c}{G}\right)^{1/2} = 2.2 \times 10^{-8}\,\text{kg}, \tag{1.2}$$

and the Planck time,

$$t_P \equiv \left(\frac{G\hbar}{c^5}\right)^{1/2} = 5.4 \times 10^{-44}\,\text{s}. \tag{1.3}$$

Using Einstein's relation between mass and energy, we can also define the Planck energy,

$$E_P = M_P c^2 = 2.0 \times 10^9\,\text{J} = 1.2 \times 10^{28}\,\text{eV}. \tag{1.4}$$

By bringing the Boltzmann constant, $k = 8.6 \times 10^{-5}\,\text{eV}\,\text{K}^{-1}$, into the act, we can also define the Planck temperature,

$$T_P = E_P/k = 1.4 \times 10^{32}\,\text{K}. \tag{1.5}$$

When distance, mass, time, and temperature are measured in the appropriate Planck units, then $c = k = \hbar = G = 1$. This is convenient for individuals who have difficulty in remembering the numerical values of physical constants. However, using Planck units can have potentially confusing side effects. For instance, many cosmology texts, after noting that $c = k = \hbar = G = 1$ when Planck units are used, then proceed to omit c, k, $\hbar$, and/or G from all equations. For instance, Einstein's celebrated equation, $E = mc^2$, becomes $E = m$. The blatant dimensional incorrectness of such an equation is jarring, but it simply means that

[1] The Planck mass is roughly equal to the mass of a grain of sand a quarter of a millimeter across.

the rest energy of an object, measured in units of the Planck energy, is equal to its mass, measured in units of the Planck mass. In this book, however, I will retain all factors of c, k, $\hbar$, and G, for the sake of clarity.

Here we will deal with distances ranging from the Planck length to 10^4 Mpc or so, a span of some 61 orders of magnitude. Dealing with such a wide range of length scales requires a stretch of the imagination, to be sure. However, cosmologists are not permitted to let their imaginations run totally unfettered. Cosmology, I emphasize strongly, is based ultimately on observation of the universe around us. Even in ancient times, cosmology was based on observations; unfortunately, those observations were frequently imperfect and incomplete. Ancient Egyptians, for instance, looked at the desert plains stretching away from the Nile valley and the blue sky overhead. Based on their observations, they developed a model of the universe in which a flat Earth (symbolized by the earth god Geb in Figure 1.2) was covered by a solid dome (symbolized by the sky goddess Nut). Greek cosmology was based on more precise and sophisticated observations. Ancient Greek astronomers deduced, from their observations, that the Earth and Moon are spherical, that the Sun is much farther from the Earth than the Moon is, and that the distance from the Earth to the stars is much greater than the Earth's diameter. Based on this knowledge, Greek cosmologists devised a "two-sphere" model of the universe, in which the spherical Earth is surrounded by a much larger celestial sphere, a spherical shell to which the stars are attached. Between the Earth and the celestial sphere, in this model, the Sun, Moon, and planets move on their complicated apparatus of epicycles and deferents.

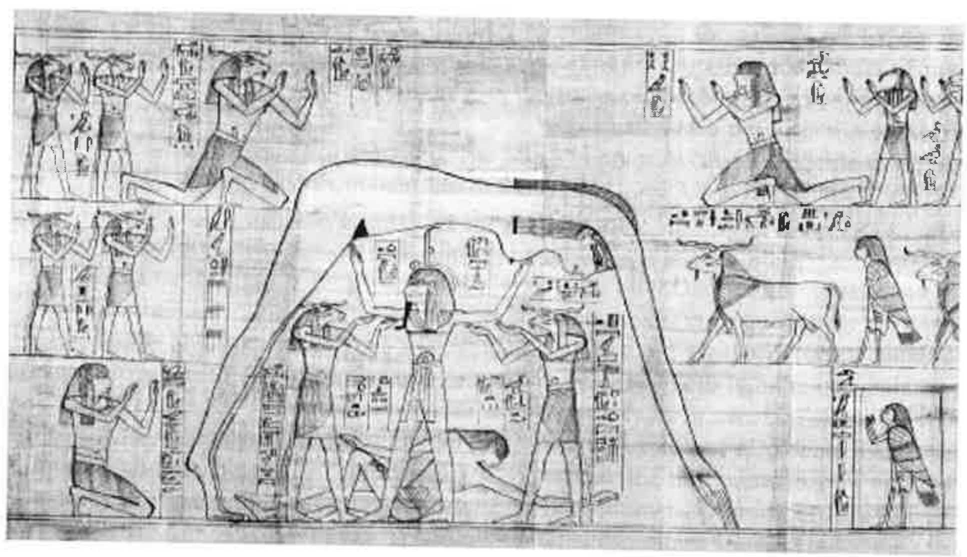


FIGURE 1.2 The ancient Egyptian view of the cosmos: the sky goddess Nut, supported by the air god Shu, arches over the earth god Geb (from the Greenfield Papyrus, ca. 1025 BC). [©Copyright The British Museum]

Although cosmology is ultimately based on observation, sometimes observations temporarily lag behind theory. During periods when data are lacking, cosmologists may adopt a new model for aesthetic or philosophical reasons. For instance, when Copernicus proposed a new Sun-centered model of the universe, to replace the Earth-centered two-sphere model of the Greeks, he didn't base his model on new observational discoveries. Rather, he believed that putting the Earth in motion around the Sun resulted in a conceptually simpler, more appealing model of the universe. Direct observational evidence didn't reveal that the Earth revolves around the Sun, rather than vice versa, until the discovery of the aberration of starlight in the year 1728, nearly two centuries after the death of Copernicus. Foucault didn't demonstrate the rotation of the Earth, another prediction of the Copernican model, until 1851, over *three* centuries after the death of Copernicus. However, although observations sometimes lag behind theory in this way, every cosmological model that isn't eventually supported by observational evidence must remain pure speculation.

The current standard model for the universe is the "Hot Big Bang" model, which states that the universe has expanded from an initially hot and dense state to its current relatively cool and tenuous state, and that the expansion is still going on today. To see why cosmologists have embraced the Hot Big Bang model, let us turn, in the next chapter, to the fundamental observations on which modern cosmology is based.

SUGGESTED READING

Full references are given in the Annotated Bibliography on page 235.

Cox (2000): Accurate values of physical and astronomical constants

Harrison (2000), ch. 1–4: A history of early (pre-Einstein) cosmology

CHAPTER 2

Fundamental Observations

Some of the observations on which modern cosmology is based are highly complex, requiring elaborate apparatus and sophisticated data analysis. However, other observations are surprisingly simple. Let's start with an observation that is deceptive in its extreme simplicity.

2.1 ■ THE NIGHT SKY IS DARK

Step outside on a clear, moonless night, far from city lights, and look upward. You will see a dark sky, with roughly two thousand stars scattered across it. The fact that the night sky is dark at visible wavelengths, instead of being uniformly bright with starlight, is known as *Olbers' Paradox*, after the astronomer Heinrich Olbers, who wrote a scientific paper on the subject in 1826. As it happens, Olbers was not the first person to think about Olbers' Paradox. As early as 1576, Thomas Digges mentioned how strange it is that the night sky is dark, with only a few pinpoints of light to mark the location of stars.[1]

Why should it be paradoxical that the night sky is dark? Most of us simply take for granted the fact that daytime is bright and nighttime is dark. The darkness of the night sky certainly posed no problems to the ancient Egyptians or Greeks, to whom stars were points of light stuck to a dome or sphere. However, the cosmological model of Copernicus required that the distance to stars be very much larger than an astronomical unit; otherwise, the parallax of the stars, as the Earth goes around on its orbit, would be large enough to see with the naked eye. Moreover, since the Copernican system no longer requires that the stars be attached to a rotating celestial sphere, the stars can be at different distances from the Sun. These liberating realizations led Thomas Digges, and other post-Copernican astronomers, to embrace a model in which stars are large glowing spheres, like the Sun, scattered throughout infinite space.

