The large-scale distribution of galaxies

Since the early 1980s, multi-object spectrographs, CCD detectors, and some dedicated telescopes have allowed the mass production of galaxy redshifts. These large surveys have revealed a very surprising picture of the luminous matter in the Universe. Many astronomers had imagined roughly spherical galaxy clusters floating amongst randomly scattered field galaxies, like meatballs in sauce. Instead, they saw galaxies concentrated into enormous walls and long filaments, surrounding huge voids that appear largely empty. The galaxy distribution has been compared to walls of soapy water, surrounding bubbles of air in a basinful of suds; linear filaments appear where the walls of different soap bubbles join, and rich clusters where three or more walls run into each other. A more accurate metaphor is that of a sponge; the voids are interlinked by low-density ‘holes’ in the walls. Sometimes we think of the filaments as forming a cosmic web.

For a star like the Sun in the disk of our Milky Way, the task of finding where it formed is essentially hopeless, because it has already made many orbits about the galaxy, and the memory of its birthplace is largely lost. But the large structures that we discuss in this chapter are still under construction, and the regions where mass is presently concentrated reveal where denser material was laid down in the early Universe. The peculiar motions of groups and clusters of galaxies, their speeds relative to the uniformly expanding cosmos, are motions of infall toward larger concentrations of mass. So the problem of understanding the large-scale structure that we see today becomes one of explaining small variations in the density of the early Universe.

We begin in Section 8.1 by surveying the galaxies around us, mapping out both the local distribution and the larger structures stretching over hundreds of megaparsecs. The following sections discuss the history of our expanding Universe, within which the observed spongy structures grew, and how the expansion and large-scale curvature affect our observations of galaxies. In Section 8.4 we discuss fluctuations in the cosmic microwave background and what they tell us about the initial irregularities that might have given rise to the galaxies that we
Fig. 8.1. Positions of 14,650 bright galaxies, in Galactic longitude $l$ and latitude $b$. Many lie near the supergalactic plane, approximately along the Great Circle $l = 140^\circ$ and $l = 320^\circ$ (heavy line); V marks the Local Void. Galaxies near the plane $b = 0$ of the Milky Way's disk are hidden by dust – T. Kolatt and O. Lahav.
The large-scale distribution of galaxies

see today. We will see how we can use the observed peculiar motions of galaxies to estimate how much matter is present. The final section asks how dense systems such as galaxies developed from these small beginnings.


### 8.1 Large-scale structure today

As we look out into the sky, it is quite clear that galaxies are not spread uniformly through space. Figure 8.1 shows the positions on the sky of almost 15,000 bright galaxies, taken from three different catalogues compiled from optical photographs. Very few of them are seen close to the plane of the Milky Way’s disk at \( b = 0 \), and this region is sometimes called the *Zone of Avoidance*. The term is unfair: surveys in the 21 cm line of neutral hydrogen, and in far-infrared light, show that galaxies are indeed present, but their visible light is obscured by dust in the Milky Way’s disk. Dense areas on the map mark rich clusters: the Virgo cluster is close to the north Galactic pole, at \( b = 90^\circ \). Few galaxies are seen in the Local Void, stretching from \( l = 40^\circ, b = -20^\circ \) across to \( l = 0, b = 30^\circ \).

The galaxy clusters themselves are not spread evenly on the sky: those within about 100 Mpc form a rough ellipsoid lying almost perpendicular to the Milky Way’s disk. Its midplane, the *supergalactic plane*, is well defined in the northern Galactic hemisphere \( (b > 0) \), but becomes rather scruffy in the south. The pole or \( Z \) axis of the supergalactic plane points to \( l = 47.4^\circ, b = 6.3^\circ \). We take the supergalactic \( X \) direction in the Galactic plane, pointing to \( l = 137.3^\circ, b = 0^\circ \), while the \( Y \) axis points close to the north Galactic pole at \( b = 90^\circ \), so that \( Y \approx 0 \) along the Zone of Avoidance. The supergalactic plane is close to the Great Circle through Galactic longitude \( l = 140^\circ \) and \( l = 320^\circ \), shown as a heavy line in Figure 8.1. It passes through the Ursa Major group of Figures 5.6 and 5.8, and the four nearby galaxy clusters described in Section 7.2; the Virgo cluster at right ascension \( \alpha = 12^h \), declination \( \delta = 12^\circ \); Perseus at \( \alpha = 3^h, \delta = 40^\circ \); Fornax at \( \alpha = 4^h, \delta = -35^\circ \); and the Coma cluster at \( \alpha = 13^h, \delta = 28^\circ \), almost at the north Galactic pole.

Figure 8.2 shows the positions of the elliptical galaxies within about 20 Mpc, tracing out the Virgo and Fornax clusters. The distances to these galaxies have been found by analyzing *surface brightness fluctuations*. Even though they are too far away for us to distinguish individual bright stars, the number \( N \) of stars falling within any arcsecond square on the image has some random variation.
8.1 Large-scale structure today

Fig. 8.2. Positions of nearby elliptical galaxies on the supergalactic $X-Y$ plane; $\times$ shows the Milky Way. Shading indicates recession velocity $V_r$ – J. Tonry.

So the surface brightness of any square fluctuates about some average value. The closer the galaxy, the fewer stars lie within each square, and the stronger the fluctuations between neighboring squares: when $N$ is large, the fractional variation is proportional to $1/\sqrt{N}$. So if we measure the surface brightness fluctuations in two galaxies where we know the relative luminosity of the bright stars that emit most of the light, we can find their relative distances.

This method works only for relatively nearby galaxies in which the stars are at least 3–5 Gyr old. In these, nearly all the light comes from stars close to the tip of the red giant branch. As we noted in Section 1.1, these have almost the same luminosity for all stars below about $2M_\odot$. These stars are old enough that they have made many orbits around the center, so they are dispersed smoothly through the galaxy. The observations are usually made in the $I$ band near 8000 Å,
or in the $K$ band at 2.2 $\mu$m, to minimize the contribution of the younger bluer stars. The technique fails for spiral galaxies, because their brightest stars are younger: they are red supergiants and the late stages of intermediate-mass stars, and their luminosity depends on the stellar mass, and so on the average age of the stellar population. Since that average changes across the face of the galaxy, so does the luminosity of those bright stars. Also, the luminous stars are too short-lived to move far from the stellar associations where they formed. Their clumpy distribution causes much stronger fluctuations in the galaxy’s surface brightness than those from random variations in the number of older stars.

In Figure 8.2, we see that the Virgo cluster is roughly 16 Mpc away. It appears to consist of two separate pieces, which do not coincide exactly with the two velocity clumps around the galaxies M87 and M49 that we discussed in Section 7.2. Here, galaxies in the northern part of the cluster, near M49, lie mainly in the nearer grouping, while those in the south near M86 are more distant: M87 lies between the two clumps. The Fornax cluster, in the south with $Y<0$, is at about the same distance as Virgo. Both these clusters are part of larger complexes of galaxies. Because galaxy groups contain relatively few ellipticals, they do not show up well in this figure. The Local Group is represented only by the elliptical and dwarf elliptical companions of M31, as the overlapping circles just to the right of the origin.

**Problem 8.1** In Section 4.5 we saw that the motions of the Milky Way and M31 indicate that the Local Group’s mass exceeds $3 \times 10^{12} M_\odot$: taking its radius as 1 Mpc, what is its average density? Show that this is only about $3 h^{-2}$ of the critical density $\rho_{\text{crit}}$ defined in Equation 1.30 – the Local Group is only just massive enough to collapse on itself.

**Problem 8.2** The free-fall time $t_{ff} = 1/\sqrt{G\rho}$ of Equation 3.23 provides a rough estimate of the time taken for a galaxy or cluster to grow to density $\rho$. In Problem 4.7 we saw that, for the Milky Way, with average density of $10^5 \rho_{\text{crit}}$ within the Sun’s orbit, this minimum time is $\sim 3 \times 10^8$ years or $0.03 \times t_{H}$, the expansion age given by Equation 1.28. Show that a cluster of galaxies with density $200 \rho_{\text{crit}}$ can barely collapse within the age of the Universe. This density divides structures like the Local Group that are still collapsing from those that might have settled into an equilibrium.

To probe further afield, we use a ‘wedge diagram’ like Figure 8.3 from the 2dF survey, which measured redshifts of galaxies in two large slices across the sky. If we ignore peculiar motions, Equation 1.27 tells us that the recession speed $V_r \approx c z \approx H_0 d$, where $H_0$ is the Hubble constant; the redshift is roughly proportional to the galaxy’s distance $d$ from our position at the apex of each wedge. So this figure gives us a somewhat distorted map of the region out to $600 h^{-1}$ Mpc.
8.1 Large-scale structure today

The three-dimensional distribution of galaxies in Figure 8.3 has even more pronounced structure than Figure 8.1. We see dense linear features, the walls and stringlike filaments of the cosmic web; at their intersections there are complexes of rich clusters. Between the filaments we find large regions that are almost empty of bright galaxies: these voids are typically \( \gtrsim 50h^{-1}\) Mpc across. The galaxies appear to thin out beyond \( z = 0.15 \), because redshifts were measured only for objects that exceeded a fixed apparent brightness. Figure 8.4 shows that, at large distances, just the rarest and most luminous systems have been included. When a solid angle \( \Omega \) on the sky is surveyed, the volume between distance \( d \) and \( d + \Delta d \) is \( \Delta V = \Omega d^2 \Delta d \), which increases rapidly with \( d \). Accordingly, we see few galaxies nearby; most of the measured objects lie beyond 25 000 km s\(^{-1}\).

**Problem 8.3** The Local Group moves at 600 km s\(^{-1}\) relative to the cosmic background radiation. At this speed, show that an average galaxy would take \(~40h^{-1}\) Gyr to travel from the center to the edge of a typical void. Whatever process removed material from the voids must have taken place very early, when the Universe was far more compact.

In Figures 8.3 and 8.4, the walls appear to be several times denser than the void regions. But ignoring the peculiar motions has exaggerated their narrowness and
sharpness; they would appear less pronounced if we could plot the true distances of the galaxies. The extra mass in a wall or filament attracts nearby galaxies in front of the structure, pulling them toward it and away from us. So the radial velocities of those objects are increased above that of the cosmic expansion, and we overestimate their distances, placing them further from us and closer to the wall. Conversely, galaxies behind the wall are pulled in our direction, reducing their redshifts; these systems appear nearer to us and closer to the wall than they really are. In fact, the walls are only a few times denser than the local average.

By contrast, dense clusters of galaxies appear elongated in the direction toward the observer. The cores of these clusters have completed their collapse, and galaxies are packed tightly together in space. They orbit each other with random speeds as large as 1500 km s$^{-1}$, so their distances inferred from Equation 1.27 have random errors of up to 15$h^{-1}$ Mpc. In a wedge diagram, rich clusters appear as dense ‘fingers’ that point toward the observer.

**Problem 8.4** How long is the narrow ‘finger’ in the left panel of Figure 8.5 near $z = 0.045$ and 14$^h$30$^m$? Show that this represents a small galaxy cluster with $\sigma_r \approx 600$ km s$^{-1}$.

Figure 8.5 shows wedge diagrams for red galaxies, with spectra that show little sign of recent star formation, and blue galaxies, with spectral lines characteristic
8.1 Large-scale structure today

Fig. 8.5. About 27 000 red galaxies (left) with spectra like those of elliptical galaxies, and the same number of star-forming blue galaxies (right), in a slice $-32^\circ < \delta < -28^\circ$ from the 2dF survey. These are luminous galaxies, with $-21 < M(B_J) < -19$. The elliptical and S0 galaxies cluster more strongly than the spiral-like systems.

of young massive stars and the ionized gas around them. In Figure 8.4 we see clumps of galaxies around $13^h20^m$ at $z = 0.05, 0.08, 0.11$, and $0.15$. These are quite clear in the left wedge of Figure 8.5, but much weaker in the right wedge. Similarly, the void at $12^h$ and $z = 0.08$ is emptier in the left wedge – why? The left-wing galaxies are red elliptical and S0 systems, while on the right are spirals and irregulars. As we saw in Section 7.2, elliptical galaxies live communally in the cores of rich clusters, where spiral galaxies are rare. Accordingly, the clustering of the red galaxies is stronger than that of the bluer systems of the right panel. Whether galaxies are spiral or elliptical is clearly related to how closely packed they are: we see a morphology–density relation.

In fact, we should not talk simply of ‘the distribution of galaxies’, but must be careful to specify which galaxies we are looking at. We never see all the galaxies in a given volume; our surveys are always biased by the way we choose systems for observation. For example, if we select objects that are large enough on the sky that they appear fuzzy and hence distinctly nonstellar, we will omit the most compact galaxies. A survey that finds galaxies by detecting the 21 cm radio emission of their neutral hydrogen gas will readily locate optically dim but gas-rich dwarf irregular galaxies, but will miss the luminous ellipticals which usually lack HI gas. The Malmquist bias of Problem 2.11 is present in any sample selected by apparent magnitude. Even more insidious are the ways in which the bias changes with redshift and apparent brightness. Mapping even the luminous matter of the Universe is no easy task.
8.1.1 Measures of galaxy clustering

One way to describe the tendency of galaxies to cluster together is the two-point correlation function $\xi(r)$. If we make a random choice of two small volumes $\Delta V_1$ and $\Delta V_2$, and the average spatial density of galaxies is $n$ per cubic megaparsec, then the chance of finding a galaxy in $\Delta V_1$ is just $n \Delta V_1$. If galaxies tend to clump together, then the probability that we then also have a galaxy in $\Delta V_2$ will be greater when the separation $r_{12}$ between the two regions is small. We write the joint probability of finding a particular galaxy in both volumes as

$$\Delta P = n^2 [1 + \xi(r_{12})] \Delta V_1 \Delta V_2;$$  \hspace{1cm} (8.1)

if $\xi(r) > 0$ at small $r$, then galaxies are clustered, whereas if $\xi(r) < 0$, they tend to avoid each other. We generally compute $\xi(r)$ by estimating the distances of galaxies from their redshifts, making a correction for the distortion introduced by peculiar velocities. On scales $r \lesssim 10h^{-1} \text{Mpc}$, it takes roughly the form

$$\xi(r) \approx (r/r_0)^{-\gamma},$$  \hspace{1cm} (8.2)

with $\gamma > 0$. When $r < r_0$, the correlation length, the probability of finding one galaxy within radius $r$ of another is significantly larger than for a strictly random distribution. Since $\xi(r)$ represents the deviation from an average density, it must at some point become negative as $r$ increases.

