24 History of the Universe

Knowing how the scale factor $a(t)$ grew in the past and predicting how it will change in the future is an important goal of cosmologists. The Friedmann equation tells us that the growth of the scale factor is related to the energy density of the universe. It is useful to divide the energy content into radiation (relativistic particles), matter (nonrelativistic particles), and a cosmological constant. This is because each of these components has an energy density with a different dependence on the scale factor.

A cosmological constant has an energy density u_Λ that is constant with time. To see how the energy density of radiation and matter behaves as the universe expands, consider a volume V that expands with the universe, so that $V(t) \propto a(t)^3$. If particles are neither created nor destroyed, then the number density of particles, n, is diluted by the expansion of the universe at the rate $n(t) \propto V(t)^{-1} \propto a(t)^{-3}$, as illustrated in Figure 24.1. The energy of the nonrelativistic particles is contributed entirely by their rest mass, $\varepsilon = mc^2$, which remains constant as the universe expands. Thus, for nonrelativistic particles, a.k.a. "matter," the energy density has the dependence

$$u_m(t) = n(t)\varepsilon = n(t)mc^2 \propto a(t)^{-3}. \tag{24.1}$$

The energy of relativistic particles, such as photons, has the dependence $\varepsilon(t) = hc/\lambda(t) \propto a(t)^{-1}$. Thus, for relativistic particles, a.k.a. "radiation," the energy density has the dependence

$$u_r(t) = n(t)\varepsilon(t) = n(t)hc/\lambda(t) \propto a(t)^{-4}. \tag{24.2}$$

Given the different rates of decrease for the energy density, we find that the total energy density u was contributed mainly by radiation at early times, when $a \ll 1$ (Figure 24.2). In the language of cosmologists, the early universe was "radiation dominated." If the universe has a positive cosmological constant Λ, then it becomes "lambda dominated" if it reaches a sufficiently large scale factor.

24.1 ▪ THE CONSENSUS MODEL

In recent years, cosmologists (ordinarily a contentious bunch) have found themselves approaching an approximate consensus on the curvature, contents, and age of the universe. The curvature is flat (or nearly so), implying that the energy density today is close to the

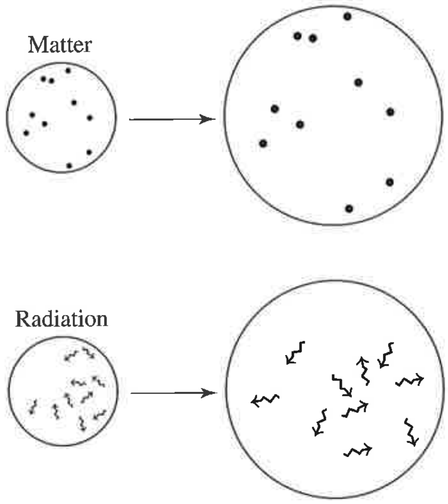

FIGURE 24.1 Dilution of nonrelativistic particles ("matter") and relativistic particles ("radiation").

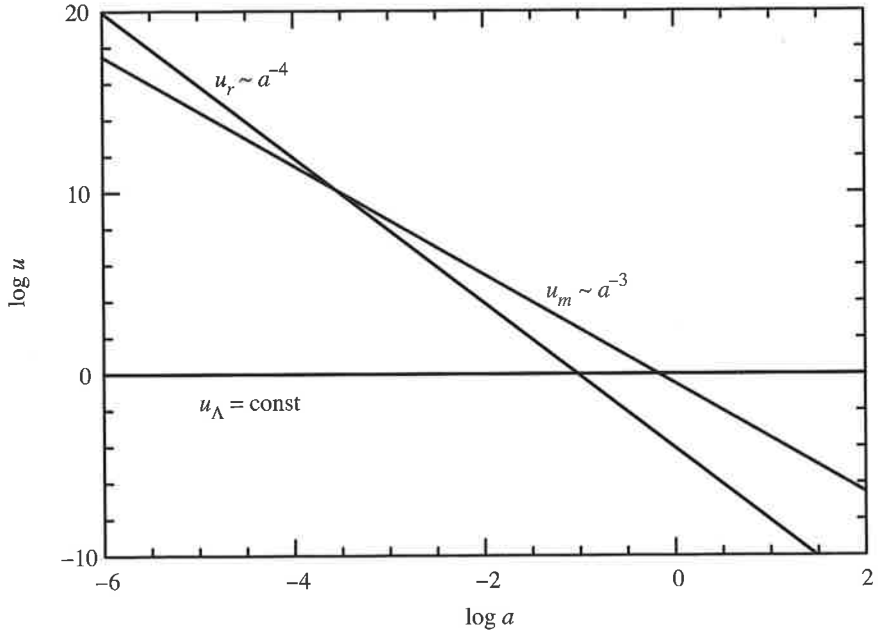

FIGURE 24.2 Dependence of energy density u on the scale factor $a(t)$.

critical density $u_0 \approx u_{c,0} \approx 5200$ MeV m^{-3}. To see how this energy density is allocated among the different components, let's do a census of the universe.

Most of the energy density of *photons* is provided by the cosmic microwave background; although stars have been shining away for ~ 13 Gyr, starlight still provides less than 10% of the total photon energy of the universe.[1] The current energy density of the CMB, as computed in equation (23.18), is $u_{\text{cmb},0} = 0.260$ MeV m^{-3}. The contribution of the CMB to the critical density is thus

$$\Omega_{\text{cmb},0} = \frac{u_{\text{cmb},0}}{u_{c,0}} = \frac{0.260 \text{ MeV m}^{-3}}{5200 \text{ MeV m}^{-3}} = 5.0 \times 10^{-5}. \quad (24.3)$$

The CMB is a relic of the time when the universe was hot and dense enough to be opaque to photons. If we extrapolate to earlier times and smaller scale factors, we reach a time when the universe was hot and dense enough to be opaque to *neutrinos*. Thus, there should be a cosmic neutrino background (CNB) analogous to the CMB. A detailed statistical mechanics calculation (of which we omit the details) reveals that the energy density of the CNB should be

$$u_{\nu,0} = 0.68 u_{\text{cmb},0} = 0.177 \text{ MeV m}^{-3}, \quad (24.4)$$

if the neutrinos are still traveling fast enough to qualify as relativistic particles today (that is, if the energy per particle, ε_ν, is much larger than the rest energy $m_\nu c^2$). The CNB has not yet been detected. The energy per neutrino is comparable to the energy per photon in the CMB—less than 10^{-3} eV. Detecting such low-energy neutrinos is not yet technically feasible.

If neutrinos are low enough in mass to be relativistic today, the present density parameter in radiation is

$$\Omega_{r,0} = \Omega_{\text{cmb},0} + \Omega_{\nu,0} = 1.68\Omega_{\text{cmb},0} = 8.4 \times 10^{-5}. \quad (24.5)$$

Thus, photons and neutrinos contribute a small fraction of the critical density today; about 1 part in 12,000. Most of the density must currently be provided by nonrelativistic matter and/or a cosmological constant (or by some other, hitherto unsuspected, component).

The energy density of the CMB has been measured with high precision; the energy density of the CNB has been computed using well-understood principles of physics. The energy density of matter is not as well determined. If we add together the mass of all the clusters of galaxies in our neighborhood, we find that the density of clustered matter is

$$\Omega_{\text{cluster},0} \approx 0.2. \quad (24.6)$$

This number doesn't include any smoothly distributed matter in the intercluster voids. The best estimate for the current density of nonrelativistic matter, using all available data, is

$$\Omega_{m,0} \approx 0.3. \quad (24.7)$$

[1] Thus, although photon number is not strictly conserved, as we assumed when computing $u_r \propto a^{-4}$, it's a better approximation than we might guess.