Let's compute how bright we expect the night sky to be in an infinite universe. Let n be the average number density of stars in the universe, and let L be the average stellar luminosity. The flux received here at Earth from a star of luminosity

[1]The name "Olbers' Paradox" is thus a prime example of what historians of science jokingly call the law of misonomy: nothing is ever named after the person who really discovers it.

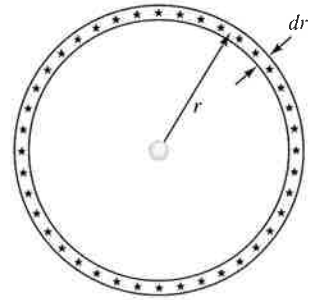


FIGURE 2.1 A star-filled spherical shell, of radius r and thickness dr, centered on the Earth.

L at a distance r is given by an inverse square law:

$$f(r) = \frac{L}{4\pi r^2}. \tag{2.1}$$

Now consider a thin spherical shell of stars, with radius r and thickness dr, centered on the Earth (Figure 2.1). The intensity of radiation from the shell of stars (that is, the power per unit area per steradian of the sky) is

$$dJ(r) = \frac{L}{4\pi r^2} \cdot n \cdot r^2 dr = \frac{nL}{4\pi} dr. \tag{2.2}$$

The total intensity of starlight from a shell thus depends only on its thickness, not on its distance from us. We can compute the total intensity of starlight from *all* the stars in the universe by integrating over shells of all radii:

$$J = \int_{r=0}^{\infty} dJ = \frac{nL}{4\pi} \int_0^{\infty} dr = \infty. \tag{2.3}$$

Thus, I have demonstrated that the night sky is infinitely bright.

This is utter nonsense.

Therefore, one (or more) of the assumptions that went into the above analysis of the sky brightness must be wrong. Let's scrutinize some of the assumptions. One assumption that I made is that we have an unobstructed line of sight to every star in the universe. This is not true. In fact, since stars have a finite angular size as seen from Earth, nearby stars will hide more distant stars from our view. Nevertheless, in an infinite distribution of stars, every line of sight should end at the surface of a star; this would imply a surface brightness for the sky equal to the surface brightness of a typical star. This is an improvement on an infinitely bright sky, but is still distinctly different from the dark sky we actually see. Heinrich Olbers himself tried to resolve Olbers' Paradox by proposing that distant stars are hidden from view by interstellar matter that absorbs starlight. This resolution does not work, because the interstellar matter would be heated by starlight until it had the same temperature as the surface of a star. At that point, the interstellar matter would emit as much light as it absorbs, and glow as brightly as the stars themselves.

A second assumption I made is that the number density n and mean luminosity L of stars are constant throughout the universe; more accurately, the assumption made in equation (2.3) is that the product nL is constant as a function of r. This might not be true. Distant stars might be less luminous or less numerous than nearby stars. If we are in a clump of stars of finite size, then the absence of stars at large distances will keep the night sky from being bright. Similarly, if distant stars are sufficiently low in luminosity compared to nearby stars, then they won't contribute significantly to the sky brightness. In order for the integrated intensity in equation (2.3) to be finite, the product nL must fall off more rapidly than $nL \propto 1/r$ as $r \to \infty$.

A third assumption is that the universe is infinitely large. This might not be true. If the universe extends only to a maximum distance $r_{\rm max}$ from us, then the total intensity of starlight we see in the night sky will be $J \sim nLr_{\rm max}/(4\pi)$. Note that this result will also be found if the universe is infinite in space, but is devoid of stars beyond a distance $r_{\rm max}$.

A fourth assumption, slightly more subtle than the previous ones, is that the universe is infinitely old. This might not be true. Because the speed of light is finite, when we look farther out in space, we are looking farther out in time. Thus, we see the Sun as it was 8.3 minutes ago, Proxima Centauri as it was 4 years ago, and M31 as it was 2 million years ago. If the universe has a finite age t_0, the intensity of starlight we see at night will be at most $J \sim nLct_0/(4\pi)$. Note that this result will also be found if the universe is infinitely old, but has only contained stars for a finite time t_0.

A fifth assumption is that the flux of light from a distant source is given by the inverse square law of equation (2.1). This might not be true. The assumption that $f \propto 1/r^2$ would have seemed totally innocuous to Olbers and other nineteenth-century astronomers; after all, the inverse square law follows directly from Euclid's laws of geometry. However, in the twentieth century, Albert Einstein, that great questioner of assumptions, demonstrated that the universe might not obey the laws of Euclidean geometry. In addition, the inverse square law assumes that the source of light is stationary relative to the observer. If the universe is systematically expanding or contracting, then the light from distant sources will be redshifted to lower photon energies or blueshifted to higher photon energies.

Thus, the infinitely large, eternally old, Euclidean universe that Thomas Digges and his successors pictured simply does not hold up to scrutiny. This is a textbook, not a suspense novel, so I'll tell you right now: the primary resolution to Olbers' Paradox comes from the fact that the universe has a finite age. The stars beyond some finite distance, called the horizon distance, are invisible to us because their light hasn't had time to reach us yet. A particularly amusing bit of cosmological trivia is that the first person to hint at the correct resolution of Olbers' Paradox was Edgar Allen Poe.[2] In his essay "Eureka: A Prose Poem," completed in 1848,

[2]That's right, the "Nevermore" guy.

Poe wrote, "Were the succession of stars endless, then the background of the sky would present us an [*sic*] uniform density . . . since there could be absolutely no point, in all that background, at which would not exist a star. The only mode, therefore, in which, under such a state of affairs, we could comprehend the voids which our telescopes find in innumerable directions, would be by supposing the distance of the invisible background so immense that no ray from it has yet been able to reach us at all."

2.2 ■ ON LARGE SCALES, THE UNIVERSE IS ISOTROPIC AND HOMOGENEOUS

What does it mean to state that the universe is isotropic and homogeneous? Saying that the universe is *isotropic* means that there are no preferred directions in the universe; it looks the same no matter which way you point your telescope. Saying that the universe is *homogeneous* means that there are no preferred locations in the universe; it looks the same no matter where you set up your telescope. Note the very important qualifier: the universe is isotropic and homogeneous *on large scales*. In this context, large scales means that the universe is only isotropic and homogeneous on scales of roughly 100 Mpc or more.

The isotropy of the universe is not immediately obvious. In fact, on small scales, the universe is blatantly anisotropic. Consider, for example, a sphere 3 m in diameter, centered on your navel (Figure 2.2a). Within this sphere, there is a preferred direction; it is the direction commonly referred to as "down." It is easy to determine the vector pointing down. Just let go of a small dense object. The object doesn't hover in midair, and it doesn't move in a random direction; it falls down, toward the center of the Earth.