Figure 8.6 shows the two-point correlation function $\xi(r)$ for galaxies in the 2dF survey. The correlation length $r_0 \approx 5h^{-1} \text{Mpc}$; it is $6h^{-1} \text{Mpc}$ for the ellipticals, which are more strongly clustered, and smaller for the star-forming galaxies. The slope $\gamma \approx 1.7$ around $r_0$. For $r \gtrsim 50h^{-1} \text{Mpc}$, which is roughly the size of the largest wall or void features, $\xi(r)$ oscillates around zero: the galaxy distribution is fairly uniform on larger scales.

Unfortunately, the correlation function is not very useful for describing the one-dimensional filaments or two-dimensional walls of Figure 8.3. If our volume $\Delta V_1$ lies in one of these, the probability of finding a galaxy in $\Delta V_2$ is high only when it also lies within the structure. Since $\xi(r)$ is an average over all possible placements of $\Delta V_2$, it will not rise far above zero once the separation $r$ exceeds the thickness of the wall or filament. We can try to overcome this by defining the three-point and four-point correlation functions, which give the joint probability of finding that number of galaxies with particular separations; but this is not very satisfactory. We do not yet have a good statistical method to describe the strength and prevalence of walls and filaments.

The Fourier transform of $\xi(r)$ is the power spectrum $P(k)$:

$$P(k) \equiv \int \xi(r) \exp(i k \cdot r) d^3 r = 4\pi \int_0^\infty \xi(r) \frac{\sin(kr)}{kr} r^2 dr;$$  \hspace{1cm} (8.3)

so that small $k$ corresponds to a large spatial scale.
8.1 Large-scale structure today

![Graph](image.png)

*Fig. 8.6.* Left, the correlation function $\xi(s)$ for the 2dF galaxies, at small (circles, left logarithmic scale) and large (triangles, right linear scale) separations; vertical bars show uncertainties. $\xi(s)$ is calculated assuming that Hubble’s law holds exactly: the ‘fingers’ of Figure 8.5 reduce $\xi(s)$ on scales $r \approx 1$ Mpc, but infall to the walls makes clustering look stronger on scales near $r_0$. The dashed line shows $\xi(r)$, corrected for these effects. Right, the variance $\sigma_R$ of Equation 8.4, describing how much the average density varies between regions of size $R - S$. Maddox and S. Cole.

Since $\xi(r)$ is dimensionless, $P(k)$ has the dimensions of a volume. The function $\sin(kr)/kr$ is positive for $|kr| < \pi$, and it oscillates with decreasing amplitude as $kr$ becomes large; so, very roughly, $P(k)$ will have its maximum when $k^{-1}$ is close to the radius where $\xi(r)$ drops to zero. In Figure 8.17 we show $P(k)$ calculated by combining the 2dF galaxy survey with observations of the cosmic microwave background.

**Problem 8.5** Prove the last equality of Equation 8.3. One method is to write the volume integral for $P(k)$ in spherical polar coordinates $r, \theta, \phi$ and set $k \cdot r = kr \cos \theta$. Show that, because $\xi(r)$ describes departures from the mean density, Equation 8.1 gives $\int_0^\infty \xi(r)r^2 \, dr = 0$, and hence $P(k) \to 0$ as $k \to 0$. ●

**Problem 8.6** Show that the power spectrum $P(k) \propto k^n$ corresponds to a correlation function $\xi(r) \propto r^{-(3+n)}$. Hence $\gamma \approx 1.5$ implies $n \approx -1.5$. Figure 8.17 shows that when $k$ is large $P(k)$ declines roughly as $k^{-1.8}$, about as expected. ●

Another way to describe the non-uniformity of the galaxy distribution is to ask how likely we are to find a given deviation from the average density. We can write the local density at position $x$ as a multiple of the mean level $\bar{\rho}$, as $\rho(x) = \bar{\rho}(1 + \delta(x))$, and let $\delta_R$ be the fractional deviation $\delta(x)$ averaged within a sphere of radius $R$. When we take the average $\langle \delta_R \rangle$ over all such spheres, this must
The large-scale distribution of galaxies

be zero. Its variance \( \sigma_R^2 = \langle \delta_R^2 \rangle \) measures how clumpy the galaxy distribution is on this scale. We can relate it to \( k^3 P(k) \), the dimensionless number prescribing the fluctuation in density within a volume \( k^{-1} \text{Mpc} \) in radius. If clumps of galaxies with size \( k^{-1} \approx R \) are placed randomly in relation to those on larger or smaller scales (the random-phase hypothesis), we have

\[
\sigma_R^2 \approx \frac{k^3 P(k)}{2\pi^2} \equiv \Delta_k^2, \quad \text{so, if } P(k) \propto k^n, \quad \text{then } \sigma_R \propto R^{-(n+3)/2}. \quad (8.4)
\]

Thus, if \( n > -3 \), the Universe is lumiest on small scales.

Figure 8.6 shows \( \sigma_R \) (or \( \Delta_k \)) for the 2dF galaxies. It increases with \( k \): the smaller the region we consider, the greater the probability of finding a very high density of galaxies. We often parametrize the clustering by \( \sigma_8 \), the average fluctuation on a scale \( R = 8 h^{-1} \text{Mpc} \): Figure 8.6 shows that \( \sigma_8 \approx 0.9 \). The wiggles at \( k \sim 0.1 \) correspond to the `baryon oscillations' that influence the cosmic background radiation: see Section 8.5. If \( P(k) \propto k^n \) and \( \sigma(M) \) is the variance in density over a region containing a mass \( M \approx 4\pi R^3 \bar{\rho}/3 \), then, since \( M \propto R^3 \), we have \( \sigma(M) \propto M^{-(n+3)/6} \). Cosmological models for the development of structure can predict \( P(k) \); we return to this topic in Section 8.5.

**Problem 8.7** The quantity \( (\Delta_k^2)^{1/2} \) gives the expected fractional deviation \( |\delta(x)| \) from the mean density in an overdense or diffuse region of size \( 1/k \). Write \( \delta(x) \) and \( \Phi(x) \) as Fourier transforms and use Equation 3.9, Poisson’s equation, to show that these lumps and voids cause fluctuations \( \Delta \Phi_k \) in the gravitational potential, where \( k^2 |\Delta \Phi_k| \sim 4\pi G \bar{\rho} (\Delta_k^2)^{1/2} \). Show that, when \( P(k) \propto k \), the *Harrison–Zel’dovich* spectrum, \( |\Delta \Phi_k| \) does not depend on \( k \): the potential is equally ‘rippled’ on all spatial scales. •

In this section we have seen that the present-day distribution of galaxies is very lumpy and inhomogeneous on scales up to \( 50h^{-1} \text{Mpc} \). But measurements of the cosmic background radiation show that its temperature is the same in all parts of the sky to within a few parts in 100 000. As we saw in Section 1.5, before the time of recombination when \( z = z_{\text{rec}} \approx 1100 \), the cosmos was largely filled with ionized gas. Since light could not stream freely through the charged particles, the gas was opaque and glowing like a giant neon tube. The Universe became largely neutral and transparent only after recombination. Because the cosmic background radiation is smooth today, we know that the matter and radiation were quite smoothly distributed at that time. How could our present highly structured Universe of galaxies have arisen from such uniform beginnings? To understand what might have happened, we must look at how the Universe expanded following the Big Bang, and how concentrations of galaxies could form within it.
8.2 Expansion of a homogeneous Universe

Because the cosmic background radiation is highly uniform, we infer that the Universe is isotropic – it is the same in all directions. We believe that on a large scale the cosmos is also homogeneous – it would look much the same if we lived in any other galaxy. Then, mathematics tells us that the length $s$ of a path linking any two points at time $t$ must be given by integrating the expression

$$\Delta s^2 = R^2 \left( \frac{\Delta \sigma^2}{1 - k \sigma^2} + \sigma^2 \Delta \theta^2 + \sigma^2 \sin^2 \theta \Delta \phi^2 \right),$$

where $\sigma, \phi, \theta$ are spherical polar coordinates in a curved space. The origin $\sigma = 0$ looks like a special point, but in fact it is not. Just as at the Earth’s poles where lines of longitude converge, the curvature here is the same as everywhere else, and we can equally well take any point to be $\sigma = 0$. The constant $k$ specifies the curvature of space. For $k = 1$, the Universe is closed, with positive curvature and finite volume, analogous to the surface of a sphere; $R$ is the radius of curvature. If $k = -1$, we have an open Universe, a negatively curved space of infinite volume, while $k = 0$ describes familiar unbounded flat space.

Near the origin, where $\sigma \ll 1$, the formula for $\Delta s$ is almost the same for all values of $k$; on a small enough scale, curvature does not matter. If we look at a tiny region, the relationships among angles, lengths, and volumes will be the same as they are in flat space. We can call the comoving coordinate $\sigma$ of Equation 8.5 an ‘area radius’, because at time $t$ the area of a sphere around the origin at radius $\sigma$ is

$$A(\sigma, t) = 4\pi R^2 \sigma^2.$$  

(8.6)

**Problem 8.8** In ordinary three-dimensional space, using cylindrical polar coordinates we can write the distance between two nearby points $(R, \theta, z)$ and $(R + \Delta R, \theta + \Delta \theta, z + \Delta z)$ as $\Delta s^2 = \Delta R^2 + R^2 \Delta \theta^2 + \Delta z^2$. The equation $R^2 + z^2 = R^2$ describes a sphere of radius $R$: show that, if our points lie on this sphere, then the distance between them is

$$\Delta s^2 = \Delta R^2 (1 + R^2 / z^2) + R^2 \Delta \theta^2 = R^2 \left( \frac{\Delta \sigma^2}{1 - \sigma^2} + \sigma^2 \Delta \phi^2 \right).$$

(8.7)

where $\sigma = R / R$. Integrate from a point P, at radius $R$ and with $z > 0$, to the ‘north pole’ at $z = R$ to show that the distance $s = R \arcsin \sigma$. Show that the circumference $2\pi R$ of the ‘circle of latitude’ through P is always larger than $2\pi s$, but approaches it when $s \ll R$. When $k = 1$ in Equation 8.5, any surface of constant $\phi$ is the surface of a sphere of radius $R$. $ullet$

The cosmic expansion is described by setting $R = R(t)$, allowing the radius of curvature to grow with time. Apart from their small peculiar speeds, galaxies
The large-scale distribution of galaxies

remain at points with fixed values of $\sigma, \phi, \theta$; so these are called comoving coordinates. The separation $d$ of two galaxies is just the length $s$ of the shortest path between them. So, if they stay at fixed comoving coordinates, $d$ expands proportionally to $R(t)$: Hubble’s law is just one symptom of the expansion of curved space. Equation 8.5 tells us that the two systems are carried away from each other at a speed

$$V_t = \dot{d} = \frac{\dot{R}(t)}{R(t)}d \equiv H(t)d.$$  \hspace{1cm} (8.8)

Here $H(t)$ is the Hubble parameter, which at present has the value $H_0$. Equation 8.8 describes the average motion of the galaxies; we defer discussion of the peculiar motions that develop from the gravitational pull of near neighbors until Section 8.4 below.

General Relativity tells us that the distance between two events happening at different times and in different places depends on the motion of the observer. But all observers will measure the same proper time $\tau$ along a path through space and time connecting the events, which is found by integrating

$$\Delta \tau^2 = \Delta t^2 - \Delta s^2/c^2.$$  \hspace{1cm} (8.9)

Light rays always travel along paths of zero proper time, $\Delta \tau = 0$. If we place ourselves at the origin of coordinates, then the light we receive from a galaxy at comoving distance $\sigma_e$ has followed the radial path

$$\frac{c}{\mathcal{R}(t)} = -\frac{\Delta \sigma}{\sqrt{1 - k\sigma^2}}.$$  \hspace{1cm} (8.10)

The light covers less comoving distance per unit of time as the scale length $\mathcal{R}(t)$ of the Universe grows. We can integrate this equation for a wavecrest that sets off at time $t_e$, arriving at our position now at time $t_0$:

$$c \int_{t_e}^{t_0} \frac{dt}{\mathcal{R}(t)} = \int_0^{\sigma_e} \frac{d\sigma}{\sqrt{1 - k\sigma^2}}.$$  \hspace{1cm} (8.11)

Suppose that another wavecrest sets off later, by a time $\Delta t_e$. We receive it at time $t_0 + \Delta t$, given by the same equation with the new departure and arrival times. The galaxy’s comoving position $\sigma_e$, and the integral on the right-hand side of Equation 8.11, have not changed. So the left-hand side also stays constant:

$$\int_{t_e + \Delta t_e}^{t_0 + \Delta t} \frac{dt}{\mathcal{R}(t)} = \int_{t_e}^{t_0} \frac{dt}{\mathcal{R}(t)} \quad \text{so} \quad \frac{\Delta t_e}{\mathcal{R}(t_e)} = \frac{\Delta t}{\mathcal{R}(t_0)},$$  \hspace{1cm} (8.12)

as long as $\Delta t \ll \mathcal{R}(t)/\dot{\mathcal{R}}(t)$. Thus all processes in the distant galaxy appear to be slowed down by the factor $\mathcal{R}(t_0)/\mathcal{R}(t_e)$. If $\Delta t_e$ is the time between two
consecutive crests emitted with wavelength $\lambda_e = c \Delta t_e$, that light is received with $\lambda_{\text{obs}} = c \Delta t$. So the wavelength grows along with the scale length $R(t)$, while the frequency, momentum, and energy of each photon decay proportionally to $1/R(t)$. The measured redshift $z$ of a distant galaxy tells us how much expansion has taken place since the time $t_e$ when its light set off on its journey to us. This is the cosmological redshift of Equation 1.34:

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_e} = \frac{R(t_0)}{R(t_e)}.$$  

(8.13)

We often use redshift $z(t)$ as a substitute for time $t_e$ or comoving distance $\sigma_e$. The time $t_e$ corresponding to a given redshift depends on how fast the Universe has expanded, while Equation 8.11 tells us the comoving distance $\sigma_e$ from which the light would have started.