The best estimate for the current density of *baryonic* matter (that is, stuff made of protons, neutrons, and electrons) is

$$\Omega_{\text{bary},0} = 0.04. \tag{24.8}$$

(We see how this number is determined in Section 24.3.) The majority of the matter in the universe must consist of nonbaryonic dark matter, such as WIMPs.

The total mass density of baryonic matter today is

$$\rho_{\text{bary},0} = 0.04\rho_{c,0} = 3.7 \times 10^{-28}\ \text{kg m}^{-3}. \tag{24.9}$$

The number density of baryons is thus $n_{\text{bary},0} \approx \rho_{\text{bary},0}/m_p \approx 0.22\ \text{m}^{-3}$. This is much lower than the number density of photons. The photon-to-baryon ratio in the universe is approximately

$$\frac{n_{\text{cmb},0}}{n_{\text{bary},0}} \approx \frac{4.11 \times 10^8\ \text{m}^{-3}}{0.22\ \text{m}^{-3}} \approx 2 \times 10^9. \tag{24.10}$$

Baryons are badly outnumbered by photons, by a ratio of 2 billion to 1.

The available observational evidence has led cosmologists to a **Consensus Model of the Universe**. This model is flat and contains radiation, nonrelativistic matter, and a cosmological constant (a.k.a. Λ or "lambda," a.k.a. "vacuum energy," a.k.a. "dark energy"). Some of the current properties of the Consensus Model are listed in Table 24.1.

For the Consensus Model, with its mix of radiation, matter, and cosmological constant, the Friedmann equation (eq. 23.71) is

$$H(t)^2 = \frac{8\pi G}{3c^2}\left[\frac{u_{r,0}}{a(t)^4} + \frac{u_{m,0}}{a(t)^3} + u_\Lambda\right], \tag{24.11}$$

where $H(t) \equiv \dot{a}/a$. Dividing by H_0^2, and using the definition of the critical density (equation 23.73), we find that

$$\frac{H(t)^2}{H_0^2} = \frac{1}{u_{c,0}}\left[\frac{u_{r,0}}{a(t)^4} + \frac{u_{m,0}}{a(t)^3} + u_\Lambda\right], \tag{24.12}$$

TABLE 24.1 The Consensus Model

Component	Property
photons	$\Omega_{\gamma,0} = 5.0 \times 10^{-5}$
neutrinos	$\Omega_{\nu,0} = 3.4 \times 10^{-5}$
total radiation	$\Omega_{r,0} = 8.4 \times 10^{-5}$
baryonic matter	$\Omega_{\text{bary},0} = 0.04$
nonbaryonic dark matter	$\Omega_{\text{dm},0} = 0.26$
total matter	$\Omega_{m,0} = 0.30$
cosmological constant	$\Omega_{\Lambda,0} \approx 0.70$

or, in terms of the dimensionless density parameter Ω (equation 23.76),

$$\frac{H(t)^2}{H_0^2} = \frac{\Omega_{r,0}}{a(t)^4} + \frac{\Omega_{m,0}}{a(t)^3} + \Omega_\Lambda. \tag{24.13}$$

The Friedmann equation thus provides us with a differential equation for the scale factor $a(t)$:

$$\frac{da}{dt} = H_0 \left[\frac{\Omega_{r,0}}{a(t)^2} + \frac{\Omega_{m,0}}{a(t)} + \Omega_{\Lambda,0} a(t)^2 \right]^{1/2}. \tag{24.14}$$

Given values for H_0, $\Omega_{r,0}$, $\Omega_{m,0}$, and $\Omega_{\Lambda,0}$, equation (24.14) can be integrated to yield the scale factor as a function of time, given our usual normalization $a(t_0) \equiv 1$. Unfortunately, the solution of equation (24.14) doesn't have a simple analytic form. However, since the right-hand side of equation (24.14) is always positive for the Consensus Model, we can immediately predict that the universe will continue to expand forever. There is no Big Crunch for the Consensus Model.

Since the three components (radiation, matter, and Λ) have different dependences on scale factor, there will be long stretches in the history of the universe when one component dominates the energy density. At the moment, $u_\Lambda > u_{m,0} \gg u_{r,0}$. At an earlier time, and a smaller scale factor $a_{m\Lambda}$, the density of matter u_m was equal to u_Λ. This equality took place when

$$u_\Lambda = \frac{u_{m,0}}{a_{m\Lambda}^3}, \tag{24.15}$$

or

$$a_{m\Lambda} = \left(\frac{u_{m,0}}{u_\Lambda}\right)^{1/3} = \left(\frac{\Omega_{m,0}}{\Omega_{\Lambda,0}}\right)^{1/3} = \left(\frac{0.7}{0.3}\right)^{1/3} = 0.75. \tag{24.16}$$

When we observe a galaxy with redshift $z = 1/a_{m\Lambda} - 1 = 0.33$, we are looking back to a time when matter was equal in density to the cosmological constant.

If we go to earlier times, there was a scale factor a_{rm} at which the density of radiation u_r was equal to the density of matter u_m. This equality took place when

$$\frac{u_{m,0}}{a_{rm}^3} = \frac{u_{r,0}}{a_{rm}^4}, \tag{24.17}$$

or

$$a_{rm} = \frac{u_{r,0}}{u_{m,0}} = \frac{\Omega_{r,0}}{\Omega_{m,0}} = \frac{8.4 \times 10^{-5}}{0.3} = 2.8 \times 10^{-4}. \tag{24.18}$$

This scale factor corresponds to a redshift $z = 1/a_{rm} - 1 = 3600$. This is higher than the redshift at which the universe became transparent ($z \approx 1100$), so we cannot directly see the time of radiation–matter equality.

Early in the history of the universe, when the scale factor was small ($a \ll a_{rm} \approx 0.00028$), the universe was **radiation-dominated**. That is, the vast majority of the density was provided by photons and highly relativistic particles such as neutrinos. The Friedmann equation (eq. 24.14) in a radiation-dominated universe reduces to the form

$$\frac{da}{dt} = \frac{\Omega_{r,0}^{1/2} H_0}{a(t)}. \tag{24.19}$$

This equation has the solution

$$a(t) = [2\Omega_{r,0}^{1/2} H_0 t]^{1/2}, \tag{24.20}$$

as the reader can verify by substitution. Since $a(t) \propto t^{1/2}$ in the early universe, the horizon size

$$\ell_{\rm hor}(t_0) = \int_0^{t_0} \frac{dt}{a(t)} \tag{24.21}$$

does not diverge as $t \to 0$, and we live in a universe with a finite horizon. The acceleration in the early universe was negative:

$$\ddot{a} = -\frac{1}{4t^2} a(t) < 0, \tag{24.22}$$

indicating that the expansion of the early universe was slowed by gravity acting on photons and other relativistic particles.

At intermediate scale factors, when $a_{rm} \ll a \ll a_{m\Lambda}$, or $0.00028 \ll a \ll 0.75$, the universe was **matter-dominated**. That is, the majority of the density was provided by nonrelativistic particles, such as WIMPs, protons, and neutrons. During the matter-dominated era, the Friedmann equation (eq. 24.14) takes the simplified form

$$\frac{da}{dt} = \frac{\Omega_{m,0}^{1/2} H_0}{a(t)^{1/2}}, \tag{24.23}$$

which has the solution

$$a(t) = [\frac{3}{2}\Omega_{m,0}^{1/2} H_0 t]^{2/3}, \tag{24.24}$$

again verifiable by substitution. Since $a(t) \propto t^{2/3}$ during the matter-dominated era, the acceleration was

$$\ddot{a} = -\frac{2}{9t^2} a(t) < 0. \tag{24.25}$$

A matter-dominated universe, like a radiation-dominated universe, is decelerating.