On significantly larger scales, the universe is still anisotropic. Consider, for example, a sphere 3 AU in diameter, centered on your navel (Figure 2.2b). Within this sphere, there is a preferred direction; it is the direction pointing toward the Sun, which is by far the most massive and most luminous object within the sphere. It is easy to determine the vector pointing toward the Sun. Just step outside on a sunny day, and point to that really bright disk of light up in the sky.

On still larger scales, the universe is *still* anisotropic. Consider, for example, a sphere 3 Mpc in diameter, centered on your navel (Figure 2.2c). This sphere contains the Local Group of galaxies, a small cluster of some 40 galaxies. By far the most massive and most luminous galaxies in the Local Group are our own galaxy and M31, which together contribute about 86% of the total luminosity within the 3 Mpc sphere. Thus, within this sphere, our galaxy and M31 define a preferred direction. It is fairly easy to determine the vector pointing from our galaxy to M31; just step outside on a clear night when the constellation Andromeda is above the horizon, and point to the fuzzy oval in the middle of the constellation.

It isn't until you get to considerably larger scales that the universe can be considered as isotropic. Consider a sphere 200 Mpc in diameter, centered on your navel. Figure 2.2d shows a slice through such a sphere, with superclusters of

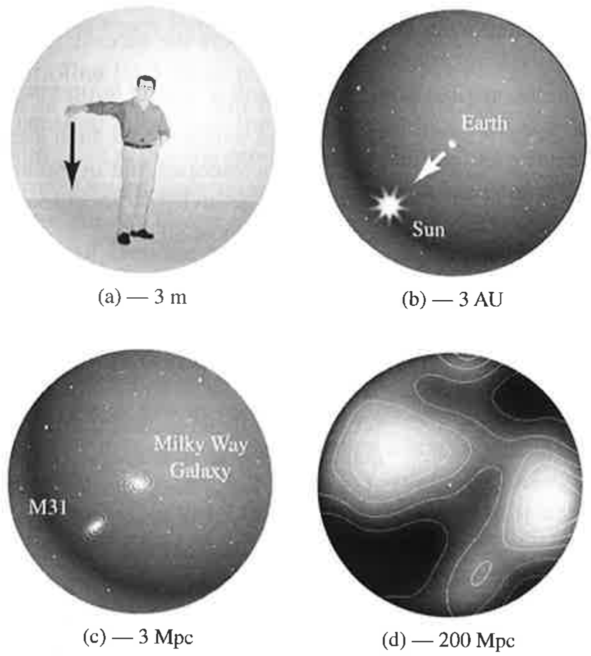
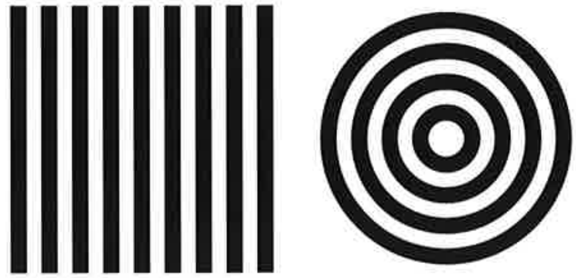


FIGURE 2.2 (a) A sphere 3 m in diameter, centered on your navel. (b) A sphere 3 AU in diameter, centered on your navel. (c) A sphere 3 Mpc in diameter, centered on your navel. (d) A sphere 200 Mpc in diameter, centered on your navel. Shown is the number density of galaxies smoothed with a Gaussian of width 17 Mpc. The heavy contour is drawn at the mean density; darker regions represent higher density, lighter regions represent lower density.

galaxies indicated as light patches. The Perseus–Pisces supercluster is on the right, the Hydra–Centaurus supercluster is on the left, and the edge of the Coma supercluster is just visible at the top of Figure 2.2d. Superclusters are typically $\sim 100\,\text{Mpc}$ along their longest dimensions, and are separated by voids (low density regions) which are typically $\sim 100\,\text{Mpc}$ across. These are the largest structures in the universe, it seems; surveys of the universe on still larger scales don't find "superduperclusters."

On small scales, the universe is obviously inhomogeneous, or lumpy, in addition to being anisotropic. For instance, a sphere 3 m in diameter, centered on your navel, will have an average density of $\sim 100\,\text{kg}\,\text{m}^{-3}$, in round numbers. However, the average density of the universe as a whole is $\rho_0 \sim 3 \times 10^{-27}\,\text{kg}\,\text{m}^{-3}$. Thus, on a scale $d \sim 3\,\text{m}$, the patch of the universe surrounding you is more than 28 orders of magnitude denser than average.

On significantly larger scales, the universe is still inhomogeneous. A sphere 3 AU in diameter, centered on your navel, has an average density of $4 \times 10^{-5}\,\text{kg}\,\text{m}^{-3}$; that's 22 orders of magnitude denser than the average for the universe.



(a) (b)

FIGURE 2.3 (a) A pattern that is anisotropic, but is homogeneous on scales larger than the stripe width. (b) A pattern that is isotropic about the origin, but is inhomogeneous.

On still larger scales, the universe is *still* inhomogeneous. A sphere 3 Mpc in diameter, centered on your navel, will have an average density of $\sim 3 \times 10^{-26}\,\text{kg}\,\text{m}^{-3}$, still an order of magnitude denser than the universe as a whole. It's only when you contemplate a sphere $\sim 100\,\text{Mpc}$ in diameter that a sphere centered on your navel is not overdense compared to the universe as a whole.

Note that homogeneity does not imply isotropy. A sheet of paper printed with stripes (Figure 2.3a) is homogeneous on scales larger than the stripe width, but it is not isotropic. The direction of the stripes provides a preferred direction by which you can orient yourself. Note also that isotropy around a single point does not imply homogeneity. A sheet of paper printed with a bullseye (Figure 2.3b) is isotropic around the center of the bullseye, but is it not homogeneous. The rings of the bullseye look different far from the center than they do close to the center. You can tell where you are relative to the center by measuring the radius of curvature of the nearest ring.

In general, then, saying that something is homogeneous is quite different from saying it is isotropic. However, modern cosmologists have adopted the *cosmological principle*, which states "There is nothing special about our location in the universe." The cosmological principle holds true only on large scales (of 100 Mpc or more). On smaller scales, your navel obviously is in a special location. Most spheres 3 m across don't contain a sentient being; most spheres 3 AU across don't contain a star; most spheres 3 Mpc across don't contain a pair of bright galaxies. However, most spheres over 100 Mpc across do contain roughly the same pattern of superclusters and voids, statistically speaking. The universe, on scales of 100 Mpc or more, appears to be isotropic around us. Isotropy around any point in the universe, such as your navel, combined with the cosmological principle, implies isotropy around every point in the universe; and isotropy around every point in the universe *does* imply homogeneity.

The cosmological principle has the alternate name of the "Copernican principle" as a tribute to Copernicus, who pointed out that the Earth is not the center of the universe. Later cosmologists also pointed out the Sun is not the center, that our galaxy is not the center, and that the Local Group is not the center. In fact, there is no center to the universe.

2.3 ■ GALAXIES SHOW A REDSHIFT PROPORTIONAL TO THEIR DISTANCE

When we look at a galaxy at visible wavelengths, we detect primarily the light from the stars that the galaxy contains. Thus, when we take a galaxy's spectrum at visible wavelengths, it typically contains absorption lines created in the stars' relatively cool upper atmospheres.[3] Suppose we consider a particular absorption line whose wavelength, as measured in a laboratory here on Earth, is λ_{em}. The wavelength we measure for the same absorption line in a distant galaxy's spectrum, λ_{ob}, will not, in general, be the same. We say that the galaxy has a redshift z, given by the formula

$$z \equiv \frac{\lambda_{ob} - \lambda_{em}}{\lambda_{em}}. \tag{2.4}$$

Strictly speaking, when $z < 0$, this quantity is called a blueshift, rather than a redshift. However, the vast majority of galaxies have $z > 0$.