The rate at which the Universe expands is set by the gravitational pull of matter and energy within it. We first use Newtonian physics to calculate the expansion, and then discuss how General Relativity modifies the result. Consider a small sphere of radius $r$, at a time $t$ when our homogeneous Universe has density $\rho(t)$; we take $r \ll R(t)$, so that we can neglect the curvature of space. Everything is symmetric about the origin $r = 0$, so we appeal to Newton’s first theorem in Section 3.1: the gravitational force at radius $r$ is determined only by the mass $M(<r)$ within the sphere. If our sphere is large enough that gas pressure forces are much smaller than the pull of gravity (see Section 8.5 below), then Equation 3.20 gives the force on a gas cloud of mass $m$ at that radius:

$$m \frac{d^2r}{dt^2} = -\frac{GmM(<r)}{r^2} = -\frac{4\pi Gm}{3} \rho(t)r.$$  

(8.14)

Our sphere of matter is expanding along with the rest of the Universe, so its radius $r(t) \propto R(t)$. The mass $m$ of the cloud cancels out, giving

$$\dot{R}(t) = -\frac{4\pi G}{3} \rho(t)R(t);$$  

(8.15)

the higher the density, the more strongly gravity slows the expansion.

Nothing enters or leaves our sphere, so the mass within it does not change: $\rho(t)R^3(t)$ is constant. Multiplying by $\dot{R}(t)$ tells us how the kinetic energy decreases as the sphere expands:

$$\frac{1}{2} \frac{d}{dt}[\dot{R}^2(t)] = -\frac{4\pi G}{3} \frac{\rho(t_0)R^3(t_0)}{R^2(t)} \dot{R}(t),$$  

(8.16)

where the time $t_0$ refers to the present day. Integrating, we have

$$\dot{R}^2(t) = \frac{8\pi G}{3} \rho(t)R^2(t) - kc^2,$$  

(8.17)
The large-scale distribution of galaxies

where $k$ is a constant of integration. Although we derived it using Newtonian theory, Equation 8.17 is also correct in General Relativity, which tells us that the constant $k$ is the same one as in Equation 8.5. According to thermodynamics, as heat $\Delta Q$ flows into a volume $V$ its internal energy $E$ must increase, or it expands and does work against pressure:

$$\Delta Q = \Delta E + p \Delta V = V \Delta \rho c^2 + (\rho c^2 + p) \Delta V,$$  \hspace{1cm} (8.18)

where the density $\rho$ includes all forms of matter and energy. The cosmos is uniform, so no volume $V$ gains heat at the expense of another:

$$\Delta Q = 0 = \Delta \rho + \left( \rho + \frac{p}{c^2} \right) \frac{\Delta V}{V}, \quad \text{or} \quad \frac{d\rho}{dt} = -\frac{3}{R(t)} \left( \rho + \frac{p}{c^2} \right).$$  \hspace{1cm} (8.19)

Differentiating Equation 8.17 and substituting for $d\rho/dt$ yields

$$\frac{dR}{dt} = -\frac{4\pi G}{3} R \left[ \rho(t) + \frac{3p(t)}{c^2} \right].$$  \hspace{1cm} (8.20)

This change to Equation 8.15 shows that, in General Relativity, the pressure $p$ adds to the gravitational attraction. Equations 8.17 and 8.20 describe the Friedmann models, telling us how the contents of the Universe determine its expansion.

For cool matter, the pressure $p \sim \rho c_s^2$, where the sound speed $c_s \ll c$. So we can safely neglect the pressure term in Equation 8.20, and Equation 8.19 tells us that the density follows $\rho(t) \propto R^{-3}(t)$. For radiation, and particles moving almost at the speed of light, pressure is important: $p \approx \rho c^2/3$, where $\rho$ is the energy density divided by $c^2$. We have $\rho(t) \propto R^{-4}(t)$ from Equation 8.19. For any mixture of matter and radiation, the term $\rho + 3p/c^2$ must be positive, so the expansion always slows down. While matter and radiation account for most of the energy, $\rho(t) R^2(t)$ decreases as $R(t)$ grows. Thus, in a closed Universe with $k = 1$, the right-hand side of Equation 8.17 becomes negative at large $R$. But $R^2$ cannot be negative, so the distance between galaxies cannot grow forever; $R(t)$ attains some maximum before shrinking again. If the Universe is open with $k \geq 0$, the expansion will never halt.

General Relativity also allows a vacuum energy, with constant density $\rho_{\text{VAC}} = \Lambda/(8\pi G)$. Since $\rho_{\text{VAC}}$ does not change, the rightmost term of Equation 8.19 must also be zero, so the pressure $p_{\text{VAC}} = -\Lambda c^2/(8\pi G)$. Instead of a pressure pushing inwards on the contents of our sphere, this term represents a tension pulling outwards. The vacuum energy contributes a positive term to the right-hand side of Equation 8.20, speeding up the expansion. If the Universe expands far enough, the vacuum energy must become the largest term, and $R(t)$ then grows exponentially.
Problem 8.9 By substituting into Equation 8.20, show that, when vacuum energy dominates the expansion, we have \( R(t) \propto \exp(t \sqrt{\Lambda/3}) \). •

There are reasons to believe that very early, at \( t \lesssim 10^{-32} \) s, \( \rho_{\text{VAC}} \) might have been much larger than the density of matter or radiation. During this period, \( R(t) \) inflated, growing exponentially by a factor \( \sim e^{70} \approx 10^{30} \). The almost uniform cosmos that we now observe would have resulted from the expansion of a tiny near-homogeneous region. Because this patch was so small, the curvature of space within it would have been negligible; hence devotees of inflation expect our present Universe to be nearly flat, with \( k = 0 \). We will see in Section 8.4 that the measured temperature fluctuations in the cosmic background radiation imply that this is in fact so.

Further reading: for a discussion of the physics behind inflation, see Chapter 11 of Ryden’s Introduction to Cosmology.

In the borderline case when space is flat and \( k = 0 \), Equation 8.17 requires that the density \( \rho \) is equal to the critical value

\[
\rho(t) = \rho_{\text{crit}}(t) \equiv \frac{3H^2(t)}{8\pi G}. \quad (8.21)
\]

At the present day, the critical density \( \rho_{\text{crit}}(t_0) = 3.3 \times 10^{11} h^2 \mathcal{M}_\odot \text{Mpc}^{-3} \). We can measure the mass content of the Universe as a fraction of this critical density, defining the density parameter \( \Omega(t) \) as

\[
\Omega(t) \equiv \frac{\rho(t)}{\rho_{\text{crit}}(t)} , \quad (8.22)
\]

and writing \( \Omega_0 \) for its present-day value. Equation 8.17 then becomes

\[
H^2(t)[1 - \Omega(t)] = -k c^2 / R^2(t). \quad (8.23)
\]

If the Universe is closed, with \( k = 1 \), then \( \Omega(t) > 1 \) and the density always exceeds the critical value, whereas if \( k = -1 \), we always have \( \Omega(t) < 1 \). If the density is now equal to the critical value, then Equation 8.23 tells us that \( \Omega(t) \) must be unity at all times. The present value \( \Omega(t_0) \) is often written as \( \Omega_{\text{tot}} \) or (especially when \( \Lambda = 0 \)) as \( \Omega_0 \).

We already saw in Section 1.5 that normal (baryonic) matter, largely neutrons and protons, makes up only 4%–5% of the critical density. Including the dark matter, in Section 8.4 below we arrive at only \((0.2-0.3) \rho_{\text{crit}}\). Radiation contributes hardly anything. So space can be flat only if there is a nonzero vacuum energy. This is often called the dark energy. It probably has a different physical origin from
The large-scale distribution of galaxies

Fig. 8.7. For the benchmark cosmology with $\Omega_r = 8.42 \times 10^{-5}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$, the fraction of the critical density contributed at each time by radiation (dotted line), matter (solid), and the dark energy (dashed). For this model, matter–radiation equality occurs at $z_{eq} = 3570$, $t_{eq} = 0.05$ Myr; recombination is complete at $z_{rec} = 1100$, $t_{rec} = 0.35$ Myr. The present age $t_0 = 13.4$ Gyr.

the vacuum energy causing inflation at early times. We describe the contributions of matter and dark energy at the present time by writing

$$\Omega_m = \frac{\rho_m(t_0)}{\rho_{crit}(t_0)} \quad \text{and} \quad \Omega_\Lambda = \frac{\rho_{VAC}}{\rho_{crit}(t_0)}.$$  \hspace{1cm} (8.24)

In the benchmark model illustrated in Figure 8.7, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$; thus space is flat. The baryon density $\Omega_B = 0.045$, with cold dark matter (see Section 8.5) making up the remainder of $\Omega_m$. A warning: although $\Omega_m$, $\Omega_\Lambda$, and similar quantities do not carry the subscript 0, they always refer to the present day.

After inflation ended, the Universe was radiation-dominated. It was so extremely hot that its energy was almost entirely due to radiation and relativistic particles, moving so close to light speed that their energy, momentum, and pressure are related in the same way as for photons. From Equation 8.19, we know that the energy density $\rho_r c^2$ is proportional to $R^{-4}(t)$. (If few photons are created or destroyed, the number per unit volume is proportional to $1/R^3(t)$, while by Equation 8.12 the energy of each falls as $1/R(t)$.) For a blackbody spectrum, the temperature $T \propto 1/R(t)$: recall Problem 1.17.

As expansion proceeded the density of matter followed $\rho_m(t) \propto R^{-3}(t)$, falling more slowly than the radiation density. At the time $t_{eq}$ of matter–radiation equality, about a million years after the Big Bang, its energy density $\rho_m(t)c^2$ exceeded that in radiation, and the Universe became matter-dominated. Figure 8.7 shows that, in the benchmark model, the matter density has only recently dropped below that of the dark energy.
8.2 Expansion of a homogeneous Universe

Problem 8.10 The cosmic background radiation is now a blackbody of temperature \( T = 2.73 \) K: show that its energy density \( \rho c^2 = 4.2 \times 10^{-13} \text{ erg cm}^{-3} \), so \( \Omega_r = 4.1 \times 10^{-5} h^{-2} \). From Equation 1.30, the matter density \( \rho_m = 1.9 \times 10^{-29} \Omega_m h^2 \text{ g cm}^{-3} \). Show that the energy density \( \rho_m c^2 \) was equal to that in radiation at redshift \( z_{eq} \approx 40000 \Omega_m h^2 \). This is well before the redshift of recombination, when the gas becomes largely neutral. If the neutrinos \( \nu_e, \nu_\mu, \) and \( \nu_\tau \) have masses \( m_\nu \ll k_B T_{eq}/c^2 \), where \( T_{eq} \) is the temperature at the time \( t_{eq} \), then at earlier times they are relativistic. The energy density of ‘radiation’ is increased by a factor of 1.68, and equalization is delayed until \( z_{eq} \approx 24000 \Omega_m h^2 \), or \( z \approx 3600 \) in the benchmark model.

To measure the cosmic expansion relative to the present day, we define the dimensionless scale factor \( a(t) = R(t)/R(t_0) \). We use Equation 8.23 to rewrite \( k/R^2(t_0) \) in terms of the present-day quantities \( H_0 \) and \( \Omega_{tot} = \Omega(t_0) \). Then Equation 8.17 becomes

\[
\frac{kc^2}{R^2(t_0)} = H_0^2(1 - \Omega_{tot}) = a^2(t) \left[ H^2(t) - \frac{8\pi G}{3} \rho(t) \right].
\]

(8.25)

Adding up the contributions to \( \rho(t) \) and recalling that \( 1+z = 1/a(t) \), we have

\[
H^2(t) = H_0^2[\Omega_r (1+z)^4 + \Omega_m (1+z)^3 + (1 - \Omega_{tot})(1+z)^2 + \Omega_\Lambda].
\]

(8.26)

Problem 8.11 Blackbody radiation and relativistic particles provide most of the energy density at \( t \ll t_{eq} \). Show that Equation 8.26 then implies that \( H(t) = \dot{a}/a \propto a^{-2} \), so \( \dot{a} \propto 1/a(t) \); hence \( R(t) \propto t^{1/2} \), and \( H(t) = 1/(2t) \). Early on, the leftmost term of Equation 8.25 is tiny, so \( H^2(t) \approx 8\pi G \rho(t)/3 \); show that the temperature \( T(t) \) is given by Equation 1.38.

Problem 8.12 Use Equation 8.20 to show that, in the benchmark model, cosmic expansion has speeded up since \( \rho_m(t) = 2\rho_{VAC} \), at redshift \( z = 0.67 \). According to this model, we live during that small fraction of cosmic history in which normal gravity and the cosmological constant have roughly equal influence on the expansion. At \( z \geq 1 \), the \( \Lambda \) term hardly affects the expansion, whereas when \( R(t) \geq 2R(t_0) \), the gravitational pull of matter will become irrelevant.

Problem 8.13 Show that, when cool matter accounts for most of the energy density, and the Universe is flat with \( k = 0 \), we have

\[
\dot{a} \propto a^{-1/2}, \quad \text{and} \quad a(t) \propto t^{2/3}.
\]

(8.27)
Show that, if \( k = 0 \), then Equation 8.27 holds when \( z_{\text{eq}} \gg 1 + z \gg (\Omega_\Lambda / \Omega_m)^{1/3} \), whereas at very late times \( a(t) \propto \exp(t \sqrt{\Lambda}/3) \). If the curvature is negative, with \( \Omega_{\text{tot}} = \Omega_r + \Omega_m + \Omega_\Lambda < 1 \), the third term of Equation 8.26 exceeds the second when \( 1 + z < (1 - \Omega_{\text{tot}})/\Omega_m \). Expansion then proceeds almost at a constant speed, with \( a(t) \propto t \); it is barely slowed by the matter, and not accelerated by dark energy until \( (1 + z)^2 < \Omega_\Lambda/(1 - \Omega_{\text{tot}}) \).

Equation 8.27 is a good description of the expansion over much of cosmic history. In particular, most of the structure of galaxy clusters and voids that we see today developed after the Universe became matter-dominated, but before the dark energy became important.

**Problem 8.14** Even if the cosmos has infinite volume, we can observe only a finite part of it because light travels at a finite speed. From Equation 8.11, light reaching us at time \( t \) has travelled no further than

\[
c \int_0^t \frac{dt'}{R(t')} = \int_0^{\sigma_H} \frac{d\sigma}{\sqrt{1 - k\sigma^2}},
\]

so we cannot see beyond our horizon, at comoving radius \( \sigma_H \). Explain why only points within \( \sigma_H(t) \) of each other can exchange signals or particles before time \( t \). Use Equation 8.27 to show that, while the Universe was matter-dominated, \( R(t)\sigma_H \approx 3ct \); at the time \( t_{\text{rec}} \) of recombination \( R(t_{\text{rec}})\sigma_H \approx 0.43 \text{ Mpc} \) in the benchmark model of Figure 8.7. We will see below that a sphere of matter with this diameter would cover only about 2° on the sky. So it is very surprising that the cosmic microwave background has almost the same spectrum across the whole sky.