In the future, when the scale factor becomes large ($a \gg a_{m\Lambda} \approx 0.75$), the universe will become **lambda-dominated**. When the cosmological constant is the only significant contributor to the energy density, the Friedmann equation (eq. 24.14) takes the form

$$\frac{da}{dt} = \Omega_{\Lambda,0}^{1/2} H_0 a(t). \tag{24.26}$$

This equation has an exponential solution:

$$a(t) \propto e^{Kt}, \tag{24.27}$$

where

$$K = \Omega_{\Lambda,0}^{1/2} H_0 = \frac{1}{16.7\ \text{Gyr}}. \tag{24.28}$$

When the cosmological constant takes over, the universe will expand exponentially, with an e-folding time of 16.7 Gyr. In the lambda-dominated universe, the Hubble parameter will be

$$H = \frac{\dot{a}}{a} = K = \Omega_{\Lambda,0}^{1/2} H_0. \tag{24.29}$$

The Hubble constant really will be constant with time. In addition, when the universe is lambda-dominated, the acceleration will be *positive*:

$$\ddot{a} = \Omega_{\Lambda,0} H_0^2 a(t) > 0. \tag{24.30}$$

In a relativistic universe, a cosmological constant $\Lambda > 0$ plays the role of a *repulsive* force in a Newtonian universe; that is, it causes the relative speed of any two points to increase with time.[2]

The Friedmann equation for the Consensus Model can be integrated numerically to find $a(t)$ for all times, not just those special epochs when a single component is dominant. The resulting scale factor is shown in Figure 24.3. Note that the transitions from radiation to matter domination, and from matter to lambda domination, are smooth and gradual.

With a complete knowledge of $a(t)$, the time corresponding to any scale factor can be computed. The scale factor of radiation–matter equality, $a_{rm} = 0.00028$, corresponds to a time

$$t_{rm} = 3.3 \times 10^{-6} H_0^{-1} = 47{,}000\ \text{yr}. \tag{24.31}$$

Despite the brevity of the radiation-dominated era, a lot of interesting physics was going on back then, and cosmologists have focused a great deal of attention on it. The scale factor of matter–lambda equality, $a_{m\Lambda} = 0.75$, corresponds to a time

$$t_{m\Lambda} 0.70 H_0^{-1} = 9.8\ \text{Gyr}. \tag{24.32}$$

This should be compared to the current age of the universe in the Consensus Model, which turns out to be

$$t_0 = 0.964 H_0^{-1} = 13.5\ \text{Gyr}. \tag{24.33}$$

[2] It can be shown that if a component of the universe has an energy density $u \propto a^{-n}$, then if $n < 2$, it will cause $\ddot{a} > 0$. In general, components that cause $\ddot{a} > 0$ are given the generic name of "dark energy." The "cosmological constant" is a special case of "dark energy," with $n = 0$.

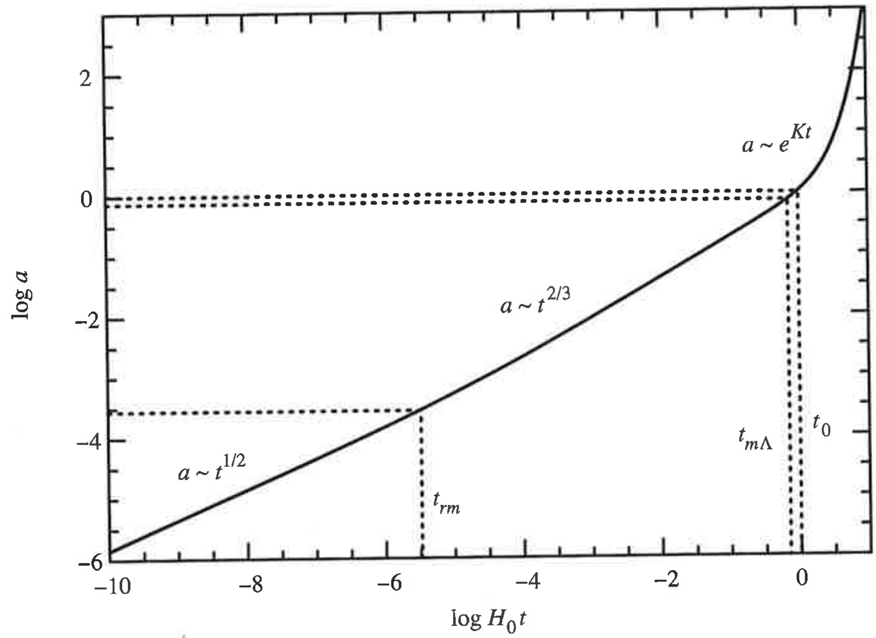

FIGURE 24.3 Scale factor as a function of time (in units of the Hubble time) for the Consensus Model.

What's amusing is that the period of deceleration (when the universe was dominated by radiation and matter) was almost exactly balanced by the later period of positive acceleration (when the universe was dominated by Λ). The net effect is that the age of the universe is nearly equal to H_0^{-1}, the naïve result you would get by assuming no acceleration at all.

With $a(t)$ known, other properties of the Consensus Model can be computed. For instance, Figure 24.4 shows the current proper distance, $\ell_p(t_0)$, to a galaxy with measured redshift z. The bold, solid line shows the results for the Consensus Model. For comparison, the dotted line shows the proper distance in a flat, matter-only universe and the dot-dash line shows the proper distance in a flat, lambda-only universe. As $z \to \infty$, the proper distance in the Consensus Model reaches a limiting value, $\ell_p(t_0) \to 3.24c/H_0$. Thus, the Consensus Model has a finite horizon distance,

$$\ell_{\rm hor}(t_0) = 3.24\frac{c}{H_0} = 3.12ct_0 = 14{,}000 \text{ Mpc}. \tag{24.34}$$

In the matter-only universe, the horizon distance is $\ell_{\rm hor} = 2c/H_0$; in the lambda-only universe, the horizon distance is infinite.

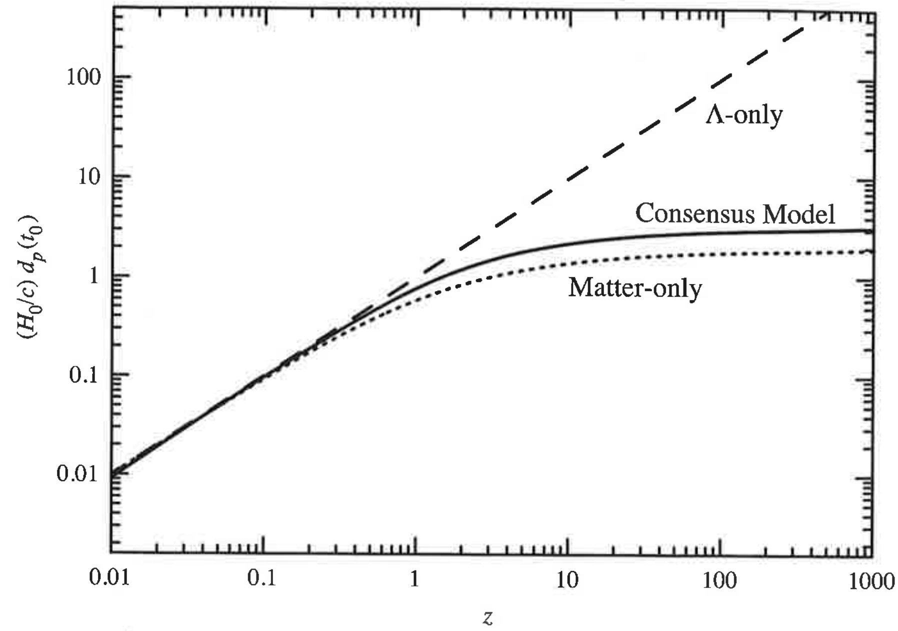

FIGURE 24.4 Current proper distance (in units of the Hubble distance) to a galaxy with redshift z.