The fact that the light from galaxies is generally redshifted to longer wavelengths, rather than blueshifted to shorter wavelengths, was not known until the twentieth century. In 1912, Vesto Slipher at the Lowell Observatory measured the shift in wavelength of the light from M31; this galaxy, as it turns out, is one of the few that exhibits a blueshift. By 1925, Slipher had measured the shifts in the spectral lines for approximately 40 galaxies, finding that they were nearly all redshifted; the exceptions were all nearby galaxies within the Local Group.

By 1929, enough galaxy redshifts had been measured for the cosmologist Edwin Hubble to make a study of whether a galaxy's redshift depends on its distance from us. Although measuring a galaxy's redshift is relatively easy, and can be done with high precision, measuring its distance is difficult. Hubble knew z for nearly 50 galaxies, but had estimated distances for only 20 of them. Nevertheless, from a plot of redshift (z) versus distance (r), reproduced in Figure 2.4, he found the famous linear relation now known as Hubble's Law:

$$z = \frac{H_0}{c} r, \tag{2.5}$$

where H_0 is a constant (now called the Hubble constant). Hubble interpreted the observed redshift of galaxies as being a Doppler shift due to their radial velocity away from Earth. Because the values of z in Hubble's analysis were all small ($z < 0.004$), he was able to use the classical, nonrelativistic relation for the Doppler shift, $z = v/c$, where v is the radial velocity of the light source (in this case, a galaxy). Interpreting the redshifts as Doppler shifts, Hubble's law takes the form

$$v = H_0 r. \tag{2.6}$$

The Hubble constant H_0 can be found by dividing velocity by distance, so it is customarily written in the rather baroque units of $\mathrm{km\,s^{-1}\,Mpc^{-1}}$. When Hubble

[3] Galaxies containing active galactic nuclei will also show *emission* lines from the hot gas in their nuclei.

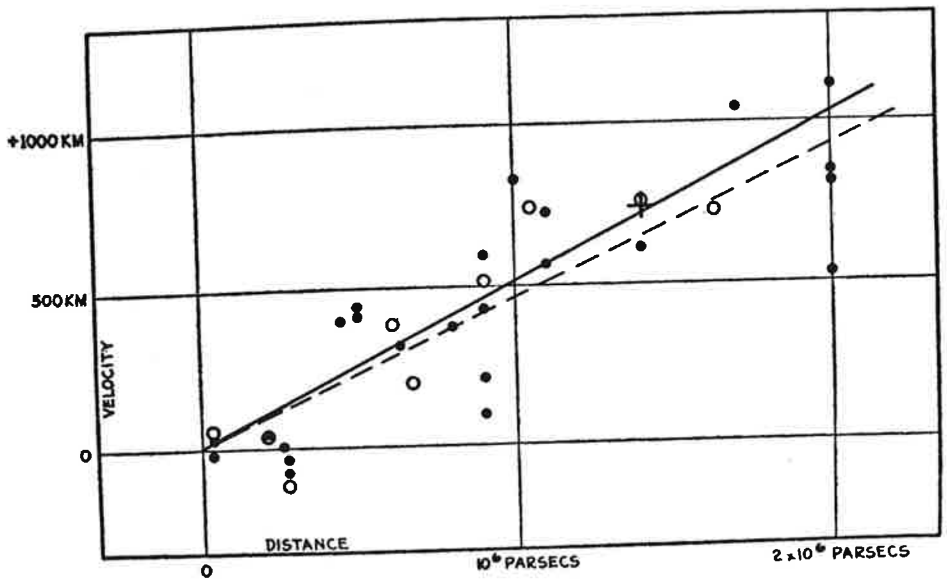


FIGURE 2.4 Edwin Hubble's original plot of the relation between redshift (vertical axis) and distance (horizontal axis). Note that the vertical axis actually plots cz rather than z, and that the units are accidentally written as km rather than km/s.

first discovered Hubble's Law, he thought that the numerical value of the Hubble constant was $H_0 = 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (see Figure 2.4). However, it turned out that Hubble was severely underestimating the distances to galaxies.

Figure 2.5 shows a more recent determination of the Hubble constant from nearby galaxies, using data obtained by (appropriately enough) the Hubble Space Telescope. The best current estimate of the Hubble constant, combining the results of different research groups, is

$$H_0 = 70 \pm 7\,\mathrm{km\,s^{-1}\,Mpc^{-1}}. \tag{2.7}$$

This is the value for the Hubble constant that we will use in the remainder of this book.

Cosmological innocents sometimes exclaim, when first encountering Hubble's Law, "Surely it must be a violation of the cosmological principle to have all those distant galaxies moving away from *us*! It looks as if we are at a special location in the universe—the point away from which all other galaxies are fleeing." In fact, what we see here in our galaxy is exactly what you would expect to see in a universe that is undergoing homogeneous and isotropic expansion. We see distant galaxies moving away from us; but observers in any other galaxy would also see distant galaxies moving away from them.

To see on a more mathematical level what we mean by homogeneous, isotropic expansion, consider three galaxies at positions $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$. They define a triangle

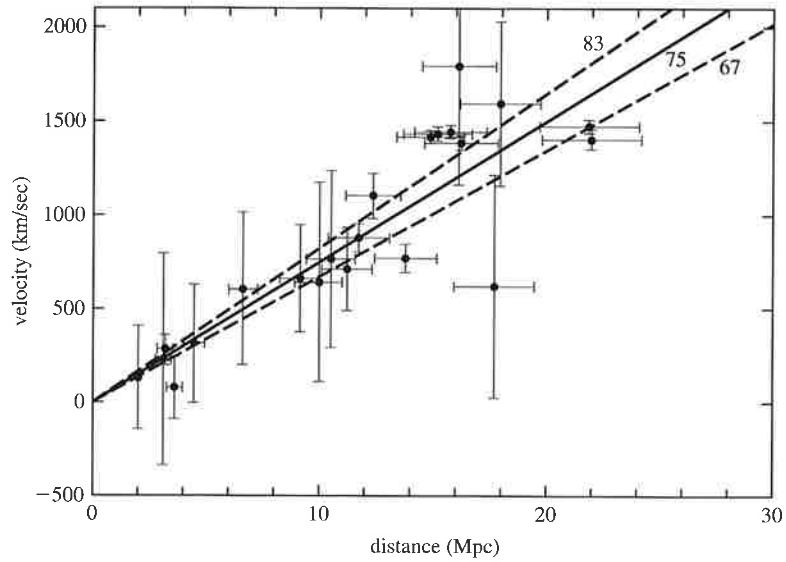


FIGURE 2.5 A more modern version of Hubble's plot, showing cz versus distance. In this case, the galaxy distances have been determined using Cepheid variable stars as standard candles, as described in Chapter 6.