An inflationary cosmology can explain this horizon problem. When \( R(t) \propto \exp(t \sqrt{\Lambda}/3) \) after some initial time \( t_i \), show that when \( t \gg t_i \)

\[
\int_{t_i}^t \frac{dt'}{R(t')} \approx \sqrt{\frac{3}{\Lambda}} \frac{1}{R(t_i)} \text{ so } R(t)\sigma_H \approx c \sqrt{\frac{3}{\Lambda}} \frac{R(t)}{R(t_i)}.
\]

For \( t_i \sim 10^{-35} \text{ s} \) and \( \sqrt{3/\Lambda} \sim t_i \) as in most inflationary models, show that, when inflation ends at \( t_f \approx 70t_i \), the horizon distance \( R(t_f)\sigma_H \approx ct_f \times e^{70/70} \approx 3.6 \times 10^{28}ct_f \) or \( \sim 7 \text{ km} \). After inflation, the Universe is radiation-dominated, so \( a(t) \propto t^{1/2} \). Show that, by the time of matter–radiation equality at \( \sim 0.35 \text{ Myr} \), this distance has grown by \( z(t_{\text{eq}})/z(t_f) \) to about 30 Gpc. Inflation expands a region within which light signals could be exchanged, until it is much larger than the entire Universe that we can see today.

**Further reading:** on cosmic horizons, see Chapter 10 of the book by Padmanabhan.
8.2 Expansion of a homogeneous Universe

8.2.1 How old is that galaxy? Lookback times and ages

Because light travels at a finite speed, we see a younger cosmos as we look toward more distant galaxies at higher redshifts. As we observe a galaxy at redshift \( z \), at what time \( t_e \) did those photons leave on their journey toward us? Equation 8.13 tells us that they arrive today at time \( t_0 \) with redshift \( 1 + z = \frac{R(t_0)}{R(t_e)} \). Light from another galaxy at a smaller redshift \( z - \Delta z \) must have been emitted slightly later by a small interval \( \Delta t_e \), when \( R(t_e) \) was larger by \( \Delta R = \frac{\dot{R}(t_e)}{R(t_e)} \Delta t_e \). Since

\[
\Delta z = -\frac{\Delta R}{R^2} \frac{R(t_0)}{R(t_e)}, \quad \text{we have} \quad \Delta t_e = \frac{1}{H(t_e)} \frac{\Delta z}{1 + z}, \quad (8.30)
\]

where \( H(t) = \frac{\dot{R}(t)}{R(t)} \) is the Hubble parameter. Integrating the second relation gives us the lookback time \( t_0 - t_e \):

\[
t_0 - t_e = \int_0^z \frac{1}{H(t)} \frac{dz'}{1 + z'} \quad (8.31)
\]

At redshifts \( z \ll z_{eq} \), when the density of radiation is no longer important, we can calculate this time simply for two special cases: a flat Universe with \( k = 0 \), and one with matter only, so that \( \Lambda = 0 \). When the densities of dark energy and of matter must add to give \( \Omega_m + \Omega_\Lambda = 1 \),

\[
t_e = \int_0^\infty \frac{1}{H(z')} \frac{dz'}{1 + z'} = \int_0^\infty \frac{dz'}{H_0(1 + z')(\Omega_m(1 + z')^3 + \Omega_\Lambda)^{1/2}}, \quad (8.32)
\]

which integrates to

\[
t_e = \frac{2}{3H_0\sqrt{\Omega_\Lambda}} \ln \left( \frac{1 + \cos \theta}{\sin \theta} \right), \quad \text{where} \quad \tan^2 \theta = \frac{(1 - \Omega_\Lambda)}{\Omega_\Lambda} (1 + z)^3. \quad (8.33a)
\]

Even if \( k \neq 0 \), this formula is accurate to within a few percent if we replace \( \Omega_\Lambda \) by \( 0.3\Omega_\Lambda + 0.7(1 - \Omega_m) \), so long as this quantity is positive. For \( \Lambda = 0 \), we have

\[
t_e = \int_0^\infty \frac{1}{H_0\sqrt{1 + \Omega_0 z}(1 + z')^2},
\]

where \( \Omega_0 = \Omega_m \) is the present-day value of the density parameter \( \Omega(t) \). We have exact expressions when \( \Omega_0 = 1 \) and the density is equal to the critical value, or in an empty Universe with \( \Omega_0 = 0 \):

\[
t_0 - t_e = \begin{cases} \frac{1}{H_0} \int_0^z \frac{dz'}{(1 + z')^{3/2}} = \frac{2}{3H_0} \left[ 1 - \frac{1}{(1 + z')^{1/2}} \right] & \text{if } \Omega_0 = 1, \\ \frac{1}{H_0} \int_0^z \frac{dz'}{(1 + z')^2} = \frac{2}{H_0} \frac{z}{(1 + z)} & \text{for } \Omega_0 = 0 \end{cases} \quad (8.33b)
\]
The large-scale distribution of galaxies

![Graph showing the relationship between lookback time and redshift](image)

**Fig. 8.8.** Left, lookback time $t_0 - t_e$ to an object seen with redshift $z$, in units of the Hubble time $t_H$; the right-hand scale shows $t_e$ itself. Right, comoving volume per unit redshift, $R^{-3}(t) dV/dz$, in units $(c/H_0)^3$. The solid line is for $\Lambda = 0, \Omega_m = 1$; the dashed line shows $\Lambda = 0 = \Omega_m$; the dotted line is for the benchmark model.

Early in the matter-dominated era, the curvature term that we left out of Equation 8.32 will be tiny and the $\Omega_\Lambda/\Omega_m$ term is also small, so

$$t_e \approx \frac{2}{3H_0\sqrt{\Omega_m}} \frac{1}{(1 + z)^{3/2}} \quad \text{when} \quad \frac{\Omega_m}{\Omega_r} \gg 1 + z \gg \left(\frac{\Omega_\Lambda}{\Omega_m}\right)^{1/3} \quad \text{and} \quad \frac{1}{\Omega_m}. \quad (8.34)$$

**Problem 8.15** Nearby, Hubble’s law tells us that an object at distance $d$ recedes with speed $V_r = cz = H_0 d$. The lookback time is just the time that light takes to cover this distance, so $t_0 - t_e = d/c = z/H_0$. Show that both parts of Equation 8.33 agree with this in the limit $z \ll 1$.

**Problem 8.16** Show that, if $\Omega = 1$ in a matter-dominated Universe, then $H(t) = 2/(3t)$, so that the time $t_0$ since the Big Bang is two-thirds of our simple estimate $t_H = 1/H_0$ in Equation 1.28. From Equation 8.21, show that the density falls as $1/t^2$.

Equation 8.33 also tells us $t_0$, the present age of the Universe. When $\Lambda = 0$, this is always less than our simple estimate $t_H = 1/H_0$ in Equation 1.28. For an empty Universe with $\Lambda = 0$, $t_0 = H_0^{-1} = t_H$, whereas for $\Omega_0 = 1$ the age is only two-thirds as large. For a fixed value of $H_0$, Figure 8.8 shows that the lookback time in gigayears to a given redshift $z$ is longer in the case $\Omega_0 = 0$ than for $\Omega_0 = 1$, but the galaxies are older: the time $t_e$ is longer. When $\Lambda > 0$ and $k = 0$ so the Universe is flat, $t_0$ exceeds the Hubble time $t_H = 1/H_0$ if the matter density is low, with $\Omega_m \lesssim 0.2$. For example, $\Omega_\Lambda = 0.9, \Omega_m = 0.1$ yields $t_0 = 1.3t_H$. In the benchmark model, the age is 0.964$t_H$ or 13.4 Gyr.
8.3 Observing the earliest galaxies

Problem 8.17 Galaxies have now been observed at redshifts \( z \gtrsim 5 \): how old are they? Show that, at \( z = 5 \), \( t_e = 0.17t_0 = 1.6h^{-1}\text{Gyr} \) if the Universe is nearly empty so \( \Omega_0 \approx 0 \), but only \( 0.07t_0 = 0.44h^{-1}\text{Gyr} \) if \( \Omega_0 = 1 \), while for the benchmark model \( t_e = 0.8h^{-1}\text{Gyr} \). The time available to assemble the earliest galaxies is very short! Show that, at redshift \( z = 3 \), the time \( t_e = 2.4h^{-1}\text{Gyr} \) if \( \Omega_0 = 0 \), \( 0.82h^{-1}\text{Gyr} \) for \( \Omega_0 = 1 \), and \( 0.07t_0 = 1.5h^{-1}\text{Gyr} \) for the benchmark model.

8.3 Observing the earliest galaxies

Our view of very distant objects is complicated by the cosmic expansion, and by the curvature of the space through which their light must travel. Because of the expansion, distant galaxies look bigger than we would expect: at redshifts \( z \) larger than \( 1/\Omega_m \) a given object covers more of the sky when it is further from us. But the starlight of these big galaxies is rapidly dimmed as it spreads out in the expanding cosmos. Their ultraviolet and visible light is also shifted into the infrared, where strong emission lines from the Earth’s atmosphere (see Figure 1.15) make it more difficult to observe from the ground. To estimate how densely galaxies are scattered through space, we need to take account of both cosmic expansion and the curvature of space.

8.3.1 Luminosity, size, and surface brightness

For nearby objects, the apparent brightness \( F \) of a star or galaxy at distance \( d \) is related to its luminosity \( L \) by

\[
F = \frac{L}{4\pi d^2}. \tag{1.1}
\]

In an expanding Universe, this no longer holds. Objects appear dimmer because the cosmic expansion saps photons of their energy, and the area of the sphere over which the light rays must spread is expanding. This is just as well, because in a homogeneous static cosmos the light of distant stars would accumulate without limit, making the sky bright everywhere: this is Olbers’ paradox.

Problem 8.18 Suppose that galaxies of luminosity \( L \) are spread uniformly through space. Show that the number \( N(>F) \) that you observe to have apparent brightness larger than \( F \) varies as \( F^{-3/2} \), while the number between \( F \) and \( F + \Delta F \) varies as \( N(F) \propto F^{-5/2} \). Explain why \( N(F) \propto F^{-5/2} \) even when the galaxies have a range of luminosities, as long as they are spread uniformly and the luminosity function is the same everywhere.
The large-scale distribution of galaxies

Fig. 8.9. Left: as observer O receives the light of galaxy G, its energy is spread over a sphere of area $4\pi R(t_0)^2 \sigma_e^2$. Right: as seen by observer O, galaxy G with diameter $D$ covers $D/R(t_e)\sigma_e$ radians on the sky.

Show that the total light from all galaxies that appear brighter than $F_\star$ is proportional to $\int_{F_\star}^{\infty} F N(F) dF$. Whenever $N(F)$ rises faster than $1/F$ as $F \to 0$, this grows without limit as $F_\star$ decreases. •

To see how we should modify Equation 1.1, let $\sigma_e$ be the comoving distance or ‘area radius’ of a galaxy G that we observe today with redshift $z$. The apparent brightness of a distant galaxy is diminished by one power of $1+z$ because each photon carries less energy, and by another because those photons arrive at a slower rate. The left panel of Figure 8.9 shows that G’s light is now spread over a sphere with area $4\pi R(t_0)^2 \sigma_e^2$. The flux of energy $F$ that we receive, in W m$^{-2}$, is related to the total or bolometric luminosity $L$ by

$$F = \frac{L}{4\pi R(t_0)^2 \sigma_e^2 (1+z)^2} \equiv \frac{L}{4\pi d_L^2},$$

where $d_L = (1+z)R(t_0)\sigma_e$. (8.35)

We call $d_L$ the luminosity distance of the galaxy.

Similarly, we must modify Equation 1.2, telling us how large an object will appear on the sky. Suppose that galaxy G is $D$ kpc across. The right panel of Figure 8.9 shows that, at time $t_e$, it covered a fraction $D/[2\pi R(t_e)\sigma_e]$ of the circumference of the sky. So it extends over an angle

$$\alpha \text{ (in radians)} = \frac{D}{R(t_e)\sigma_e} \equiv \frac{D}{d_A}, \text{ so } d_A = \frac{R(t_e)\sigma_e}{1+z}.$$ (8.36)

Thus the angular-size distance $d_A$ is less than the luminosity distance $d_L$ by a factor $(1+z)^2$. We must use the distance $d_A$ when calculating the gravitational bending of light using Equations 7.14 and 7.25. A warning: some authors refer to $R(t_0)\sigma_e$ as the angular-size distance; this is $1+z$ times larger than our $d_A$. 


8.3 Observing the earliest galaxies

Fig. 8.10. Left, the measured radii $R_{25}$ of two model galaxies: an exponential disk (solid curve) and an $R^{1/4}$ spheroid (Equation 6.1: dashed curve). At $z = 0$ we take $R_{25} = 5hR$ for the disk and $R_{25} = 4R_e$ for the spheroid, then for each redshift we plot the radius where the measured surface brightness reaches this level. Right, the fraction of the total light within this radius. At small redshifts, the exponential disk shrinks less than the spheroid, but when $z \gtrsim 1$ it is more strongly affected.

Problem 8.19 Show that a galaxy of known luminosity $L$ and diameter $D$ at redshift $z$ will appear larger by a factor of $(1 + z)^2$ than one would expect by using Equation 1.1 to calculate its distance from its measured apparent brightness, and then finding the apparent diameter from Equation 1.2.

The surface brightness $I(\mathbf{x})$ of a galaxy is the flux we receive from each square arcsecond as it appears on the sky. If we integrate over all wavelengths to measure the bolometric surface brightness, then, instead of Equation 1.23, we have

$$I(\mathbf{x}) = \frac{F}{\alpha^2} = \frac{L/(4\pi d_L^2)}{D^2/d_A^2} = \frac{L}{4\pi D^2} \left(\frac{d_A}{d_L}\right)^2 = \frac{L}{4\pi D^2} \frac{1}{(1 + z)^4}. \quad (8.37)$$

The bolometric surface brightness of a nearby galaxy, with $z \ll 1$, does not diminish with distance. If we neglect changes caused by the birth of new stars or by stellar aging, then a given isophote always corresponds to a fixed radius in the galaxy, and encloses the same fraction of the galaxy’s light, regardless of the distance. But at redshifts beyond a few tenths $I(\mathbf{x})$ drops rapidly, making photometry increasingly difficult and expensive.