24.2 ▪ THE ACCELERATING UNIVERSE

The Friedmann equation for the Consensus Model can be written in the form

$$\dot{a} = H_0 \left[\frac{\Omega_{r,0}}{a^2} + \frac{\Omega_{m,0}}{a} + \Omega_{\Lambda,0} a^2 \right]^{1/2} . \tag{24.35}$$

By taking the derivative with respect to t, then doing a bit of algebra, we find an equation for the second time-derivative of the scale factor:

$$\ddot{a} = H_0^2 \left[-\frac{\Omega_{r,0}}{a^3} - \frac{\Omega_{m,0}}{2a^2} + \Omega_{\Lambda,0} a \right] . \tag{24.36}$$

Note that on the right-hand side of equation (24.36), the terms involving radiation and matter are negative (they slow down the expansion), while the term involving Λ is positive (it speeds up the expansion). At present, $a(t_0) = 1$, so the acceleration of the expansion is

$$\ddot{a}_0 = H_0^2[-\Omega_{r,0} - \Omega_{m,0}/2 + \Omega_{\Lambda,0}] = 0.55 H_0^2 \tag{24.37}$$

for the Consensus Model. The speeding up of the expansion is a remarkable—and in the context of Newtonian gravity, counterintuitive—result. What led cosmologists to embrace the accelerating universe? The conclusion was based largely on measuring the flux of standard candles at high redshifts.

Suppose we are looking at a standard candle, of known luminosity L, whose current proper distance is $r = \ell_p(t_0)$. In a static, flat universe, the observed flux would be given by an inverse square law:

$$F_{\text{static}} = \frac{L}{4\pi r^2}. \tag{24.38}$$

If the universe is expanding rather than static, the observed flux of the standard candle will be *lower* than this value, for two reasons.

First, the expansion of the universe causes the energy of each photon from the standard candle to decrease. The photon begins with an energy ε_e when it is emitted at time t_e. By the time we observe the photon at time t_0, its energy will have dropped to the value

$$\varepsilon_0 = \varepsilon_e \frac{a(t_e)}{a(t_0)} = \frac{\varepsilon_e}{1+z}, \tag{24.39}$$

where z is the measured redshift of the standard candle.

Second, the expansion of the universe will cause the time between photon detections to increase. If two photons are emitted in the same direction separated by a time interval δt_e, the proper distance between them will initially be $\delta r_e = c(\delta t_e)$. However, by the time we detect the two photons at the later time t_0, the proper distance between them will be stretched to $\delta r_0 = c(\delta t_e)(1+z)$, and we will detect them separated by a time interval $\delta t_0 = \delta t_e(1+z)$.

The net result of these two effects—lower energy photons and a longer time interval between photons—is that the observed flux f in an expanding (but spatially flat) universe will be

$$F_{\text{expand}} = \frac{L}{4\pi r^2 (1+z)^2}. \tag{24.40}$$

Converting from fluxes to apparent magnitudes, we can also write down the observed apparent magnitude m in an expanding universe:

$$m = M + 5\log[r(1+z)] - 5, \tag{24.41}$$

where r is in parsecs. The distance modulus for a standard candle in an expanding (but spatially flat) universe is thus

$$m - M = 5\log r + 5\log(1+z) - 5, \tag{24.42}$$

where r is the current proper distance $\ell_p(t_0)$ to the standard candle. Consider the proper distances shown in Figure 24.4 for three different flat universes. Since the exponentially expanding, lambda-only universe has the largest proper distance r for a given redshift z, it will have the *faintest* standard candles at that redshift. Figure 24.5 shows the distance

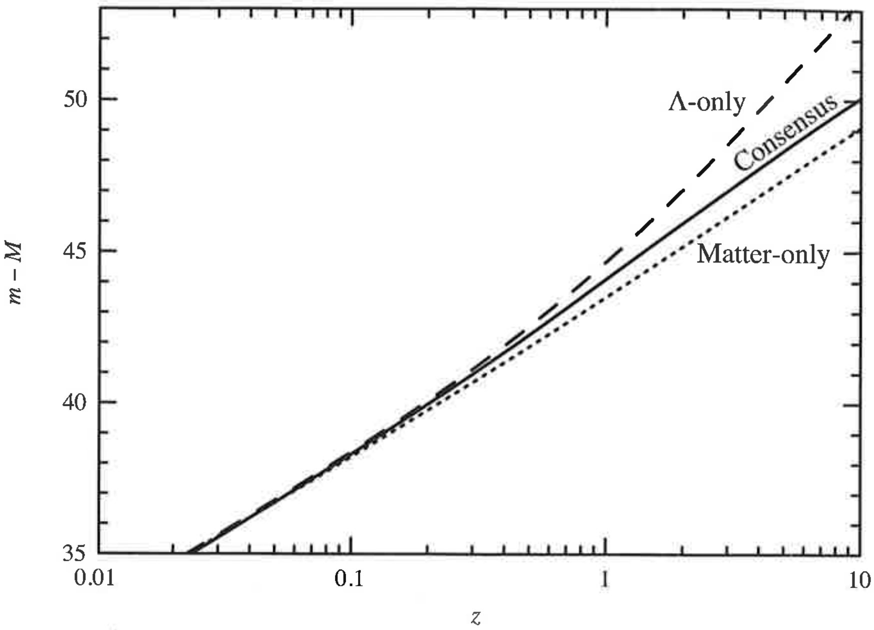

FIGURE 24.5 Distance modulus for an object with redshift z. The solid line represents the Consensus Model; the dotted line, a flat, matter-only universe; and the dot-dash line, a flat, lambda-only universe.

modulus for standard candles in our three different flat universes. At a very small redshift ($z \ll 1$), the distance modulus reduces to

$$m - M \approx 5 \log \left(\frac{c}{H_0} z \right) - 5 \approx 43.17 - 5 \log z, \tag{24.43}$$

regardless of the values of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$. It is only at $z > 1$ that the differences between models becomes large.

As an example, consider a type Ia supernova with an absolute magnitude $M = -20.0$ mag. If it is seen at a redshift $z = 1$, then its apparent magnitude in the Consensus Model will be $m = 24.1$ mag. Its apparent magnitude in the flat, lambda-only model will be $m = 24.7$ mag, 0.6 mag fainter than in the Consensus Model. Its apparent magnitude in the flat, matter-only model will be $m = 23.5$ mag, 0.6 mag brighter than in the Consensus Model.

Using the apparent magnitude of distant type Ia supernovae to distinguish among different models requires accurate photometry of apparently faint sources. It's difficult, but it can be done. Figure 24.6 shows the results from two different surveys of type Ia supernovae. The observational results are compared to three different models. In Figure 24.6a, the top line is the result expected in the Consensus Model; the bottom

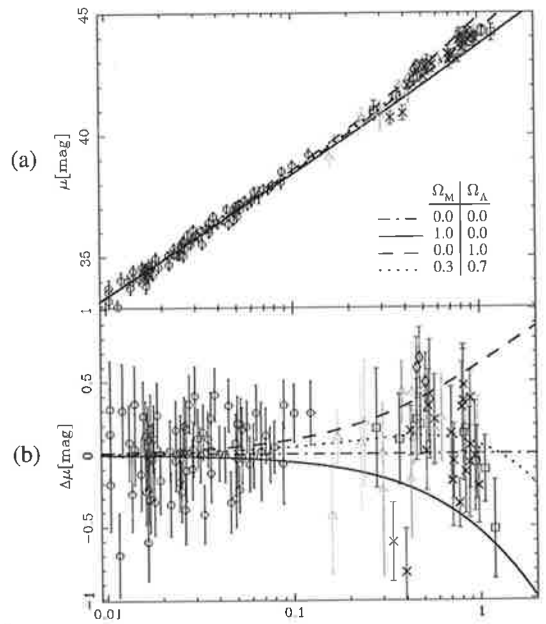

FIGURE 24.6 (a) Distance modulus versus redshift for type Ia supernovae. (b) Difference between the data and the predictions for an empty ($\Omega = 0$) universe.

line is the result for a flat, matter-only universe; and the middle line is for a negatively curved universe with $\Omega_{m,0} = 0.3$ and $\Lambda = 0$. The data are best fitted by the Consensus Model; this is better seen in Figure 24.6b, which shows the difference between the data and the predictions of the negatively curved $\Omega_{m,0} = 0.3$ model.