(Figure 2.6) with sides of length

$$r_{12} \equiv |\vec{r}_1 - \vec{r}_2| \tag{2.8}$$

$$r_{23} \equiv |\vec{r}_2 - \vec{r}_3| \tag{2.9}$$

$$r_{31} \equiv |\vec{r}_3 - \vec{r}_1| \,. \tag{2.10}$$

Homogeneous and uniform expansion means that the shape of the triangle is preserved as the galaxies move away from each other. Maintaining the correct relative lengths for the sides of the triangle requires an expansion law of the form

$$r_{12}(t) = a(t) r_{12}(t_0) \tag{2.11}$$

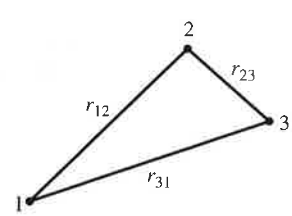


FIGURE 2.6 A triangle defined by three galaxies in a uniformly expanding universe.

$$r_{23}(t) = a(t)r_{23}(t_0) \quad (2.12)$$

$$r_{31}(t) = a(t)r_{31}(t_0). \quad (2.13)$$

Here the function $a(t)$ is a *scale factor*, equal to one at the present moment ($t = t_0$) and totally independent of location or direction. The scale factor $a(t)$ tells us how the expansion (or possibly contraction) of the universe depends on time. At any time t, an observer in galaxy 1 will see the other galaxies receding with a speed

$$v_{12}(t) = \frac{dr_{12}}{dt} = \dot{a}r_{12}(t_0) = \frac{\dot{a}}{a}r_{12}(t) \quad (2.14)$$

$$v_{31}(t) = \frac{dr_{31}}{dt} = \dot{a}r_{31}(t_0) = \frac{\dot{a}}{a}r_{31}(t). \quad (2.15)$$

You can demonstrate easily that an observer in galaxy 2 or galaxy 3 will find the same linear relation between observed recession speed and distance, with $\dot{a}/a$ playing the role of the Hubble constant. Since this argument can be applied to any trio of galaxies, it implies that in any universe where the distribution of galaxies is undergoing homogeneous, isotropic expansion, the velocity–distance relation takes the linear form $v = Hr$, with $H = \dot{a}/a$.

If galaxies are currently moving away from each other, then it implies they were closer together in the past. Consider a pair of galaxies currently separated by a distance r, with a velocity $v = H_0 r$ relative to each other. If there are no forces acting to accelerate or decelerate their relative motion, then their velocity is constant, and the time that has elapsed since they were in contact is

$$t_0 = \frac{r}{v} = \frac{r}{H_0 r} = H_0^{-1}, \quad (2.16)$$

independent of the current separation r. The time H_0^{-1} is referred to as the *Hubble time*. For $H_0 = 70 \pm 7\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, the Hubble time is $H_0^{-1} = 14.0 \pm 1.4\,\mathrm{Gyr}$. If the relative velocities of galaxies have been constant in the past, then one Hubble time ago, all the galaxies in the universe were crammed together into a small volume. Thus, the observation of galactic redshifts lead naturally to a *Big Bang* model for the evolution of the universe. A Big Bang model may be broadly defined as a model in which the universe expands from an initially highly dense state to its current low-density state.

The Hubble time of $\sim$ 14 Gyr is comparable to the ages computed for the oldest known stars in the universe. This rough equivalence is reassuring. However, the age of the universe—that is, the time elapsed since its original highly dense state—is not necessarily exactly equal to the Hubble time. We know that gravity exists, and that galaxies contain matter. If gravity working on matter is the only force at work on large scales, then the attractive force of gravity will act to slow the expansion. In this case, the universe was expanding more rapidly in the past than it is now, and the universe is younger than H_0^{-1}. On the other hand, if the energy density of the universe is dominated by a cosmological constant (an entity we'll examine in more detail in Chapter 4), then the dominant gravitational force is repulsive, and the universe may be older than H_0^{-1}.

Just as Hubble time provides a natural time scale for our universe, the Hubble distance, $c/H_0 = 4300 \pm 400\,\mathrm{Mpc}$, provides a natural distance scale. Just as the age of the universe is roughly equal to H_0^{-1} in most Big Bang models, with the exact value depending on the expansion history of the universe, so the horizon distance (the greatest distance a photon can travel during the age of the universe) is roughly equal to c/H_0, with the exact value, again, depending on the expansion history. (Later chapters will deal with computing the exact values of the age and horizon size of our universe.)

Note how Hubble's Law ties in with Olbers' Paradox. If the universe is of finite age, $t_0 \sim H_0^{-1}$, then the night sky can be dark, even if the universe is infinitely large, because light from distant galaxies has not yet had time to reach us. Galaxy surveys tell us that the luminosity density of galaxies in the local universe is

$$nL \approx 2 \times 10^8\,\mathrm{L}_\odot\,\mathrm{Mpc}^{-3}. \tag{2.17}$$

By terrestrial standards, the universe is not a well-lit place; this luminosity density is equivalent to a single 40 watt light bulb within a sphere 1 AU in radius. If the horizon distance is $d_{\mathrm{hor}} \sim c/H_0$, then the total flux of light we receive from all the stars from all the galaxies within the horizon will be

$$\begin{aligned} F_{\mathrm{gal}} = 4\pi J_{\mathrm{gal}} &\approx nL \int_0^{r_H} dr \sim nL\left(\frac{c}{H_0}\right) \\ &\sim 9 \times 10^{11}\,\mathrm{L}_\odot\,\mathrm{Mpc}^{-2} \sim 2 \times 10^{-11}\,\mathrm{L}_\odot\,\mathrm{AU}^{-2}. \end{aligned} \tag{2.18}$$

By the cosmological principle, this is the total flux of starlight you would expect at any randomly located spot in the universe. Comparing this to the flux we receive from the Sun,

$$F_{\mathrm{sun}} = \frac{1\,\mathrm{L}_\odot}{4\pi\,\mathrm{AU}^2} \approx 0.08\,\mathrm{L}_\odot\,\mathrm{AU}^{-2}, \tag{2.19}$$

we find that $F_{\mathrm{gal}}/F_{\mathrm{sun}} \sim 3 \times 10^{-10}$. Thus, the total flux of starlight at a randomly selected location in the universe is less than a billionth the flux of light we receive from the Sun here on Earth. For the entire universe to be as well-lit as the Earth, it would have to be over a billion times older than it is; *and* you'd have to keep the stars shining during all that time.

Hubble's Law occurs naturally in a Big Bang model for the universe, in which homogeneous and isotropic expansion causes the density of the universe to decrease steadily from its initial high value. In a Big Bang model, the properties of the universe evolve with time; the average density decreases, the mean distance between galaxies increases, and so forth. However, Hubble's Law can also be explained by a *Steady State* model. The Steady State model was first proposed in the 1940's by Hermann Bondi, Thomas Gold, and Fred Hoyle, who were proponents of the *perfect cosmological principle*, which states that not only are there no privileged locations in space, there are no privileged moments in time. Thus, a Steady State universe is one in which the global properties of the universe, such as the mean density ρ_0 and the Hubble constant H_0, remain constant with time.