Figure 8.10 shows how an isophote defined by a fixed surface brightness shrinks, and the amount of light within it decreases. We must correct for this missing light when comparing the luminosities of distant galaxies with those of nearby galaxies. The correction would depend only on the redshift $z$ if the luminosity from each square parsec of the galaxy had remained unchanged. But
we will see in Section 9.3 that many galaxies at \( z > 0.5 \) were significantly brighter than their counterparts today. We can interpret the brightening with models to tell us how the light of new stars should fade through time; Figure 6.18 illustrates a simple model in which all the galaxy's stars are born at once. To turn this into a prediction about luminosity at a given redshift, we need Equation 8.31, which involves the expansion rate \( H(z) \) as well as the redshift.

To calculate the distances \( d_L \) and \( d_A \) to a galaxy \( G \) seen at redshift \( z \), we must know its comoving 'area' distance \( \sigma_e \). We start by asking how far its light must go to reach us. According to Equation 8.5, at time \( t \) the distance from the origin to a point with area radius \( \sigma \) is

\[
\mathcal{R}(t) \chi \equiv \int_0^{\sigma} \frac{ds}{\sqrt{1 - k \sigma^2}}.
\]

This defines the comoving 'distance radius' \( \chi \), with

\[
\chi(\sigma) = \begin{cases} 
\arcsin(\sigma) & \text{for } k = 1, \\
\sigma & \text{for } k = 0, \\
\arcsinh(\sigma) & \text{for } k = -1.
\end{cases}
\]

The two distances are the same if the Universe is flat, with \( k = 0 \). In a closed Universe, with \( k = 1 \), the area radius \( \sigma \) is smaller than the distance radius \( \chi \), just as the length of a circle of latitude on the Earth is less than \( 2\pi \) times the distance to the pole. Conversely, if \( k = -1 \), and the Universe is open, we have \( \sigma > \chi \): the perimeter of a circle, or the area of a sphere, is larger than we would expect, compared with the distance from the center to its boundary.

Moving at light speed, in time \( \Delta t \) galaxy \( G \)'s light travels a distance \( c \Delta t = \mathcal{R}(t) \Delta \chi \) toward us. Using Equation 8.30 to relate \( \Delta t \) to the redshift, the total distance is

\[
\chi_e = \int_{t_e}^{t_0} \frac{c}{\mathcal{R}(t)} \, dt = \int_0^z \frac{c \, dz'}{\mathcal{R}(t_0) H(t)}.
\]

This integral can be done exactly when \( \Lambda = 0 \):

\[
\frac{\mathcal{R}(t_0) H_0}{c} \chi_e = \begin{cases} 
\int_0^z \frac{dz'}{(1+z')^{3/2}} = 2 \left( 1 - \frac{1}{\sqrt{1+z}} \right) & \text{for } \Omega_0 = 1, \\
\int_0^z \frac{dz'}{(1+z')^2} = \ln(1+z) & \text{for } \Omega_0 = 0.
\end{cases}
\]

Now, substituting for \( \chi_e \) in Equation 8.39 gives the comoving radius \( \sigma_e \):

\[
\mathcal{R}(t_0) \sigma_e = \begin{cases} 
\frac{2z}{H_0} \left[ 1 - \frac{1}{\sqrt{1+z}} \right] & \text{for } \Omega_0 = 1, \\
\frac{c}{H_0} \frac{z(1+z/2)}{1+z} & \text{for } \Omega_0 = 0.
\end{cases}
\]
8.3 Observing the earliest galaxies

Fig. 8.11. Left, the luminosity distance \( d_L \); and right, the angular-size distance \( d_A \), for a source at redshift \( z \); both are in units of \( c/H_0 \approx 3h^{-1} \) Gpc. The solid curve (\( \Omega_0 = 1 \)) and the dashed line (\( \Omega_0 = 0 \)) are for \( \Lambda = 0 \); the dotted line is for the benchmark model.

An excessive quantity of algebra yields the more general Mattig formula

\[
R(t_0)\sigma_e = \frac{c}{H_0} \frac{2}{\Omega_0^2(1+z)} \left[ \Omega_0 z + (\Omega_0 - 2)(\sqrt{1 + \Omega_0 z} - 1) \right]. \tag{8.42b}
\]

Notice that at large redshift we have \( R(t_0)\sigma_e \to 2c/(H_0\Omega_0) \). When the density is small, with \( \Omega_0 \ll 1 \), a more convenient form is

\[
R(t_0)\sigma_e = \frac{c}{H_0} \frac{z}{1 + \sqrt{1 + \Omega_0 z} + z} \tag{8.42c}
\]

For the flat Universe with \( \Omega_{\text{tot}} = 1 \) there is no general solution, but

\[
R(t_0)\sigma_e \approx 2c/\left( H_0 z \Omega_m^{0.4} \right) \text{ as } z \to \infty. \tag{8.43}
\]

Figure 8.11 shows \( d_L \) and \( d_A \) for some special cases. Nearby, both grow linearly with the redshift, according to Problem 8.21. But the luminosity distance \( d_L \) always increases faster than this linear relation, more strongly so when the density \( \Omega_{\text{tot}} \) is low: this rescues us from Olbers’ paradox. The angular-size distance \( d_A \) grows more slowly; it reaches a maximum at \( z \sim 1/\Omega_0 \) (or \( z \sim 1/\Omega_m \) in the flat model with \( \Lambda = 0 \)), and then declines. In our benchmark model, at redshifts \( z \gtrsim 1.5 \) a source looks larger as it is moved further away.

---

**Problem 8.20** Show from Equation 8.39 that \( \sigma \approx \chi \) when both are much less than unity, and that both \( d_A \) and \( d_L \) are then much less than \( c/H_0 \approx 2990h^{-1} \) Mpc. (We do not have to worry about the curvature of space when measuring nearby galaxies, just as we are not concerned with the Earth’s curvature when crossing a town.)
Problem 8.21 Use Equation 8.40 to show that, when the redshift \( z \ll 1 \), both \( d_L \) and \( d_A \) tend to the distance that we would normally calculate from the redshift \( z \): \( d_A = R(t_c)\sigma_e \rightarrow c/H_0 \approx d_L \). From Equation 8.42, show that, when \( \Lambda = 0 \) and the redshift is large, \( d_A \rightarrow 2c/(H_0 z \Omega_m) \).

Problem 8.22 With \( M_{\text{bol}} \odot = 4.75 \), \( L_\star = 2 \times 10^{10} L_\odot \), and \( H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1} \), show that, if \( \Lambda = 0 \), an \( L_\star \) galaxy at redshift \( z = 3 \) would have \( m_{\text{bol}} = 25.2 \) if \( \Omega_0 = 1 \), or \( m_{\text{bol}} = 26.6 \) for \( \Omega_0 = 0 \), whereas \( m_{\text{bol}} = 26.1 \) in the benchmark cosmology.

Problem 8.23 Show that, in a low-density Universe with \( \Omega_0 = 0 \), \( d_A \rightarrow c/(2H_0) \) at large redshifts. Use Figure 8.11 to show that the angle subtended by an object 10 kpc across is always at least 1.4 \( h \) arcsec for \( \Omega_0 = 0 \), and \( \gtrsim 3h \) arcsec if \( \Omega_0 = 1 \). (The star-forming regions of Figure 9.14 have radii of \( 0.1 - 0.2 \) arcseconds; so these bright patches are \( \lesssim 1 \) kpc across.)

### 8.3.2 Galaxy spectra and photometric redshifts

In practice we do not measure bolometric luminosities; we measure the apparent brightness in a specific band of wavelength or frequency. Cosmic expansion changes a galaxy’s color: when we observe in a particular waveband, the light we now see was radiated in a bluer part of the spectrum. Suppose that, at time \( t_c \), a galaxy that we see at redshift \( z \) has luminosity \( L_\lambda(t_c, \lambda) \Delta\lambda \) in the wavelength range from \( \lambda \) to \( \lambda + \Delta\lambda \). The apparent brightness \( F_\lambda \) in a bandpass \( \Delta\lambda \) that transmits all the light between wavelengths \( \lambda_1 \) and \( \lambda_2 \) is given by Equation 8.35, if we take the luminosity \( L \) to be the energy emitted at wavelengths that will pass through our bandpass when we receive it. So

\[
F_{\lambda} = \frac{1}{4\pi d_L^2} \int_{\lambda_1/(1+z)}^{\lambda_2/(1+z)} L_\lambda(t_c, \lambda) \, d\lambda
\]

For a nearby galaxy, Equation 1.15 tells us how to calculate its apparent magnitude at any distance. But at large redshift we must add two more terms: the apparent magnitude \( m_{\lambda BP} \) in this bandpass is

\[
m_{\lambda BP} = M_{\lambda BP} + 5 \log_{10} \left( \frac{d_L}{10 \text{ pc}} \right) + k_{\lambda BP}(z) + e_{\lambda BP}(z).
\]
8.3 Observing the earliest galaxies

The term $k(z)$, historically known as the *k correction*, represents the effect of shifting the galaxy’s light in wavelength. The *evolutionary* term $e(z)$, which we will discuss in Section 9.3, allows for changes in the galaxy’s luminosity between the time that its light was emitted and the present day.

From Equation 8.44, we have

$$k_{BP}(z) \equiv 2.5 \log_{10}(1 + z) - 2.5 \log_{10} \left\{ \frac{\int_{\lambda_1}^{\lambda_2} L_\lambda(\lambda/(1 + z), t_0) d\lambda}{\int_{\lambda_1}^{\lambda_2} L_\lambda(\lambda, t_0) d\lambda} \right\}, \quad (8.46)$$

where $L_\lambda(\lambda, t_0)$ is the present-day spectrum. Of course, we cannot measure the spectrum of a distant galaxy as it is today; that light is still on its way to us. But we can calculate how a present-day galaxy would appear if we observed it at a redshift $z$. Figure 8.12 shows that, if we observe in the $B$ band at 4400 Å, an elliptical galaxy will dim rapidly as its redshift increases, since our passband moves into the ultraviolet region where its stars emit little light. A starburst galaxy will fade much less, because its hot young stars are bright at short wavelengths.

At $z \approx 1$ both galaxies fade much less in the red light of the $I$ band at 8000 Å than in the $B$ band. So the $B - I$ color becomes progressively redder at higher redshift. At redshifts $z \gtrsim 2$, the light of both galaxies has largely moved into the infrared region.

**Problem 8.24** If a galaxy emits a spectrum $L_\nu \propto \nu^{-\alpha}$, show that $L_\lambda \propto \lambda^{\alpha - 2}$, and that $k(z) = (\alpha - 1) \times 2.5 \log_{10}(1 + z)$. The $k$ correction is zero if $\nu L_\nu$ is nearly constant so that $\alpha \approx 1$, as it is for many quasars (see Section 9.1). When the spectrum declines more rapidly than $L_\nu \propto \nu^{-1}$ toward high frequencies, $k(z) > 0$, and the object appears dimmer.

A *photometric redshift* is an estimate of a galaxy’s redshift made by comparing its apparent brightness in several bandpasses with that predicted by a diagram like Figure 8.12. For example, an elliptical galaxy at $z \sim 0.5$ has already become very red in the $B - I$ color, but it is less so in $V - I$. At $z \sim 1$, it is fading rapidly in the $I$ band, so the $I - H$ color starts to redden. With 17 filters at wavelengths from 3640 Å to 9140 Å, the COMBO-17 team could estimate redshifts to $\Delta z \approx 0.05$ over the range $0.2 < z < 1.2$. The most spectacular use of photometric redshifts has been to find galaxies at $z > 3$. These Lyman break galaxies almost disappear at wavelengths less than $912(1 + z)$ Å, where intergalactic atoms of neutral hydrogen absorb nearly all their light (see Section 9.4).

**Problem 8.25** Explain why, if we base a galaxy survey on images in the $B$ band, then at $z \gtrsim 0.5$ we will fail to include many of the systems with red spectra similar to present-day ellipticals.
8.3.3 How many galaxies? Space densities

To trace the formation of galaxies through cosmic history, we must take account of the expansion in counting the number within any given volume. The number of objects that we will see between redshifts \( z \) and \( z + \Delta z \) is proportional to the corresponding volume of space \( \Delta V \). This is just the product of the area \( A(\sigma_e, t_e) = 4\pi R^2(t_e)\sigma_e^2 \) of the sphere containing the galaxy at the time its light was emitted and the distance \( c|\Delta t_e| \) that the light travels toward us in the time corresponding to this interval in redshift. From Equations 8.30 and 8.31, we have

\[
\frac{A c|\Delta t_e|}{\Delta z} \approx \frac{dV}{dz} = \frac{4\pi cR^2(t_0)\sigma_e^2}{H(z)(1 + z)^3},
\]

where we replaced \( R(t_e) \) by \( R(t_0)/(1 + z) \) in the last step.

The volume \( \Delta V \) at redshift \( z \) will expand to fill a volume \( \Delta V(1 + z)^3 \) by the present day: we refer to \( \Delta V(1 + z)^3 \) as the comoving volume. If the number of galaxies in the Universe had always remained constant, then the comoving density, the number in each unit of comoving volume, would not change. If there are presently \( n_0 \) of a particular galaxy type in each cubic gigaparsec, then between
8.3 Observing the earliest galaxies

redshift \( z \) and \( z + \Delta z \) we would expect \( (d\mathcal{N}_\Delta//dz)\Delta z \) of them, where

\[
\frac{d\mathcal{N}_\Delta}{dz} = n_0(1+z)^3 \frac{dV}{dz} = n_0c \frac{4\pi R^2(t_0)\sigma_e^2}{H(z)}.
\] (8.48)

Comparing the measured number of galaxies \( d\mathcal{N}/dz \) at each redshift with \( d\mathcal{N}_\Delta//dz \) from Equation 8.48 tells us how the comoving density has changed.

The left panel of Figure 8.8 shows the comoving volume \( (1+z)^3 dV/dz \) between redshifts \( z \) and \( z + \Delta z \). It is much larger in the open Universe with \( \Lambda = \Omega_0 = 0 \) than it is in the flat model with \( \Omega_0 = 1 \). So we expect to see relatively more galaxies at high redshift in the open model. The benchmark model has slightly more volume at low redshift than the \( \Lambda = \Omega_0 = 0 \) model, but less at \( z \approx 2 \).