Instead of fitting just three models to the supernova data, we can ask more generally, What values of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ give the best fits to the data.[3] After choosing values for $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$, the relation between distance modulus and redshift can be computed, then compared to the supernova data. The results of fitting the model universes are shown in Figure 24.7. This is a rather busy plot that repays careful scrutiny. Since the radiation density is negligible, the criterion for flatness is $\Omega_m + \Omega_\Lambda = 1$, represented in Figure 24.7 by the dashed line running diagonally downward from left to right. Positively curved universes (labeled "Closed") lie above and to the right; negatively curved universes (labeled "Open") lie below and to the left. The solid line that runs diagonally upward from left to right divides universes with $\ddot{a}_0 > 0$ (labeled "Acceleration") from universes with $\ddot{a}_0 < 0$ (labeled "Deceleration"). Finally, the slightly curved line that runs nearly horizontally from ($\Omega_m = 0$, $\Omega_\Lambda = 0$) divides the Big Chill

[3] At $z < 1$, the role of radiation is negligible, so the supernova fluxes tell us nothing about the density of radiation, $\Omega_{r,0}$.

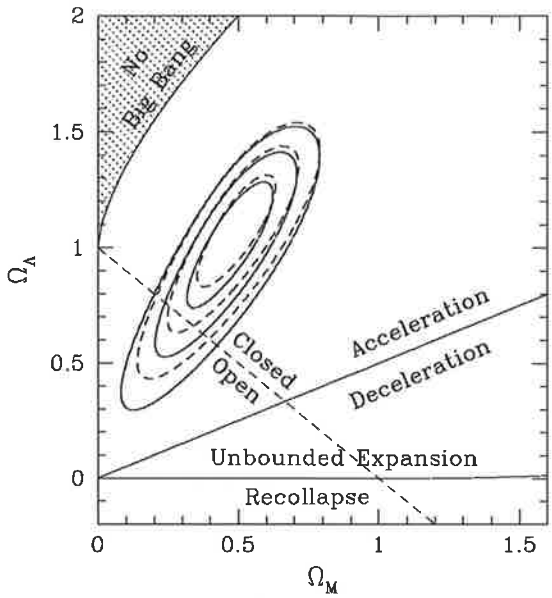

FIGURE 24.7 The values of $\Omega_{m,0}$ and $\Omega_{\Lambda,0}$ that best fit the relation between $m - M$ and z for type Ia supernovae. The solid and dashed lines represent two slightly different samples of supernovae.

universes (labeled "Unbounded Expansion") from the Big Crunch universes (labeled "Recollapse").[4]

The concentric ovals in Figure 24.7 show the region of parameter space that gives the best fit to the available supernova data. (The smallest, innermost oval gives the best fit, but the largest, outermost oval cannot be excluded at the 99.5% confidence level.) Decelerating universes can be strongly ruled out by the supernova data, as can Big Crunch universes. It is the supernova data that have led cosmologists to conclude that we live in a universe whose expansion is accelerating, leading to an exponentially chilly future for our universe.

Notice, however, that the supernova data cannot by themselves distinguish between positively curved, flat, or negatively curved universes. The curvature of the universe is constrained by looking at the angular size of distant objects, as outlined in Section 23.3. The most distant things we can see in the universe are hot and cold spots in the CMB. The angular size of these spots has been measured by the *Wilkinson Microwave Anisotropy Probe (WMAP)* and by ground-based and balloon-borne experiments. It is the preferred

[4] Your curiosity may be piqued by the wedge labeled "No Big Bang" in the upper left corner. These models, when extrapolated backward in time, have $\dot{a} = 0$ when $a > 0$; that is, they started their expansion in a state where the density was low compared to the extraordinarily high initial density we expect in a true Big Bang universe.

angular size of the structure in the CMB that provides the best evidence for the flatness of the universe ($\Omega \approx 1$). It is only when we combine the CMB results (the universe is flat) with the supernova results (the universe is accelerating) that we reach the Consensus Model, with $\Omega_m \approx 0.3$ and $\Omega_\Lambda \approx 0.7$.

If the cosmological constant is truly constant with time, then we face an accelerating future. The Local Supercluster will remain gravitationally bound (we don't have to worry about the Virgo Cluster and the Local Group being yanked apart), but more distant superclusters will move away from us with exponentially increasing velocity.

24.3 ▪ THE EARLY UNIVERSE

To understand the origins of the universe, we want to look as far back in time as possible. The oldest photons we see today are the photons of the CMB. As described in Section 23.1, when the initially hot and dense universe became sufficiently cool, protons and electrons combined to form neutral hydrogen atoms:

$$p + e^- \to H + \gamma. \tag{24.44}$$

At this time, the universe became transparent, since the photons of the cosmic background radiation no longer scattered off free electrons.

As we look outward in space, we look backward in time. Thus, we (and every other observer in the universe) are surrounded by a spherical **last scattering surface**, illustrated in Figure 24.8. The last scattering surface is where photons underwent their last scattering from a free electron before streaming freely through the newly transparent universe. The last scattering surface is the surface of the glowing, opaque ionized gas that filled the early universe.[5]

The universe became transparent, and photons underwent their last scattering, at a temperature $T_{\rm ls} = 3000$ K. The scale factor at the time of last scattering was $a_{\rm ls} = T_0/T_{\rm ls} = 2.725\ {\rm K}/3000\ {\rm K} = 9.1 \times 10^{-4}$, corresponding to a redshift $z_{\rm ls} = 1/a_{\rm ls} - 1 = 1100$. In the Consensus Model, the time of last scattering was $t_{\rm ls} = 2.5 \times 10^{-5} H_0^{-1} = 0.4$ Myr. Thus, the CMB gives us a glimpse of what the universe was like 400,000 years after the Big Bang.

At every point of the sky, the CMB has a blackbody spectrum. Although the average temperature of the CMB is $T_0 = 2.725$ K, the actual temperature varies slightly across the celestial sphere. Color Figure 30 shows a plot of the temperature of the CMB, as derived from *WMAP* data. The temperatures show a *dipole* distortion, with one hemisphere of the sky being blueshifted to higher temperatures, and the other hemisphere being redshifted to lower temperatures. This dipole distortion is simply a Doppler shift, caused by the motion of *WMAP* through space.[6] Once we subtract away the orbital motion of *WMAP* about the Sun ($v \approx 30$ km s^{-1}), the orbital motion of the Sun about the Galactic Center

[5] We can think of it as an inside-out photosphere, since the photosphere of a star is also the surface of a glowing, opaque, ionized gas.

[6] *WMAP* is at the Earth's L_2 point (see Figure 11.3), and not in a low Earth orbit.