In a Steady State universe, the velocity–distance relation

$$\frac{dr}{dt} = H_0 r \tag{2.20}$$

can be easily integrated, since H_0 is constant with time, to yield an exponential law:

$$r(t) \propto e^{H_0 t}. \tag{2.21}$$

Note that $r \to 0$ only in the limit $t \to -\infty$; a Steady State universe is infinitely old. If there existed an instant in time at which the universe started expanding (as in a Big Bang model), that would be a special moment, in violation of the assumed "perfect cosmological principle." The volume of a spherical region of space, in a Steady State model, increases exponentially with time:

$$V = \frac{4\pi}{3} r^3 \propto e^{3H_0 t}. \tag{2.22}$$

However, if the universe is in a steady state, the density of the sphere must remain constant. To have a constant density of matter within a growing volume, matter must be continuously created at a rate

$$\dot{M}_{ss} = \rho_0 \dot{V} = \rho_0 3 H_0 V. \tag{2.23}$$

If our own universe, with matter density $\rho_0 \sim 3 \times 10^{-27}\,\mathrm{kg\,m^{-3}}$, happened to be a Steady State universe, then matter would have to be created at a rate

$$\frac{\dot{M}_{ss}}{V} = 3H_0\rho_0 \sim 6 \times 10^{-28}\,\mathrm{kg\,m^{-3}\,Gyr^{-1}}. \tag{2.24}$$

This corresponds to creating roughly one hydrogen atom per cubic kilometer per year.

During the 1950s and 1960s, the Big Bang and Steady State models battled for supremacy. Critics of the Steady State model pointed out that the continuous creation of matter violates mass-energy conservation. Supporters of the Steady State model pointed out that the continuous creation of matter is no more absurd than the instantaneous creation of the entire universe in a single "Big Bang" billions of years ago.[4] The Steady State model finally fell out of favor when observational evidence increasingly indicated that the perfect cosmological principle is not true. The properties of the universe *do*, in fact, change with time. The discovery of the

[4] The name "Big Bang" was actually coined by Fred Hoyle, a supporter of the Steady State model.

Cosmic Microwave Background, discussed in section 2.5, is commonly regarded as the observation that decisively tipped the scales in favor of the Big Bang model.

2.4 ■ THE UNIVERSE CONTAINS DIFFERENT TYPES OF PARTICLES

It doesn't take a brilliant observer to confirm that the universe contains a variety of different things: shoes, ships, sealing wax, cabbages, kings, galaxies, and what have you. From a cosmologist's viewpoint, though, cabbages and kings are nearly indistinguishable—the main difference between them is that the mean mass per king is greater than the mean mass per cabbage. From a cosmological viewpoint, the most significant difference between the different components of the universe is that they are made of different elementary particles. The properties of the most cosmologically important particles are summarized in Table 2.1.

The material objects that surround us in our everyday life are made up of *protons*, *neutrons*, and *electrons*.[5] Protons and neutrons are examples of *baryons*, where a baryon is defined as a particle made of three quarks. A proton (p) contains two "up" quarks, each with an electrical charge of $+\frac{2}{3}$, and a "down" quark, with charge $-\frac{1}{3}$. A neutron (n) contains one "up" quark and two "down" quarks. Thus a proton has a net positive charge of $+1$, while a neutron is electrically neutral. Protons and neutrons also differ in their mass—or equivalently, in their rest energies. The proton mass is $m_p c^2 = 938.3\,\text{MeV}$, while the neutron mass is $m_n c^2 = 939.6\,\text{MeV}$, about 0.1% greater. Free neutrons are unstable, decaying into protons with a decay time of $\tau_n = 890\,\text{s}$, about a quarter of an hour. By contrast, experiments have put a lower limit on the decay time of the proton, which is very much greater than the Hubble time. Neutrons can be preserved against decay by binding them into an atomic nucleus with one or more protons.

Electrons (e^-) are examples of *leptons*, a class of elementary particles that are not made of quarks. The mass of an electron is much smaller than that of a neutron or proton; the rest energy of an electron is $m_e c^2 = 0.511\,\text{MeV}$. An electron has an electric charge equal in magnitude to that of a proton, but opposite in sign. On large scales, the universe is electrically neutral; the number of electrons is equal to the number of protons. Since protons outmass electrons by a factor of

TABLE 2.1 Particle Properties

Particle	Symbol	Rest energy (MeV)	Charge
proton	p	938.3	+1
neutron	n	939.6	0
electron	e^-	0.511	−1
neutrino	ν_e, ν_μ, ν_τ	?	0
photon	γ	0	0
dark matter	?	?	0

[5]For that matter, we ourselves are made of protons, neutrons, and electrons.

1836 to 1, the mass density of electrons is only a small perturbation to the mass density of protons and neutrons. For this reason, the component of the universe made up of ions, atoms, and molecules is generally referred to as *baryonic matter*, since only the baryons (protons and neutrons) contribute significantly to the mass density. Protons and neutrons are 800-pound gorillas; electrons are only 7-ounce bushbabies.

About three-fourths of the baryonic matter in the universe is currently in the form of ordinary hydrogen, the simplest of all elements. In addition, when we look at the remainder of the baryonic matter, it is primarily in the form of helium, the next simplest element. The Sun's atmosphere, for instance, contains 70% hydrogen by mass, and 28% helium; only 2% is contributed by more massive atoms. When astronomers look at a wide range of astronomical objects—stars and interstellar gas clouds, for instance—they find a minimum helium mass fraction of 24%. The baryonic component of the universe can be described, to lowest order, as a mix of three parts hydrogen to one part helium, with only minor contamination by heavier elements.

Another type of lepton, in addition to the electron, is the *neutrino* (ν). The most poetic summary of the properties of the neutrino was made by John Updike, in his poem "Cosmic Gall":[6]

> Neutrinos, they are very small.
> They have no charge and have no mass
> And do not interact at all.
> The earth is just a silly ball
> To them, through which they simply pass,
> Like dustmaids down a drafty hall
> Or photons through a sheet of glass.

In truth, Updike was using a bit of poetic license here. It is definitely true that neutrinos have no charge.[7] However, it is not true that neutrinos "do not interact at all"; they actually are able to interact with other particles via the weak nuclear force. The weak nuclear force, though, is very weak indeed; a typical neutrino emitted by the Sun would have to pass through a few parsecs of solid lead before having a 50% chance of interacting with a lead atom. Since neutrinos pass through neutrino detectors with the same facility with which they pass through the Earth, detecting neutrinos from astronomical sources is difficult.

There are three types, or "flavors," of neutrinos: electron neutrinos, muon neutrinos, and tau neutrinos. What Updike didn't know in 1960, when he wrote his poem, is that some or all of the neutrino types probably have a small mass. The evidence for massive neutrinos comes indirectly, from the search for neutrino oscillations. An *oscillation* is the transmutation of one flavor of neutrino into another. The rate at which two neutrino flavors oscillate is proportional to the difference

[6]From COLLECTED POEMS 1953–1993 by John Updike, ©1993 by John Updike. Used by permission of Alfred A. Knopf, a division of Random House, Inc.
[7]Their name, given them by Enrico Fermi, means "little neutral one" in Italian.

of the squares of their masses. Observations of neutrinos from the Sun are most easily explained if electron neutrinos (the flavor emitted by the Sun) oscillate into some other flavor of neutrino, with the difference in the squares of their masses being $\Delta(m_\nu^2 c^4) \approx 5 \times 10^{-5}\,\mathrm{eV}^2$. Observations of muon neutrinos created by cosmic rays striking the upper atmosphere indicate that muon neutrinos oscillate into tau neutrinos, with $\Delta(m_\nu^2 c^4) \approx 3 \times 10^{-3}\,\mathrm{eV}^2$ for these two flavors. Unfortunately, knowing the differences of the squares of the masses doesn't tell us the values of the masses themselves.