**Problem 8.26** Use Equations 8.23 and 8.42 to show that, if \( \Omega_0 = 0 \), then redshift \( z = 5 \) corresponds to \( R(t_0)\sigma_e = 2.92 c/H_0 \), whereas for \( \Omega_0 = 1 \), \( R(t_0)\sigma_e = 1.18 c/H_0 \). For any given density \( n(z) \), use Equation 8.48 to show that, if \( \Omega_0 = 0 \), then at \( z = 5 \) we would expect to find roughly 15 times as many objects within a small redshift range \( \Delta z \) as we would see if \( \Omega_0 = 1 \). What is this ratio at \( z = 3 \)?

Quasars, the extremely luminous ‘active’ nuclei of galaxies which we discuss in the following chapter, are so bright that we can see them across most of the observable Universe. They also have strong emission lines that make it easy to measure their redshifts. Figure 1.16 told us that each cubic gigaparsec now contains \( \sim 10^6 \) galaxies with \( L \approx L_\star \), where \( L_\star \approx 2 \times 10^{10} L_\odot \) is the luminosity of a bright galaxy defined by Equation 1.24. At present, each cubic gigaparsec contains about one very luminous quasar with \( L \gg 100 L_\star \); bright quasars are much rarer than luminous galaxies. But Figure 8.13 shows that, at redshifts \( z \approx 2 \), the brightest quasars were about 100 times more common than they are today. There was roughly one quasar for every 10,000 present-day giant galaxies. What happened to them?

If quasars represent the youth of a galactic nucleus, then at least one in 10,000 luminous galaxies must have been bright quasars in the past. The fraction could be as high as 100% if the nuclear activity lasts much less than a gigayear. A period with a quasar nucleus might be a normal part of a galaxy’s early development.

We will see in Section 9.1 that quasars shine by the energy released as gas falls into a hugely massive black hole of \( \sim 10^9 M_\odot \). In Section 6.4 we found that today’s luminous galaxies harbor massive black holes at their centers; perhaps these remain from an early quasar phase. It takes time to build up the black hole as it consumes gas, so quasars are rare during the first quarter of cosmic history, before \( z \sim 2 \). It is perhaps more surprising that we begin to see them already at \( z > 6 \), less than a gigayear after the Big Bang.
8.4 Growth of structure: from small beginnings

The cosmic background radiation is almost, but not quite, uniform: across the sky, its temperature differs by a few parts in $10^5$. These tiny differences tell us how lumpy the cosmos was at the time $t_{\text{rec}}$ of recombination, when the radiation cooled enough for neutral atoms to form. Quantum fluctuations in the field responsible for inflation left their imprint as irregularities in the density of matter and radiation. Most versions of inflation predict that fluctuations should obey the random-phase hypothesis, and that the power spectrum $P(k) \propto k$ (see Problem 8.7): we will call these the benchmark initial fluctuations. Equation 8.4 then tells us that the density varies most strongly on small spatial scales or large $k$.

The largest features, extending a degree or more across the sky, tell us about that early physics. Smaller-scale irregularities are modified by the excess gravitational pull toward regions of high density and by the pressure of that denser gas. Observing them tells us about the geometry of the Universe and its matter content. After recombination, dense regions rapidly became yet denser as surrounding matter fell into them. By observing the peculiar motions of infalling galaxies, we can probe the large-scale distribution of mass today and compare it with what is revealed by the light of the galaxies.

8.4.1 Fluctuations in the cosmic microwave background radiation

How did the distribution of matter affect the cosmic background radiation as we observe it today? To reach us from an overdense region, radiation has to climb
8.4 Growth of structure: from small beginnings

out of a deeper gravitational potential. In doing this, it suffers a gravitational redshift proportional to $\Delta \Phi_g$, the excess depth of the potential: its temperature $T$ changes by $\Delta T$, where $\Delta T / T \sim \Delta \Phi_g / c^2$. The temperature is reduced where the potential is unusually deep, since $\Delta \Phi_g$ is negative there. But time also runs more slowly within the denser region by a fraction $\Delta t / t \sim \Delta \Phi_g / c^2$, so we see the gas at an earlier time when it was hotter. The radiation temperature decreases as $T \propto 1 / a(t)$, so

$$\frac{\Delta T}{T} = - \frac{\Delta a}{a} = - \frac{2 \Delta t}{3 t} = - \frac{2 \Delta \Phi_g}{3 c^2},$$

(8.49)

where we have used $a \propto t^{2/3}$ from Equation 8.27. This partly cancels out the gravitational redshift to give $\Delta T / T \sim \Delta \Phi_g / (3c^2)$. At these early times, the average density $\bar{\rho}$ is very nearly equal to the critical density of Equation 8.21. If our region has density $\bar{\rho}(1 + \delta)$ and radius $R$, its excess mass is $\Delta M = 4\pi \bar{\rho} R^3 \delta / 3$. We can write

$$3c^2 \frac{\Delta T}{T} = \Delta \Phi_g \sim - \frac{2G \Delta M}{R} = - \frac{8\pi}{3} G \bar{\rho} R^2 \delta \approx - \delta(t) [\dot{H}(t) R^2].$$

(8.50)

Radiation reaching us from denser regions is cooler.

The best current measurements of the cosmic microwave background on scales larger than $0.3^\circ$ are from the WMAP satellite, which was launched in June 2001. WMAP’s first observing year confirmed that the background radiation has the form of a blackbody everywhere on the sky: only its temperature differs slightly from point to point. This is exactly what we would expect if it was affected by non-uniformities in the matter density. We can describe the temperature variations by choosing some polar coordinates $\theta, \phi$ on the sky. As we look in a given direction, we can write the difference $\Delta T$ from the mean temperature $T$ using the spherical harmonic functions $Y_l^m$:

$$\Delta T(\theta, \phi) = \sum_{l>1} \sum_{-l \leq m \leq l} a_l^m Y_l^m(\theta, \phi).$$

(8.51)

Since $Y_l^m$ has $l$ zeros as the angle $\theta$ varies from 0 to $\pi$, the $a_l^m$ measure an average temperature difference between points separated by an angle $\left(\frac{180}{l}\right)^\circ$ on the sky. Apart from the $l = 1$ terms which reflect our motion relative to the background radiation, all the $a_l^m$ must average to zero; their squared average measures how strongly $T$ fluctuates across the sky. Theorists aim to predict $C_l = \langle |a_l^m|^2 \rangle$ averaged over all the $m$-values, since this does not depend on which direction we chose for $\theta = 0$, our ‘north pole’. Figure 8.14 shows $\Delta T(l)$, defined by $\Delta T^2 = T^2 l(l + 1) C_l / (2\pi)$.

Gravity, which makes denser regions even denser, and pressure forces that tend to even out the density, have modified the fluctuations left behind after inflation. These forces cannot propagate faster than light, so they act only within the horizon...
temperature fluctuations $\Delta T$ in the cosmic microwave background: triangular points combine data from many experiments, circles are from WMAP. Horizontal bars show the range of angular scales. Curves show predictions for the benchmark model (solid), for a flat model with half as many baryons (dotted), and for $\Omega_0 = 0.3$, $\Lambda = 0$ (dashed). The second, third, and subsequent peaks correspond to regions that sound waves could cross twice, three times, etc., before recombination. When $l$ is small, we have few $a_l^m$ to average over (only five for $l = 2$), and vertical bars indicate larger uncertainties – M. Tegmark, CMBFAST.

scale of Problem 8.14. When the gas became transparent, the comoving distance $\sigma_H$ to the horizon was

$$R(t_{rec})\sigma_H = 3c_t_{rec} = \frac{2c}{H(t_{rec})} \approx \frac{2c}{H_0 \sqrt{\Omega_m(1 + z_{rec})^{3/2}}} ,$$

where we used Equation 8.26 in the last step. (Why can we ignore $\Omega_\Lambda$?) A region of this size will expand to $184/(h^2\Omega_m)^{1/2}$ Mpc by today. Because inflation left us with $P(k) \propto k$, on larger scales (at small $l$), we expect $\Delta T(l)$ to rise smoothly as $l$ increases.

The angle $\theta_H$ that the horizon covers on the sky depends on $\Omega_\Lambda$ and $\Omega_m$ through the angular-size distance $d_A$. When $\Lambda = 0$ and $\Omega_0 z \gg 1$, Equation 8.42 tells us that $d_A \rightarrow 2c/(H_0 z \Omega_0)$. So only points separated by less than the angle

$$\theta_H \approx \frac{R(t_{rec}) \sigma_H}{d_A(t_{rec})} \approx \sqrt{\frac{\Omega_0}{z_{rec}}} \approx 2^\circ \times \sqrt{\Omega_0}$$

can communicate before time $t_{rec}$. The lower the matter density, the smaller this angle should be. Detailed calculation shows that, if $\Omega_0 = 1$, $\Delta T$ is largest on scales just less than a degree, where we see the main acoustic peak in Figure 8.14.
8.4 Growth of structure: from small beginnings

A model with $\Omega_0 = 0.3$ places the peak at roughly half this angle. Setting $\Omega_m + \Omega_\Lambda = 1$ changes the way that the distance radius $d_A$ depends on the matter density. From Equation 8.43 we have $d_A \propto 1/\Omega_m^{0.4}$ at large redshift, so the angular size of the ripples is almost independent of $\Omega_m$. The observed position of the acoustic peak is the most powerful current evidence in favor of dark energy.

We will see in the following section that, before recombination, irregularities built up most strongly in the dark matter. The mixed fluid of baryons and radiation then simply fell into the denser regions under gravity. The maximum distance through which that mixed fluid can fall by $t_{rec}$ sets the position of the first peak in $\Delta T_a$, at $l = 220$. In the benchmark cosmology, this distance corresponds to 105 Mpc today, or a cube containing $2.5 \times 10^{16} M_\odot$. The second peak, at $l = 540$, corresponds to a smaller lump of dark matter, where the fluid has time to fall in and be pushed out again by its own increased pressure. The third peak corresponds to ‘in–out–in’, the fourth to ‘in–out–in–out’, and so on: hence the label of ‘acoustic peaks’. The more dark matter is present, the stronger its gravity causes its irregularities to become, and so the greater the height of the main peak. The mass of the baryonic matter ‘helps’ the baryon–radiation fluid to fall into dense regions of dark matter, but hinders its ‘bouncing’ out again; this strengthens the odd-numbered peaks relative to the even peaks. The benchmark model, with $\Omega_B \approx 0.045$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$, gives correct predictions for the abundance of deuterium and lithium (see Section 1.5), the motions of the galaxies, and fluctuations in the cosmic background radiation.

Further reading: see Chapter 6 of the book by Padmanabhan.

8.4.2 Peculiar motions of galaxies

One way that we can explore the largest structures is to map out the galaxies, as in Figure 8.3; but this samples only the luminous matter. Another is to look at the peculiar motions of galaxies, their deviation from the uniform flow described by Equation 8.8. Peculiar motions grow because of the extra tug of gravity from denser regions. In the Local Group, the Milky Way and the Andromeda galaxy M31 approach each other under their mutual gravitational attraction (Section 4.5), while groups of galaxies fall into nearby clusters (Section 7.2). We saw how to use these motions to weigh the groups and clusters. Similarly, we can use the observed peculiar motions on larger scales to reconstruct the distribution of mass, most of which is dark.

We can see the peculiar motions of the nearby elliptical galaxies in Figure 8.2. Although the Fornax cluster is roughly as far away as the Virgo cluster, the galaxies of Fornax on average are moving more rapidly away from us. It appears that the Local Group, and the galaxies nearby, are falling toward the complex of galaxies around Virgo. To examine the Virgo-centric infall, in Figure 8.15 we look at the
average radial velocity with which each group of galaxies in Figure 8.2 recedes from us. The peculiar velocities of individual galaxies are affected by their orbits within the group, but averaging over the whole group should reveal the largerscale motions. These velocities are plotted in the left panel. The two largest white symbols represent the two clumps of Virgo cluster galaxies, around M86 and M49. The other big symbols, indicating groups close to Virgo, fall below the general linear trend.

The right panel of Figure 8.15 shows the result of subtracting out Virgocentric inflow according to a simple model, which predicts an infall speed of 270 km s\(^{-1}\) at our position. We now see roughly \(V_r \propto d\). Within about 25 Mpc of Virgo, most of the plotted values deviate from the linear trend by less than 100 km s\(^{-1}\). Peculiar motions complicate our attempts to measure \(H_0\). If we tried to do this by finding distances and velocities of galaxies in the direction of Virgo, we would underestimate the Hubble constant, because Virgocentric inflow partially cancels out the cosmic expansion. But if we had observed galaxies in the opposite direction, our value for \(H_0\) would be too high.

The best-measured peculiar motion is that of the Local Group, determined from the Sun’s velocity relative to the cosmic microwave radiation: recall Section 1.5. The Local Group now moves with \(V_{pec} \approx 630\) km s\(^{-1}\) in the direction \((l, b) \approx (276^\circ, 30^\circ)\). Most of that peculiar motion seems to be caused by the gravitational pull of very distant matter, tugging at both us and the Virgo cluster. The velocities of galaxies furthest from the Virgo cluster, which are mostly on
the opposite side of the sky, lie mainly above the sloping line in the right panel of Figure 8.15. This is what we would expect if more distant matter were pulling Virgo and Fornax apart.

Both the local velocity dispersion and the Local Group’s motion toward the Virgo cluster are significantly less than our motion relative to the cosmic microwave background. The flow of galaxies through space is ‘cold’ on small scales: galaxies within tens of megaparsecs of each other share a large fraction of their peculiar velocity.

Problem 8.27 Here you use Monte Carlo simulation to show that the peculiar velocities of nearby galaxies must be very close to that of the Milky Way, or Hubble could never have discovered the cosmic expansion from his sample of 22 local galaxies.