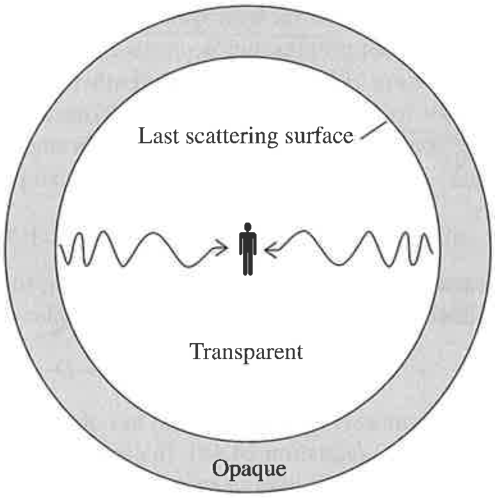

FIGURE 24.8 Observer surrounded by the "last scattering surface."

($v \approx 220$ km s^{-1}), and the orbital motion of the Galaxy relative to the center of mass of the Local Group ($v \approx 80$ km s^{-1}), we find that the Local Group is moving in the direction of Hydra, with a speed $v_{\rm lg} \approx 630$ km s^{-1}. Thus, the dipole distortion of the CMB is telling us about motion of the Local Group here and now (which is undeniably interesting but doesn't tell us directly about the early universe).

Color Figure 31 shows the remaining low-amplitude temperature fluctuations after the dipole Doppler distortion has been subtracted. The angular size of the hot and cold spots in this image are what cosmologists use to determine the curvature of space. The amplitude of the fluctuations is not large: typically, $\delta T/T \sim 10^{-5}$. The small temperature fluctuations result from small density fluctuations at the time of last scattering. A photon that happens to find itself in a dense region when the universe becomes transparent will lose energy as it climbs out of the gravitational potential well that is associated with the dense region, and will thus become redshifted to lower temperatures. Conversely, a photon that happens to be in a low-density region will be blueshifted to higher temperatures. The low-amplitude density fluctuations that were present at $t \approx 0.4$ Gyr have grown with time to the high-amplitude density fluctuations that we see at $t_0 \approx 13.5$ Gyr (superclusters, clusters, galaxies, etc.)

The opacity of the early universe draws a frustrating veil over the first 400 millennia of the history of the universe. Nevertheless, cosmologists can still deduce indirectly what was happening back then. For instance, we know that in the early universe, neutral hydrogen atoms couldn't exist because some of the cosmic background photons had energies larger than the hydrogen ionization energy ($\chi = 13.6$ eV). If we go farther back in time, we should reach a time at which bound atomic nuclei could exist because some of the cosmic background photons had energies larger than the nuclear binding energy (typically several MeV). Thus, just as there was a time when protons and electrons combined to form neutral hydrogen atoms (at $t \approx 0.4$ Myr), there must have been an earlier time when protons and neutrons combined to form atomic nuclei. This time is known as the era of **Big Bang nucleosynthesis (BBN)**.

Consider, for simplicity, a deuterium (D) nucleus. This is the simplest of all compound nuclei; it consists of a proton and neutron bound together with a binding energy $B = 2.22$ MeV. A gamma-ray photon with $\varepsilon > B$ can photodissociate deuterium:

$$D + \gamma \rightarrow p + n. \tag{24.45}$$

This reaction can run in the opposite direction, too; a proton and neutron can fuse to form a deuterium nucleus, with a gamma-ray photon carrying off the excess energy:

$$p + n \rightarrow D + \gamma. \tag{24.46}$$

Deuterium synthesis (equation 24.46) has obvious parallels to the radiative recombination of hydrogen (equation 24.44). In each case, two particles become bound together, with a photon carrying away excess energy. The most striking difference between the processes is the different energies involved. The photodissociation energy of deuterium is $B = 2.22\,\text{MeV} = (1.6 \times 10^5)(13.6\,\text{eV})$. The energy released when a deuterium nucleus is formed is 160,000 times the energy released when a neutral hydrogen atom is formed; thus, we expect the temperature at the time of nucleosynthesis to be 160,000 times greater than the temperature at the time of last scattering, when neutral hydrogen formed. This implies a nucleosynthesis temperature

$$T_{\text{nuc}} = \frac{B}{\chi} T_{\text{ls}} = (1.6 \times 10^5)(3000\,\text{K}) = 5 \times 10^8\,\text{K}. \tag{24.47}$$

In the Consensus Model, the universe had this temperature at an age $t_{\text{nuc}} \sim 400\,\text{s} \sim 7$ min.[7]

Once a significant amount of deuterium forms, it can be converted to heavier nuclei. For instance, tritium (^{3}H) is made by the reaction

$$D + n \rightarrow {}^3\text{H} + \gamma. \tag{24.48}$$

Light helium (^{3}He) is made by the reaction

$$D + p \rightarrow {}^3\text{He} + \gamma. \tag{24.49}$$

Ordinary helium (^{4}He) can be made by reactions such as

$${}^3\text{H} + p \rightarrow {}^4\text{He} + \gamma \tag{24.50}$$

and

$${}^3\text{He} + n \rightarrow {}^4\text{He} + \gamma. \tag{24.51}$$

[7] This is a slight overestimate of the age; when Steven Weinberg entitled his book on BBN *The First Three Minutes*, he was using a more accurate calculation.

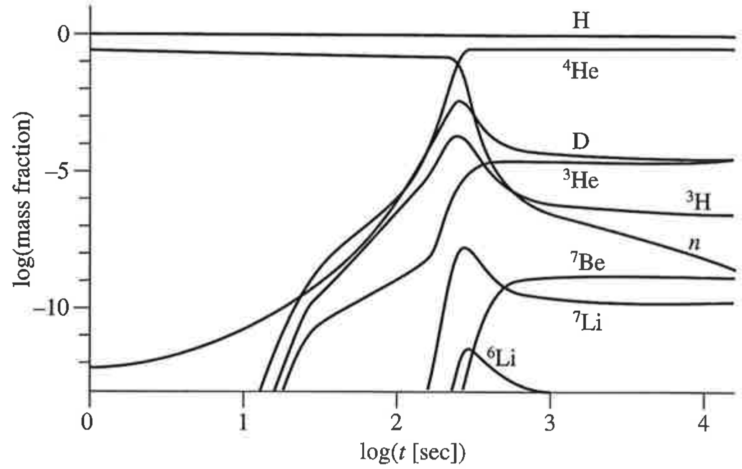

FIGURE 24.9 Mass fraction of nuclei (and free neutrons) during the time of BBN.

Once ^{4}He is reached, the orderly march of nucleosynthesis to larger atomic numbers hits a roadblock. There are no stable nuclei with atomic number 5. If we try to add a proton to ^{4}He to make ^{5}Li, it won't work; ^{5}Li is not a stable nucleus. If we try to add a neutron to ^{4}He to make ^{5}He, it won't work; ^{5}He is not a stable nucleus. We can make small amounts of lithium by the reactions

$$^4\text{He} + D \rightarrow {}^6\text{Li} + \gamma \tag{24.52}$$

and

$$^4\text{He} + {}^3\text{H} \rightarrow {}^7\text{Li} + \gamma, \tag{24.53}$$

but then we hit another roadblock. There are no stable nuclei with atomic number 8. If we try to fuse two ^{4}He nuclei together to form ^{8}Be, it won't work; ^{8}Be is not a stable nucleus.[8]

In summary, BBN works rapidly and efficiently up to ^{4}He, but few nuclei heavier than helium are produced. The precise yields of the different elements and isotopes can be computed using a computer code that takes into account the cross-sections for the different nuclear reactions. Results of a typical BBN code are shown in Figure 24.9. At $t \approx 1$ s, almost all the baryons are in the form of free protons (labeled H in the figure) and free neutrons (labeled n).

[8] As you may recall from Section 15.3, the instability of ^{8}Be is the main reason why the triple alpha process in stars requires such high temperature and density; a ^{4}He nucleus must be slammed into the ^{8}Be nucleus during the brief interval before it falls apart.