A particle which is known to be massless is the *photon*. Electromagnetic radiation can be thought of either as a wave or as a stream of particles, called photons. Light, when regarded as a wave, is characterized by its frequency f or its wavelength $\lambda = c/f$. When light is regarded as a stream of photons, each photon is characterized by its energy, $E_\gamma = hf$, where $h = 2\pi\hbar$ is the Planck constant. Photons of a wide range of energy, from radio to gamma rays, pervade the universe. Unlike neutrinos, photons interact readily with electrons, protons, and neutrons. For instance, photons can ionize an atom by kicking an electron out of its orbit, a process known as *photoionization*. Higher-energy photons can break an atomic nucleus apart, a process known as *photodissociation*.

Photons, in general, are easily created. One way to make photons is to take a dense, opaque object—such as the filament of an incandescent lightbulb—and heat it up. If an object is opaque, then the protons, neutrons, electrons, and photons that it contains frequently interact, and attain thermal equilibrium. When a system is in thermal equilibrium, the density of photons in the system, as a function of photon energy, depends only on the temperature T. It doesn't matter whether the system is a tungsten filament, an ingot of steel, or a sphere of ionized hydrogen and helium. The energy density of photons in the frequency range $f \to f + df$ is given by the *blackbody* function

$$\varepsilon(f)df = \frac{8\pi h}{c^3}\frac{f^3\,df}{\exp(hf/kT) - 1}, \tag{2.25}$$

illustrated in Figure 2.7. The peak in the blackbody function occurs at $hf_{\rm peak} \approx 2.82kT$. Integrated over all frequencies, equation (2.25) yields a total energy density for blackbody radiation of

$$\varepsilon_\gamma = \alpha T^4, \tag{2.26}$$

where

$$\alpha = \frac{\pi^2}{15}\frac{k^4}{\hbar^3 c^3} = 7.56 \times 10^{-16}\,\mathrm{J\,m^{-3}\,K^{-4}}. \tag{2.27}$$

The number density of photons in blackbody radiation can be computed from equation (2.25) as

$$n_\gamma = \beta T^3, \tag{2.28}$$

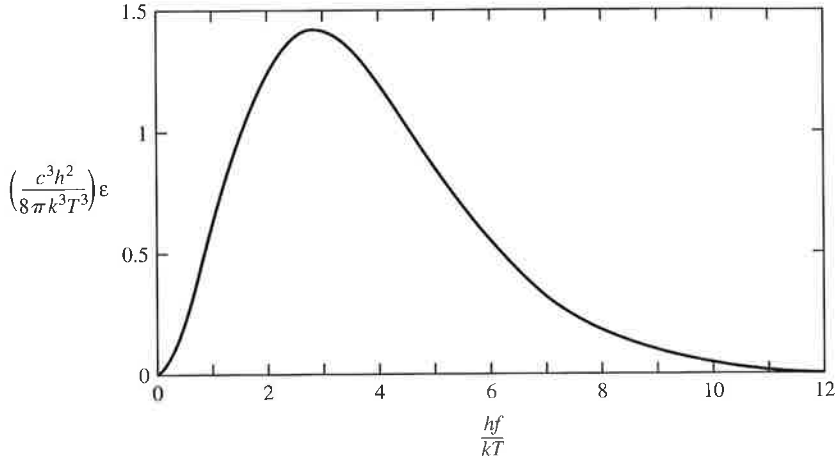


FIGURE 2.7 The energy distribution of a blackbody spectrum.

where

$$\beta = \frac{2.404}{\pi^2}\frac{k^3}{\hbar^3 c^3} = 2.03 \times 10^7\,\mathrm{m}^{-3}\,\mathrm{K}^{-3}. \tag{2.29}$$

Division of equation (2.26) by equation (2.28) yields a mean photon energy of $E_{\rm mean} = hf_{\rm mean} \approx 2.70kT$, close to the peak in the spectrum. You have a temperature of 310 K, and you radiate an approximate blackbody spectrum, with a mean photon energy of $E_{\rm mean} \approx 0.072\,\mathrm{eV}$, corresponding to a wavelength of $\lambda \approx 1.7 \times 10^{-5}\,\mathrm{m}$, in the infrared. By contrast, the Sun produces an approximate blackbody spectrum with a temperature $T_\odot \approx 5800\,\mathrm{K}$. This implies a mean photon energy $E_{\rm mean} \approx 1.3\,\mathrm{eV}$, corresponding to $\lambda \approx 9.0 \times 10^{-7}\,\mathrm{m}$, in the near infrared. Note, however, that although the mean photon energy in a blackbody spectrum is $\sim 3kT$, Figure 2.7 shows us that there is a long exponential tail to higher photon energies. A large fraction of the Sun's output is at wavelengths of $(4 \rightarrow 7) \times 10^{-7}\,\mathrm{m}$, which our eyes are equipped to detect.

The most mysterious component of the universe is *dark matter*. When observational astronomers refer to dark matter, they usually mean any massive component of the universe that is too dim to be detected readily using current technology. Thus, stellar remnants such as white dwarfs, neutron stars, and black holes are sometimes referred to as dark matter, since an isolated stellar remnant is extremely faint and difficult to detect. Substellar objects such as brown dwarfs are also referred to as dark matter, since brown dwarfs, too low in mass for nuclear fusion to occur in their cores, are very dim. Theoretical astronomers sometimes use a more stringent definition of dark matter than do observers, defining dark matter as any massive component of the universe which doesn't emit, absorb, or scatter light at all.[8] If neutrinos have mass, for instance, as the recent neutrino oscillation results indicate, they qualify as dark matter. In some extensions to the Standard Model of particle physics, there exist massive particles that interact, like neutrinos, only through the weak nuclear force and through gravity. These particles, which have not yet been detected in the laboratory, are generically referred to as Weakly Interacting Massive Particles, or WIMPs.

In this book, we will generally adopt the broader definition of dark matter as something which is too dim for us to see, even with our best available technology. Detecting dark matter is, naturally, difficult. The standard method of detecting dark matter is by measuring its gravitational effect on luminous matter, just as the planet Neptune was first detected by its gravitational effect on the planet Uranus. Although Neptune no longer qualifies as dark matter, observations of the motions of stars within galaxies and of galaxies within clusters indicate that a significant amount of dark matter is in the universe. Exactly how much there is, and what it's made of, is a topic of great interest to cosmologists.

2.5 ■ THE UNIVERSE IS FILLED WITH A COSMIC MICROWAVE BACKGROUND

The discovery of the Cosmic Microwave Background (CMB) by Arno Penzias and Robert Wilson in 1965 has entered cosmological folklore. Using a microwave antenna at Bell Labs, they found an isotropic background of microwave radiation. More recently, the Cosmic Background Explorer (COBE) satellite has revealed that the Cosmic Microwave Background is exquisitely well fitted by a blackbody spectrum (equation (2.25)) with a temperature

$$T_0 = 2.725 \pm 0.001\,\mathrm{K}. \tag{2.30}$$

The energy density of the CMB is, from equation (2.26),

$$\varepsilon_\gamma = 4.17 \times 10^{-14}\,\mathrm{J\,m^{-3}}. \tag{2.31}$$

This is roughly equivalent to a quarter of an MeV per cubic meter of space. The number density of CMB photons is, from equation (2.28),

$$n_\gamma = 4.11 \times 10^{8}\,\mathrm{m^{-3}}. \tag{2.32}$$

Thus, there are about 411 CMB photons in every cubic centimeter of the universe at the present day. The mean energy of CMB photons, however, is quite low, only

$$E_{\mathrm{mean}} = 6.34 \times 10^{-4}\,\mathrm{eV}. \tag{2.33}$$

This is too low in energy to photoionize an atom, much less photodissociate a nucleus. About all they do, from a terrestrial point of view, is cause static on

[8] Using this definition, an alternate name for dark matter might be "transparent matter" or "invisible matter." However, the name "dark matter" has received the sanction of history.

television. The mean CMB photon energy corresponds to a wavelength of 2 millimeters, in the microwave region of the electromagnetic spectrum—hence the name "Cosmic *Microwave* Background."