Your model sky consists of galaxies in regions A (1 Mpc < d < 3 Mpc), B (3 Mpc < d < 5 Mpc), C (5 Mpc < d < 7 Mpc), and D (7 Mpc < d < 9 Mpc). If the density is uniform, and you have four galaxies in region B, how many are in regions A, C, and D (round to the nearest integer)? For simplicity, put all the objects in region A at d = 2 Mpc, those in B at 4 Mpc, those in C at 6 Mpc, and those in D at 8 Mpc. Now assign peculiar velocities at random to the galaxies. For each one, roll a die, note the number N on the upturned face, and give your galaxy a radial velocity

\[ V_r = H_0 d + (N - 3.5) \times 350 \text{ km s}^{-1} \],

taking \( H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1} \). (If you like to program you can use more galaxies; place them randomly in space, and choose the peculiar velocities from a Gaussian random distribution with zero mean and standard deviation of 600 km s\(^{-1}\).)

Plot both \( V_r \) and also the average velocity in each region against the distance \( d \); is there a clear trend? How many of your model galaxies have negative radial velocities? How does your plot compare with the right panel of Figure 8.15?

Hubble found no galaxies beyond the Local Group that are approaching us.

8.4.3 How do peculiar velocities build up?

Peculiar velocities tend to die away as the Universe expands, because a moving galaxy keeps overtaking others, until it reaches the region where its motion matches that of the cosmic expansion. We can imagine two nearby comoving observers, P and Q, at rest relative to the background radiation; they recede from each other only because of cosmic expansion. A galaxy passes observer P heading toward Q with a peculiar motion \( V_{\text{pec}} \), and arrives there after a time \( \approx d/V_{\text{pec}} \). If P and Q are close enough that \( V_{\text{pec}} \gg H(t) d \), their separation remains almost constant as the galaxy travels between them. But relative to observer Q, the galaxy moves only at speed \( V_{\text{pec}} - H(t) d \). The galaxy’s speed, relative to a comoving
The large-scale distribution of galaxies

observer at its current position, has decreased at the rate

\[
\frac{dV_{\text{pec}}}{dt} = - \frac{H(t) \delta}{dV_{\text{pec}}} = -V_{\text{pec}} \frac{\dot{R}(t)}{R(t)}. \tag{8.54}
\]

Integrating this shows that \( V_{\text{pec}} \propto 1/R(t) \); a galaxy’s peculiar velocity \( V_{\text{pec}} \) falls in exactly the same way as the momentum of a photon is reduced according to Equation 8.12.

If peculiar velocities simply decreased according to Equation 8.54, then shortly after recombination at \( z \approx 1100 \), the material of the Local Group would have been moving at close to the speed of light. But this would have caused shocks in the gas and huge distortions in the cosmic microwave background. In fact, the peculiar motions of the galaxies were generated quite recently, by their mutual gravitational attraction. When some part of the Universe contains more matter than average, its increased gravity brakes the expansion more strongly. Where there is less matter than average, the expansion is faster; the region becomes even more diffuse relative to its surroundings. So the galaxies move relative to the cosmic background: they acquire peculiar motions.

To calculate how this happens, suppose that the average density of matter is \( \bar{\rho}_m(t) \), and the average expansion is described by the scale factor \( \bar{a}(t) \) and the Hubble parameter \( H(t) \). Locally, within the volume we are studying we can write

\[
\rho_m(t) = \bar{\rho}_m(t)[1 + \delta(t)], \quad \text{and} \quad a(t) = \bar{a}(t)[1 - \epsilon(t)]. \tag{8.55}
\]

If our region is approximately spherical, the matter outside will not exert any gravitational force within it; it will behave just like part of a denser, more slowly expanding, cosmos. Where \( \Omega_m[1 + \delta(t)] > 1 \) so the local density exceeds the critical value, expansion can be halted to form bound groups and clusters of galaxies.

Life is much simpler if we stay in the linear regime, where \( \delta \) and \( \epsilon \) of Equation 8.55 are much less than unity. We saw in the discussion following Equation 8.4 that this applies to structures with sizes larger than about \( 8h^{-1}\) Mpc: their density differs by only a small fraction from the cosmic average. When we substitute the expressions for \( \rho_m(t) \) and \( a(t) \) into Equation 8.25, we can then ignore terms in \( \delta^2 \), \( \delta \epsilon \), \( \epsilon^2 \), and higher powers of these variables. Remembering that \( a(t)H(t) = \dot{a}(t) \) and that terms involving only barred average quantities will cancel out, Equation 8.25 becomes

\[
\Delta[H_0^2(1 - \Omega_m)] + 2 \frac{d\dot{a}}{dt} \frac{d}{dt}[\bar{a} \epsilon(t)] + \frac{8\pi G}{3} \bar{\rho}(t) \bar{a}^2(t)[\delta(t) - 2\epsilon(t)] = 0. \tag{8.56}
\]

Here the first term represents the change in the present density and expansion rate within our denser region.
8.4 Growth of structure: from small beginnings

We saw from Figure 8.7 that, for most of the period during which galaxy clusters and groups were forming, dark energy was not important, and we can simply use Equation 8.27 to describe the average expansion \( \bar{a}(t) \). While the Universe is matter-dominated, \( \rho a^3 \) is constant, so \( \delta = 3\epsilon \). Then

\[
\delta \propto t^{2/3} \propto \bar{a}(t) \quad \text{as long as } 1 + z \gg (1 - \Omega_{\text{tot}}) / \Omega_m \cdot (\Omega_\Lambda / \Omega_m)^{1/3} \tag{8.57}
\]

is a ‘growing’ solution to Equation 8.56 (substitute back to check!). Early on, the contrast \( \delta \) grows proportionally to \( R(t) \). If \( \Omega_m + \Omega_\Lambda < 1 \), then at some point the first condition on \( z \) is violated, and the average motion becomes \( \bar{a} \propto t \): matter coasts outward with constant speed. Its gravity is too weak to have any effect on the expansion, so \( \delta \) remains fixed: the structure freezes out. In a flat model with \( \Omega_{\text{tot}} = 1 \), growth continues until \( 1 + z \sim (\Omega_\Lambda / \Omega_m)^{1/3} \). In the benchmark model, large structures continued to grow until very recently, at \( z \sim 0.3 \). In a low-density Universe with \( \Omega_m = 0.3 \) and \( \Omega_\Lambda = 0 \), they would have ceased to become denser around redshift \( z \sim 2 \).

**Further reading:** On peculiar motions, see Chapter 4 of Padmanabhan’s book.

8.4.4 Weighing galaxy clusters with peculiar motions

Any denser-than-average region pulled the surrounding galaxies more strongly toward it. While the fractional deviations \( \delta (x, t) \) from uniform density remain small, Equation 8.57 tells us that, over a given time, \( \delta (x) \) increased by an equal factor everywhere. Because the pull on a galaxy from each overdense region increases in the same proportion, its acceleration, and hence its peculiar velocity, is always parallel to the local gravitational force. So, by measuring peculiar motions, we can reconstruct the force vector, and hence the distribution of mass.

To see how this works, we can write the velocity \( u(x, t) \) of matter at point \( x \) as the sum of the average cosmic expansion directly away from the origin and a peculiar velocity \( v \):

\[
u(x, t) = \bar{H}(t)x + v(x, t). \tag{8.58}\]

The equation of mass conservation relates the velocity field \( u(x, t) \) to the density, which we write as \( \rho(x, t) = \bar{\rho}(t)[1 + \delta(x, t)] \):

\[
\left( \frac{\partial \rho}{\partial t} \right)_x + \nabla_x \cdot \rho \, u = 0. \tag{8.59}
\]

Remembering that terms involving only the barred average quantities will cancel out, and dropping terms in \( \delta^2, \delta v, \) and \( v^2 \), we have

\[
\left( \frac{\partial \delta}{\partial t} \right)_x + \bar{H}(t)x \cdot \nabla_x \delta + \nabla_x v = 0. \tag{8.60}
\]
The large-scale distribution of galaxies

Setting $x = \bar{a}(t)r$, we switch to the coordinate $r$ comoving with the average expansion. The time derivative following a point at fixed $r$ is

$$\left( \frac{\partial}{\partial t} \right)_r = \left( \frac{\partial}{\partial t} \right)_x + \bar{H}(t)x \cdot \nabla_x, \quad (8.61)$$

and, since $\bar{a}(t)\nabla_r = \nabla_x$, Equation 8.60 simplifies to

$$\left( \frac{\partial \delta}{\partial t} \right)_r + \nabla_x v = 0. \quad (8.62)$$

For a small enough volume, if we assume that the Universe beyond is homogeneous and isotropic, we can use Newton’s laws to calculate the gravitational potential $\Phi_g$ corresponding to local deviations from the average density $\bar{\rho}$. The gravitational force $F_g = -\nabla\Phi_g = dv(x, t)/dt$, so we have $0 = d(\nabla \times v)/dt$. Peculiar motions that have grown in this way from small initial fluctuations thus have $\nabla \times v \approx 0$, and we can define a velocity potential $\Phi_v$ such that $v = \nabla_x \Phi_v$. Rewriting Equation 8.62 in terms of $\Phi_v$ gives

$$\nabla^2 \Phi_v = -\left( \frac{\partial \delta}{\partial t} \right)_r. \quad (8.63)$$

Equation 3.9, Poisson’s equation, tells us that

$$\nabla^2 \Phi_g = -\nabla \cdot F_g = 4\pi G \bar{\rho} \delta(x, t) \quad (8.64)$$

– which looks suspiciously like the equation for $\Phi_v$. Equation 8.57 assures us that all perturbations grow at the same rate: if $\delta$ is twice as large, then so is $\dot{\delta}$. Thus $\delta(x, t) \propto \dot{\delta}(x, t)/\dot{\bar{a}}t$, and the right-hand sides of Equations 8.63 and 8.64 are proportional to each other. Then, as long as both $v(x, t)$ and $F_g$ diminish to zero as $|x|$ increases, they must also be proportional: the peculiar velocity is in the same direction as the force resulting from local concentrations of matter. On dividing the right-hand side of Equations 8.63 by that of 8.64, we find

$$\frac{|v(x, t)|}{|F_g|} = \frac{\bar{H}(t)f}{4\pi G \bar{\rho}(t)}, \quad \text{where } f = \frac{\bar{a}(t)}{\delta} \int \frac{\partial \delta}{\partial t} \frac{d\bar{a}}{dt}. \quad (8.65)$$

From Equation 8.57, in a matter-dominated Universe we have $f = 1$ for $\Omega_m \approx 1$, and $f \to 0$ as $\Omega_m \to 0$. In general, $f(\Omega) \approx \Omega^{0.6}$ is a good approximation when dark energy is absent, whereas in a flat Universe with $\Omega_\Lambda + \Omega_m = 1$, $f(\Omega) \approx \Omega_m^{0.23}$. Using Equation 3.5 for the force, we can write the peculiar velocity as

$$v(x, t) = \frac{\bar{H}(t)f(\Omega)}{4\pi} \int \frac{\delta(x')(x - x')}{|x - x'|^3} \, d^3x'. \quad (8.66)$$
### Problem 8.28
Show that, if the density is uniform apart from a single overdense lump at \(x = 0\), then distant galaxies move toward the origin with \(v(x, t) \propto 1/x^2\).

### Problem 8.29
In the expanding (comoving) coordinate \(r\), show that

\[
v(r, t) = \frac{\bar{H}(t)f(\Omega_0)\bar{a}(t)}{4\pi} \int \frac{\delta(r')|r - r'|^3}{|r - r'|^3} \mathrm{d}^3r'.
\]  

(8.67)

Show that, while it is early enough that we can use Equation 8.57 for \(\delta(r)\), the peculiar velocity \(v \propto t^{1/3}\). (Why did we have to transform to comoving coordinates to apply Equation 8.57?)

So, if we can measure the overdensity \(\delta(x)\) of the nearby rich galaxy clusters, and the peculiar velocities of the galaxies around them, we should be able to test Equation 8.66, and solve for the matter density \(\Omega_m\). First, we determine the average peculiar motion \(v(x)\) of our galaxies. We must assume that the Universe is homogeneous and isotropic on even larger scales, so that forces from galaxies outside our survey volume will average to zero. Inverting Equation 8.66 should then yield the product \(f(\Omega_0) \cdot \delta(x)\), from which we can find \(\Omega_m\).

But the mass distributions predicted from measured peculiar velocities do not match the observed clustering of galaxies very well. Alternatively, we could say that the forces calculated from the galaxies at their observed positions do not yield the measured peculiar motions. The pull of matter outside the volume of our present surveys appears to be significant. In particular, we still do not know that concentration of matter is responsible for most of the Local Group’s peculiar motion of \(\sim 600\) km s\(^{-1}\). Work is under way on this problem, and galaxy surveys are being extended as techniques for finding distances improve.

Locally, we can use the crude model of Figure 8.15 for the Virgocentric infall to estimate the mass density \(\Omega_m\). Let \(d_V \approx 16\) Mpc be the distance of the Local Group from the center of the Virgo cluster. Within a sphere of radius \(d_V\) about the cluster’s center, the density of luminous galaxies is roughly 2.4 times the mean; if the mass density is increased by the same factor, then the overdensity \(\delta \approx 1.4\). Although Equation 8.65 was derived for \(\delta \ll 1\), we can use it to make a rough calculation of \(f(\Omega)\).

Assuming that the Virgo cluster is roughly spherical, the additional gravitational pull on the Local Group is \(F_g \approx 4\pi Gd_V\bar{\rho}\delta/3\), just as if all the cluster’s mass had been concentrated at its center. So our peculiar motion toward Virgo is

\[
|v_{LG}| \approx \frac{(H_0d_V)\Omega_m^{0.6}\delta}{3} \approx 270\text{ km s}^{-1}.
\]  

(8.68)

Cosmic expansion is pulling the cluster away from us at a speed \(H_0d_V \approx 1400\) km s\(^{-1}\), so this yields \(\Omega_m \approx 0.2\), in reasonable agreement with the benchmark model.
8.4.5 Tidal torques: how did galaxies get their spin?

The Sun rotates for the same reason that water swirls around the plug-hole as it runs out of a sink. The material originally had a small amount of angular momentum \( \rho \mathbf{x} \times \mathbf{v} \) about its center in a random sense. This is approximately conserved as the fluid is drawn radially inward, so as \(|\mathbf{x}|\) decreases the rotation described by \( \mathbf{v} \) must speed up. But galaxies and clusters do not owe their rotation to early random motions; this peculiar motion arises from irregular lumps of matter pulling on each other by gravity, as illustrated in Figure 4.13.