Because protons have a lower rest energy than neutrons, the laws of statistical mechanics state that protons will be more numerous than neutrons in the early universe. By $t \approx 100$ s, when nucleosynthesis kicks into high gear, there are seven protons for every neutron in the universe. Consider a representative group of two neutrons and 14 protons. The two neutrons swiftly combine with two of the protons to form a single ^{4}He nucleus, leaving 12 lonely protons left over.[9] At $t \approx 10^4$ s ≈ 3 hr, the temperature has dropped too low for further nuclear reactions and the epoch of BBN is over. At this point, the mass fraction of hydrogen is $X \approx 12/16 \approx 0.75$ and the mass fraction of helium is $Y \approx 4/16 \approx 0.25$. Only tiny amounts of elements other than ^{1}H and ^{4}He are present.

A basic prediction of BBN is that helium contributed 25% of the baryon density even before the first generation of stars began to pollute the universe with heavy elements. Observations of gas and stars reveal that hydrogen is invariably mixed with helium. The helium mass fraction of the Sun is $Y = 0.250$, but the Sun is contaminated by helium formed in earlier generations of stars. When we look at interstellar gas that hasn't been run through the stellar mill, the helium mass fraction can be as low as $Y = 0.24$, but not any lower. This is in good agreement with the predictions of BBN.

24.4 ▪ THE VERY EARLY UNIVERSE

So far we have accentuated the positive when discussing the Hot Big Bang universe in general, and the Consensus Model in particular. However, the standard Hot Big Bang scenario, in which the universe was dominated by radiation at early times, has a pair of problems that have puzzled cosmologists. These are the **flatness problem** and the **horizon problem**. Let's examine the flatness problem first.

The curvature of the universe is related to its energy content by the Friedmann equation,

$$H(t)^2 = \frac{8\pi G}{3c^2}u(t) - \frac{\kappa c^2}{r_{c,0}^2}\frac{1}{a(t)^2}. \tag{24.54}$$

If we divide each side by $H(t)^2$, we can rewrite the Friedmann equation in the form

$$1 = \Omega(t) - \frac{\kappa c^2}{r_{c,0}^2}\frac{1}{a(t)^2 H(t)^2}. \tag{24.55}$$

If the density parameter is exactly equal to 1, then the universe is perfectly flat. At the present moment, the observational results are consistent with the limits

$$|1 - \Omega_0| \leq 0.1. \tag{24.56}$$

[9] The solitary life of the protons ends 400,000 years later, when they find electron sidekicks and become neutral hydrogen atoms.

Why should the value of the density parameter be so close to 1 today? We might just shrug and say, "It's a coincidence." However, when you extrapolate the value of $\Omega(t)$ back into the past, the closeness of Ω to 1 becomes more difficult to dismiss as a coincidence.

Equation (24.55) tells us

$$|1-\Omega(t)| \propto \frac{1}{a(t)^2 H(t)^2}. \quad (24.57)$$

During the matter-dominated era, $a(t) \propto t^{2/3}$ and $H(t) \equiv \dot{a}/a \propto t^{-1}$. Thus, during the matter-dominated era, the difference between Ω and 1 grew at the rate

$$|1-\Omega(t)|_m \propto t^{2/3} \propto a(t). \quad (24.58)$$

During the radiation-dominated era, $a(t) \propto t^{1/2}$ and $H(t) \equiv \dot{a}/a \propto t^{-1}$. Thus, during the radiation-dominated era, the difference between Ω and 1 grew at the rate

$$|1-\Omega(t)|_r \propto t \propto a(t)^2. \quad (24.59)$$

If we extrapolate back to the time of BBN ($t_{\rm nuc} \sim 3$ min), we compute that the deviation of Ω from one was

$$|1-\Omega(t_{\rm nuc})| \leq 10^{-14}. \quad (24.60)$$

At the time deuterium and helium were forming, the density of the universe was equal to the critical density with an accuracy of 1 part in 100 trillion. Our very existence depends on the astonishingly close match between the actual density and the critical density in the early universe. If, for instance, the deviation of Ω from 1 at the time of Big Bang nucleosynthesis had been 1 part in 100 thousand instead of one part in 100 trillion, the universe would have collapsed in a Big Crunch or expanded to a low-density Big Chill after only a few years. In either case, galaxies, stars, planets, and cosmologists would not have had time to form.

The flatness problem is simply the statement that Ω was very, very close to 1 in the early universe. It would be satisfying if we could find a physical mechanism for flattening the universe early in its history, rather than invoking a highly implausible coincidence.

The horizon problem is simply the statement that the universe is nearly homogeneous and isotropic on large scales. Why is this a problem? To see why large-scale homogeneity and isotropy is puzzling in the standard Hot Big Bang scenario, consider two antipodal points on the last scattering surface, as shown in Figure 24.10. Since the last scattering of CMB photons took place long ago ($t_{\rm ls} \approx 0.4$ Myr $\approx 3 \times 10^{-5} t_0$), the current proper distance to the last scattering surface is only slightly smaller than the horizon distance. In the Consensus Model, the last scattering surface is at a distance $\ell_p = 0.98\ell_{\rm hor}$ from us. Thus, two antipodal points on the last scattering surface are currently separated by a distance $1.96\ell_{\rm hor}$. Since the two points are farther apart than the horizon distance, they are not in causal contact. That is, they haven't had time to send messages to each other. In particular, they haven't had time to come into thermal equilibrium with each other. Nevertheless, the two points have the same temperature, once the dipole distortion is subtracted, to within 1 part in 10^5.

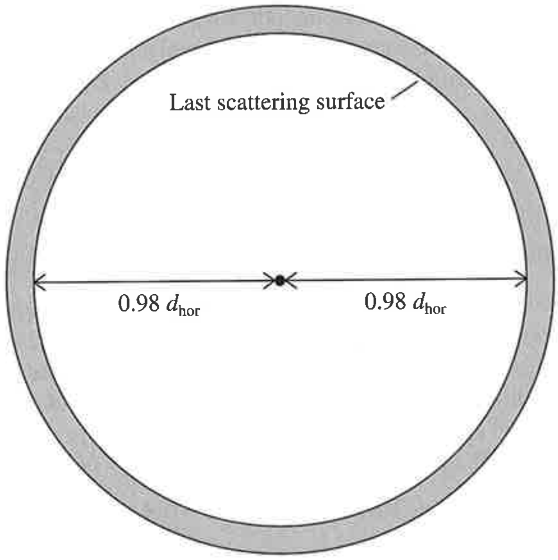

FIGURE 24.10 The distance between antipodal points on the last scattering surface.

How can two points that haven't had time to swap information be so nearly identical in their properties? It would be satisfying if we could find a physical mechanism for homogenizing the universe early in its history, rather than invoking a highly implausible coincidence.

The first satisfying solution to the flatness and horizon problems was provided by Alan Guth, who put forward the **inflationary theory** in 1981. In a cosmological context, "inflation" is the statement that there was a very early period when the acceleration of the expansion was positive ($\ddot{a} > 0$). As usually implemented, inflationary theory supposes that the universe was temporarily dominated by a cosmological constant Λ_i very much larger than the cosmological constant Λ present today.

When the universe is dominated by a cosmological constant, it expands exponentially (equation 24.27):

$$a(t) \propto e^{H_i t}, \tag{24.61}$$

where

$$H_i = \left(\frac{8\pi G u_\Lambda}{3c^2}\right)^{1/2}. \tag{24.62}$$

To see how inflation can solve the flatness and horizon problems, suppose that the universe had a period of exponential growth in the middle of its early radiation-dominated phase. For simplicity, let's suppose the exponential expansion switched on instantaneously at a time t_i, and lasted until some later time t_f, when it switched off instantaneously. In this simple case, the scale factor grows during the inflationary era ($t_i < t < t_f$) by a factor

$$\frac{a(t_f)}{a(t_i)} = e^N, \tag{24.63}$$

where N, the number of e-foldings of inflation, is

$$N \equiv H_i(t_f - t_i). \tag{24.64}$$

If the duration of inflation, $t_f - t_i$, was long compared to the Hubble time during inflation, H_i^{-1}, then N was large and the growth of the scale factor was exponentially huge.