The existence of the CMB is a very important cosmological clue. In particular, it is the clue that caused the Big Bang model for the universe to be favored over the Steady State model. In a Steady State universe, the existence of blackbody radiation at 2.725 K is not easily explained. In a Big Bang universe, however, a cosmic background radiation arises naturally if the universe was initially very hot as well as very dense. If mass is conserved in an expanding universe, then in the past the universe was denser than it is now. Assume that the early dense universe was very hot ($T \gg 10^4$ K, or $kT \gg 1$ eV). At such high temperatures, the baryonic matter in the universe was completely ionized, and the free electrons rendered the universe opaque. A dense, hot, opaque body, as described in Section 2.4, produces blackbody radiation. So, the early hot dense universe was full of photons, banging off the electrons like balls in a pinball machine, with a spectrum typical of a blackbody (equation (2.25)). However, as the universe expanded, it cooled. When the temperature dropped to ~ 3000 K, ions and electrons combined to form neutral atoms. When the universe no longer contained a significant number of free electrons, the blackbody photons started streaming freely through the universe, without further scattering off free electrons.

The blackbody radiation that fills the universe today can be explained as a relic of the time when the universe was sufficiently hot and dense to be opaque. However, at the time the universe became transparent, its temperature was ~ 3000 K. The temperature of the CMB today is 2.725 K, a factor of 1100 lower. The drop in temperature of the blackbody radiation is a direct consequence of the expansion of the universe. Consider a region of volume V that expands at the same rate as the universe, so that $V \propto a(t)^3$. The blackbody radiation in the volume can be thought of as a photon gas with energy density $\varepsilon_\gamma = \alpha T^4$. Moreover, since the photons in the volume have momentum as well as energy, the photon gas has a pressure; the pressure of a photon gas is $P_\gamma = \varepsilon_\gamma / 3$. The photon gas within our imaginary box must follow the laws of thermodynamics; in particular, the boxful of photons must obey the first law

$$dQ = dE + PdV, \tag{2.34}$$

where dQ is the amount of heat flowing into or out of the photon gas in the volume V, dE is the change in the internal energy, P is the pressure, and dV is the change in volume of the box. Since, in a homogeneous universe, there is no net flow of heat (everything is the same temperature, after all), $dQ = 0$. Thus, the first law of thermodynamics, applied to an expanding homogeneous universe, is

$$\frac{dE}{dt} = -P(t)\frac{dV}{dt}. \tag{2.35}$$

Since, for the photons of the CMB, $E = \varepsilon_\gamma V = \alpha T^4 V$ and $P = P_\gamma = \alpha T^4/3$, equation (2.35) can be rewritten in the form

$$\alpha\left(4T^3\frac{dT}{dt}V + T^4\frac{dV}{dt}\right) = -\frac{1}{3}\alpha T^4\frac{dV}{dt}, \tag{2.36}$$

or

$$\frac{1}{T}\frac{dT}{dt} = -\frac{1}{3V}\frac{dV}{dt}. \tag{2.37}$$

However, since $V \propto a(t)^3$ as the box expands, this means that the rate in change of the photons' temperature is related to the rate of expansion of the universe by the relation

$$\frac{d}{dt}(\ln T) = -\frac{d}{dt}(\ln a). \tag{2.38}$$

This implies the simple relation $T(t) \propto a(t)^{-1}$; the temperature of the cosmic background radiation has dropped by a factor of 1100 since the universe became transparent, because the scale factor $a(t)$ has increased by a factor of 1100 since then. What we now see as a Cosmic Microwave Background was once, at the time the universe became transparent, a Cosmic *Near-Infrared* Background, with a temperature slightly cooler than the surface of the star Betelgeuse.

The evidence cited so far can all be explained within the framework of a *Hot Big Bang* model, in which the universe was originally very hot and very dense, and since then has been expanding and cooling. The remainder of this book will be devoted to working out the details of the Hot Big Bang model that best fits the universe in which we live.

SUGGESTED READING

Full references are given in the Annotated Bibliography on page 235.

Bernstein (1995): The "Micropedia" that begins this text is a useful overview of the contents of the universe and the forces which work on them

Harrison (1987): The definitive treatment of Olbers' paradox

PROBLEMS

2.1. Suppose that in Sherwood Forest, the average radius of a tree is $R = 1\,\text{m}$ and the average number of trees per unit area is $\Sigma = 0.005\,\text{m}^{-2}$. If Robin Hood shoots an arrow in a random direction, how far, on average, will it travel before it strikes a tree?

2.2. Suppose you are in an infinitely large, infinitely old universe in which the average density of stars is $n_\star = 10^9\,\text{Mpc}^{-3}$ and the average stellar radius is equal to the Sun's radius: $R_\star = R_\odot = 7 \times 10^8\,\text{m}$. How far, on average, could you see in any direction before your line of sight struck a star? (Assume standard Euclidean geometry holds true in this universe.) If the stars are clumped into galaxies with a density $n_g =$

1 Mpc^{-3} and average radius $R_g = 2000\,\text{pc}$, how far, on average, could you see in any direction before your line of sight hit a galaxy?

2.3. Since you are made mostly of water, you are very efficient at absorbing microwave photons. If you were in intergalactic space, approximately how many CMB photons would you absorb per second? (If you like, you may assume you are spherical.) What is the approximate rate, in watts, at which you would absorb radiative energy from the CMB? Ignoring other energy inputs and outputs, how long would it take the CMB to raise your temperature by one nanoKelvin (10^{-9} K)? (You may assume your heat capacity is the same as pure water, $C = 4200\,\text{J kg}^{-1}\,\text{K}^{-1}$.)

2.4. Suppose that the difference between the square of the mass of the electron neutrino and that of the muon neutrino has the value $[m(\nu_\mu)^2 - m(\nu_e)^2]c^4 = 5 \times 10^{-5}\,\text{eV}^2$, and that the difference between the square of the mass of the muon neutrino and that of the tau neutrino has the value $[m(\nu_\tau)^2 - m(\nu_\mu)^2]c^4 = 3 \times 10^{-3}\,\text{eV}^2$. (This is consistent with the observational results discussed in section 2.4.) What values of $m(\nu_e)$, $m(\nu_\mu)$, and $m(\nu_\tau)$ minimize the sum $m(\nu_e) + m(\nu_\mu) + m(\nu_\tau)$, given these constraints?

2.5. A hypothesis once used to explain the Hubble relation is the "tired light hypothesis." The tired light hypothesis states that the universe is not expanding, but that photons simply lose energy as they move through space (by some unexplained means), with the energy loss per unit distance being given by the law

$$\frac{dE}{dr} = -KE, \tag{2.39}$$

where K is a constant. Show that this hypothesis gives a distance-redshift relation that is linear in the limit $z \ll 1$. What must the value of K be in order to yield a Hubble constant of $H_0 = 70\,\text{km s}^{-1}\,\text{Mpc}^{-1}$?