For concreteness, let's take one popular model for inflation. According to Grand Unified Theories of particle physics, or GUTs, there was a phase transition that took place at a time $t_{\rm GUT} \approx 10^{-35}$ s, when the strong nuclear force separated from the electroweak force. In the GUT model of inflation, exponential growth began at the GUT time, $t_i \approx t_{\rm GUT} \approx 10^{-35}$ s, with a Hubble parameter $H_i \approx t_{\rm GUT}^{-1} \approx 10^{35}\ {\rm s}^{-1}$, and lasted for $N \sim 100$ e-foldings. In the GUT model, the growth in scale factor during inflation was

$$\frac{a(t_f)}{a(t_i)} \sim e^{100} \sim 10^{43}, \tag{24.65}$$

all happening in a time $\sim 100 t_{\rm GUT} \sim 10^{-33}$ s.

How does inflation resolve the flatness problem? In an exponentially expanding universe, equation (24.57) can be written in the form

$$|1 - \Omega(t)| \propto \frac{1}{a(t)^2 H(t)^2} \propto \frac{1}{e^{2H_i t}} \propto e^{-2H_i t}. \tag{24.66}$$

If we compare Ω at the beginning of inflation ($t = t_i$) to Ω at the end of inflation ($t = t_f = t_i + N/H_i$), we find

$$|1 - \Omega(t_f)| = e^{-2N}|1 - \Omega(t_i)|. \tag{24.67}$$

If the universe were strongly curved prior to inflation, with

$$|1 - \Omega(t_i)| \sim 1, \tag{24.68}$$

then 100 e-foldings of inflation would flatten it like the proverbial pancake, and then some:

$$|1 - \Omega(t_f)| \sim e^{-200} \sim 10^{-87}. \tag{24.69}$$

The current limits on the density parameter, $|1 - \Omega_0| \leq 0.1$, imply that $N > 60$ if inflation took place at the GUT time.

How does inflation resolve the horizon problem? Consider the entire universe directly visible to us today, that is, the region bounded by the surface of last scattering (Figure 24.10). Currently, the proper distance to the surface of last scattering is

$$\ell_p(t_0) = 0.98\ell_{\rm hor}(t_0) = 14{,}000\ {\rm Mpc}. \tag{24.70}$$

If inflation ended at a time $t_f \sim 10^{-33}$ s, this corresponds to a scale factor $a_f \sim 8 \times 10^{-27}$ in the Consensus Model. Thus, immediately after inflation, the portion of the universe visible to us today was crammed into a sphere of radius

$$\begin{aligned} \ell_p(t_f) &= a_f \ell_p(t_0) \\ &\sim (8 \times 10^{-27})(1.4 \times 10^4 \text{ Mpc}) \sim 10^{-22} \text{ Mpc} \sim 4 \text{ m}. \end{aligned} \tag{24.71}$$

Immediately after inflation, all the mass–energy destined to become the hundreds of billions of galaxies we see today was contained in a sphere a few yards in radius. This may boggle your mind. If so, be prepared for additional boggling. If there were $N \sim 100$ e-foldings of inflation, then prior to the inflationary epoch, the currently visible universe was contained in a sphere of radius

$$\ell_p(t_i) \sim e^{-100} \ell_p(t_f) \sim 10^{-43} \text{ m}. \tag{24.72}$$

What matters for the solution of the horizon problem is not that this distance is small (which it certainly is!) but that it is smaller than the horizon distance at t_i, the start of inflation. Since the universe was radiation-dominated before inflation, the preinflationary scale factor was $a(t) = a_i (t/t_i)^2$, and the horizon distance at t_i was

$$\ell_{\text{hor}}(t_i) = c a_i \int_0^{t_i} \frac{dt}{a_i (t/t_i)^2} = 2ct_i \sim 6 \times 10^{-27} \text{ m}, \tag{24.73}$$

assuming that inflation began at the GUT time, $t_i \sim 10^{-35}$ s. This horizon distance is over 16 orders of magnitude bigger than the size of the currently visible universe at time t_i. Thus, everything we see today had plenty of time to swap photons back and forth prior to inflation and come to thermal equilibrium.

The detailed particle physics behind inflation is beyond the scope of this book. The usual driving mechanism behind inflation involves a scalar field being caught in a "false vacuum state" for a finite length of time. A false vacuum state is one for which the potential energy of the field is not the global minimum. It takes some length of time for the scalar field to transit to the global minimum of the potential (the true vacuum state). During the time of transition, the energy of the scalar field plays the role of a cosmological constant. A scalar field in a false vacuum state is sometimes compared to a supercooled liquid. Freezing would lower the energy of the supercooled liquid, but until some disturbance initiates the freezing, it temporarily remains in the higher-energy, liquid state. When the freezing finally occurs, the latent heat of fusion is released and warms the surroundings. Similarly, when a scalar field goes from a false vacuum to the true vacuum, the energy released in going from a higher to lower potential energy warms up the universe, returning the temperature of the universe to what it was before ~ 100 e-foldings of inflation chilled it down.

It is tempting to extrapolate the scale factor back to $t = 0$, and $a = 0$, representing an infinite density singularity. One shortcoming of general relativity, however, is that it doesn't take quantum effects into account. A complete "quantum gravity" theory has not yet been devised. However, it is speculated that time is quantized in units of the **Planck time**

$$t_{\text{Pl}} = \left(\frac{G\hbar}{c^5}\right)^{1/2} = 5 \times 10^{-44}\ \text{s}, \tag{24.74}$$

and that talking about times earlier than the Planck time may not be physically meaningful. Although invoking quantum gravity prevents us from having to contemplate infinitely dense initial conditions, the properties at $t \approx t_{\text{Pl}}$ were fairly mind-boggling in themselves. At $t \approx t_{\text{Pl}}$, the number density of particles would have been $n \sim 10^{104}\ \text{m}^{-3}$, and the average particle energy would have been $E \sim 10^{28}$ eV; that's an energy comparable to the kinetic energy of a cruising passenger jet, concentrated in a single elementary particle. From this incredibly dense, hot state evolved the complex universe we see around us today.

PROBLEMS

24.1 (a) Given that the current scale factor is $a(t_0) = 1$, at what scale factor did the temperature of the cosmic background radiation equal the temperature of the Sun's photosphere?
(b) At what scale factor did it equal the temperature of the Sun's center?
(c) If the current mass density of the universe is equal to $0.3\rho_{c,0}$, what was the mass density of the universe when the temperature was equal to that of the Sun's center? Compare this mass density to the average density of the Sun.

24.2 Explicitly calculate the redshifts for the following:
(a) The universe goes from radiation-dominated to matter-dominated.
(b) The universe goes from matter-dominated to dark-energy-dominated.

24.3 At the time this problem was written, the highest-redshift quasar known was CFHQS J2329-0301, which has a redshift $z = 6.43$.
(a) What was the scale factor a of the universe at the time the quasar light we are observing now left the quasar?
(b) How old was the universe at the time the light left the quasar?
(c) What is the distance modulus of the quasar?

24.4 Suppose that star formation stops today, everywhere in the universe.
(a) At what time t_{die} will the last stars die out?
(b) What will be the scale factor $a(t_{\text{die}})$ at that time?

24.5 Estimate how high the temperature of the universe must be for proton–proton pair production to occur. What was the approximate age of the universe when it had cooled enough for proton–proton pair production to cease?