In Problem 8.29 we saw that, while the Universe is matter-dominated, peculiar velocities grow as \( t^{1/3} \), while the distance \( d \) between galaxies follows \( a(t) \propto t^{2/3} \). So angular momentum builds up as \( d \times v \propto t \) as long as we remain in the linear regime with \( \delta(t) \ll 1 \). It stops increasing when the dense region starts to collapse on itself, as we will discuss in Section 8.5. The denser the initial lump, the sooner it collapses and the less time it has to spin up. But tidal torques are stronger in denser regions, so, in a cosmos filled with cold dark matter, objects acquire the same average angular momentum in relation to their mass and energy.

To measure how important a galaxy’s angular momentum is, we note that a galaxy of radius \( R \), mass \( M \), and angular momentum \( L \) will rotate with angular speed \( \omega \sim L/(MR^2) \). The angular speed \( \omega_c \) of a circular orbit at radius \( R \) is given by \( \omega_c^2 R \sim G M / R^3 \). The energy \( E \sim -G M^2 / R \) (see Problem 3.36 and recall the virial theorem). So the ratio

\[
\lambda = \frac{\omega}{\omega_c} = \frac{L}{MR^2} \times \frac{R^{3/2}}{\sqrt{GM}} = \frac{L|E|^{1/2}}{GM^{5/2}}
\]

(8.69)

tells us how far the galaxy is supported against collapse by rotation, rather than pressure or random motion of its stars. Gravitational \( N \)-body simulations show that the distribution of galaxies we observe would not spin up collapsing lumps very strongly: we expect \( 0.01 < \lambda < 0.1 \). This is similar to what we see in elliptical galaxies, but disk galaxies like our Milky Way have \( \lambda \approx 0.5 \). The parameter \( \lambda \) can increase if material loses energy to move inward, as a gas disk can do by radiation.

This argument already tells us that the Milky Way has a dark halo – otherwise its disk would not have time to form. Without a halo, \( L \) and \( M \) do not change as the proto-disk moves inward, so its radius must shrink 100-fold to increase \( \lambda \) by the same factor. Disk material near the Sun must originate 800 kpc from the center, but the mass \( M(<R) \) within that radius would be just what now lies between the Sun’s orbit and the Galactic center. Equation 3.20 shows that the orbital period would then be 1000 times longer than that in the Sun’s current orbit, or about 240 Gyr. The Galaxy would shrink at roughly the same rate; it would take several times longer than the age of the Universe to make the disk.

But in Problem 3.5 we saw that the Milky Way has \( M/L \approx 50 \); at least 90% of its mass is dark. Because the dark halo cannot lose energy and shrink, the gas that
8.5 Growth of structure: clusters, walls, and voids

is to become the disk originates closer to the center by a factor $M_{\text{disk}}/M_{\text{total}}$. So our disk had to collapse only to a tenth of its original size to reach $\lambda \approx 0.5$. Since the infall and orbital speeds are set by the dark halo, they would have been near today’s values. Shrinking at 200 km s$^{-1}$ from a radius of 80 kpc, the disk could have formed in $\approx 2$ Gyr.

Further reading: see Chapter 8 of Padmanabhan’s book.

8.5 Growth of structure: clusters, walls, and voids

The galaxy clusters and huge walls that we see in Figure 8.3 are visible because the density of luminous matter in them is a few times greater than that in the surrounding regions. If galaxies trace out the mass density, then the fractional variations in density are now large: in the language of Equation 8.55, $\delta(t_0) \approx 1$. How did the small fluctuations that we examined in Section 8.4 develop into the structure that we now see?

8.5.1 Pressure battles gravity: the Jeans mass

Objects like stars are supported by gas pressure, which counteracts the inward pull of gravity. The larger a body is, the more likely it is that gravity will win the fight against the outward forces holding it up. In life, the giant insects of horror movies would be crushed by their own weight. For a spherical cloud of gas, we can estimate the potential energy $\mathcal{P}E$ using the result of Problem 3.11 for a uniform sphere of radius $r$ and density $\rho$. We then compare it with the thermal energy $\mathcal{K}E$.

The sound speed $c_s$ in a gas is close to the average speed of motion of the particles along one direction, so we can write

$$\mathcal{P}E = -\frac{1}{2} \rho(x) \Phi(x) d^3 x \approx -\frac{16\pi^2}{15} G \rho^2 r^5,$$

and $\mathcal{K}E \approx \frac{3}{2} \frac{c_s^2}{4\pi r^3 \rho}$. (8.70)

In equilibrium the virial theorem, Equation 3.44, requires $|\mathcal{P}E| = 2\mathcal{K}E$; we might expect the cloud to collapse if the kinetic energy is less than this. That always happens if the cloud is big enough: $\mathcal{K}E < |\mathcal{P}E|/2$ when

$$2r \approx \sqrt{\frac{15}{\pi}} \sqrt{\frac{c_s^2}{G \rho}} \approx \lambda_J,$$

where $\lambda_J \equiv c_s \sqrt{\frac{\pi}{G \rho}}$. (8.71)

The length $\lambda_J$ is called the Jeans length. When a gas cloud is compressed, its internal pressure rises and tends to cause expansion, but the inward pull of gravity also strengthens. If its diameter is less than $\lambda_J$, the additional pressure more than
offssets the increased gravity: the cloud re-expands. In a larger cloud gravity wins, and collapse ensues.

Early on, while the Universe is radiation-dominated, the density \( \rho_r = a_B T^4/c^2 \), is low and the pressure is high, with \( c_s = c/\sqrt{3} \). So Equation 8.71 gives

\[
\lambda_J = c^2 \left( \frac{\pi}{3G a_B T^4} \right)^{1/2} \propto T^{-2}.
\]  

(8.72)

The Jeans mass \( M_J \) is the amount of matter in a sphere of diameter \( \lambda_J \):

\[
M_J \equiv \frac{\pi}{6} \lambda_J^3 \rho_m,
\]

(8.73)

where \( \rho_m \) refers only to the matter density. In the radiation-dominated period we have \( M_J \propto \rho_m T^{-6} \), with \( T \propto 1/R(t) \) and \( \rho_m \) decreasing as \( R^{-3} \). So the Jeans mass grows as \( M_J \propto R^3(t) \); the mass enclosed in a sphere of diameter \( \lambda_J \) increases as the Universe becomes more diffuse. At the time \( t_{\text{eq}} \) when the density of matter is equal to that of radiation, the temperature is \( T_{\text{eq}} \) and \( \rho_m \approx \rho_r = a_B T_{\text{eq}}^4/c^2 \).

Radiation still provides most of the pressure, so \( p \approx c^2 \rho_r/3 \) and

\[
M_J(t_{\text{eq}}) = \frac{\pi}{6} \rho_m(t_{\text{eq}}) \left( \frac{\pi c^4/3}{G a_B T_{\text{eq}}^4} \right)^{3/2} = \frac{\pi^{5/2}}{18\sqrt{3}} \frac{c^4}{G^{3/2} a_B^{1/2} T_{\text{eq}}^2}.
\]

(8.74)

If equality occurs at the redshift \( 1 + z_{\text{eq}} = 24000 \Omega_m h^2 \) of Problem 8.10, then

\[
M_J(T_{\text{eq}}) = 3.6 \times 10^{16} (\Omega_m h^2)^{-2} M_\odot.
\]

(8.75)

This is 100 times more than the Virgo cluster, or roughly the mass that we would find today in a huge cube \( 50/(\Omega_m h^2) \) Mpc on a side. This is approximately the spatial scale of some of the largest voids and complexes of galaxy clusters in Figure 8.3. Overdense regions with masses below \( M_J \) could not collapse because the outward pressure of radiation was too strong. Instead, radiation gradually diffused out of them, taking the ionized gas with it, and damping out small irregularities.

After this time, matter provides most of the mass and energy, but the pressure comes mainly from the radiation: so \( \rho \approx \rho_m \), but \( p \approx c^2 \rho_r/3 \). If a small box of the combined matter–radiation fluid is squeezed adiabatically, then, just as in the cosmic expansion, the change \( \Delta \rho_m \) in the matter density is related to \( \Delta \rho_r \) by \( 4 \Delta \rho_m/\rho_m = 3 \Delta \rho_r/\rho_r \). So the sound speed

\[
c_s^2 = \frac{\Delta p}{\Delta \rho} = \frac{c^2 \Delta \rho_r/3}{\Delta \rho_m} = \frac{c^2}{3} \frac{4 \rho_r}{\rho_m} \propto \frac{1}{R(t)}, \quad \text{so} \quad \lambda_J = c_s \frac{\pi}{G \rho_m} \propto R(t)
\]

(8.76)

and the Jeans mass of Equation 8.73 stays constant.
8.5 Growth of structure: clusters, walls, and voids

By a redshift $z_{\text{rec}} \sim 1100$ when the temperature $T_{\text{rec}} \approx 3000 \text{K}$, hydrogen atoms had recombined. Radiation streamed freely through the neutral matter, and no longer contributed to the pressure. The sound speed dropped to that of the matter:

$$c_s(t_{\text{rec}}) \approx \sqrt{\frac{k_B T}{m_p}} \approx 5 \text{ km s}^{-1}.$$  \hspace{1cm} (8.77)

Just afterward, the Jeans mass is

$$M_J = \frac{\pi}{6} \rho_m \left( \frac{\pi k_B T_{\text{rec}}}{G \rho_m m_p} \right)^{3/2} \approx 5 \times 10^4 (\Omega_m h^2)^{-1/2} M_\odot;$$  \hspace{1cm} (8.78)

it has fallen abruptly by a factor of $\sim 10^{12}$.

Radiation continues to transfer some heat to the matter, keeping their temperatures roughly equal until $z \sim 100$. Now the Jeans mass $M_J \propto T^{3/2} \rho_m^{-1/2}$, and, because the radiation cools as $T_r \propto R^{-1}$, that decrease offsets the drop in density $\rho_m$ to keep $M_J$ nearly constant. If the first dense objects formed with roughly this mass, similar to that of a globular cluster, they could subsequently have merged to build up larger bodies. Once it is no longer receiving heat, the matter cools according to $T_m \propto R^{-2}$. To see why, think of the perfect gas law relating temperature to volume, or recall that expansion reduces the random speeds of atoms according to Equation 8.54. So the Jeans mass falls further; after recombination, gas pressure is far too feeble to affect the collapse of anything as big as a galaxy.

But how can we make objects the size of galaxies or galaxy clusters, that are too small to grow before recombination? Equation 8.57 tells us that the fraction $\delta$ by which their density exceeds the average grows with time as $t^{2/3}$ or $R(t)$. To reach $\delta \gtrsim 1$ before the present, we would need $\delta(t_{\text{rec}}) \gtrsim 10^{-3}$ at $z_{\text{rec}} = 1100$. But, aside from the highest peak in Figure 8.14, $\Delta T < 50 \text{ mK}$ or $2 \times 10^{-5}$ times the average temperature. This is far too small; so why do we see any galaxies and galaxy clusters today?

8.5.2 WIMPs to the rescue!

The dark matter far outweighs the neutrons and protons. Although we have yet to detect the particles themselves, dark matter is most probably composed of weakly interacting massive particles (WIMPs). Like neutrinos, WIMPs lack strong and electromagnetic interactions – or they would not be ‘dark’ – and they have some small but nonzero mass. WIMPs can collapse into galaxy-sized lumps early on, because, unlike the baryons, they are unaffected by radiation pressure.
To describe this collapse, we can follow the same calculation as for the Jeans length and Jeans mass, but for WIMPs with density $\rho_w$ and typical random speeds $c_w$. Instead of Equation 8.73, we find that a dense region has too little kinetic energy and falls in on itself if it contains a mass larger than

$$M_{J, \text{wimp}} = \frac{\pi}{6} \rho_w \left( \frac{\pi c_w^2}{G \rho_w} \right)^{3/2}.$$  

(8.79)

While the WIMPs are relativistic, their Jeans mass is high and grows with time just as in the radiation-dominated case. A slightly overdense region that is not actively collapsing simply disperses, as WIMPs stream out of it at light speed. All structure smaller than the horizon scale of Problem 8.14 is erased in this way.

But as soon as the speed $c_w$ of random motions drops appreciably below $c/\sqrt{3}$, the Jeans mass starts to fall. Very roughly, all dense clumps of WIMPs that are larger than the horizon scale at the time when they cease moving relativistically now begin to collapse. Since inflation left behind fluctuations with a power spectrum $P(k)$ rising with $k$, lumps just larger than this will have the highest densities. The more massive the WIMPs, the smaller the horizon scale when this happens, and the smaller and denser the structures that form.

Neutrinos, with masses of a few electron-volts, remain relativistic until almost $t_{\text{eq}}$, when the comoving size of the horizon is $\sim 16(h^2\Omega_m)^{-1}$ Mpc. Such light particles are called hot dark matter. If the dark matter is hot, we still have difficulty in understanding how something as small as a galaxy or even a galaxy cluster can form. WIMPs massive enough that their sound speed $c_w$ fell below the speed of light long before the time $t_{\text{eq}}$ of matter–radiation equality are called cold dark matter. The most popular WIMP candidates have masses $\gtrsim 1$ GeV; their random motions drop well below light speed when $T < 10^{13}$ K, only $10^{-6}$ s after the Big Bang, when the mass $M_H(WIMP)$ within the horizon was less than $M_\odot$.

As it escaped from the contracting clouds of WIMPs, the radiation took the normal matter with it. So both of these should be quite evenly spread at recombination, and the temperature of the cosmic background radiation should be nearly the same across the whole sky. At recombination, as the matter became neutral and was freed of the radiation pressure, it fell into the already-dense clumps of WIMPs. Fluctuations in the density of normal matter could then grow far more rapidly than Equation 8.57 allows, building up the galaxies and clusters.

If the dark matter is cold, galaxies themselves would be built from successive merger of these smaller fragments. We call this the bottom-up picture because galaxies form early, and then fall together to form clusters and larger structures. Figure 8.16 shows results from a gravitational $N$-body simulation following the way that gravity amplifies small initial ripples in an expanding Universe of cold...
8.5 Growth of structure: clusters, walls, and voids

Fig. 8.16. A slice $20 h^{-1}$ Mpc thick, through a gravitational $N$-body simulation with cold dark matter, viewed at the present day. Side frames show magnified views of dense clumps; galaxy groups would form in these ‘dark halos’ – D. Weinberg.

dark matter. The figure shows a stage of the calculation representing the present day. Notice the profusion of small dense clumps linked by the filamentary cosmic web, and that smaller structures look like denser, scaled-down copies of larger ones. The densest regions, shown in the side boxes, have ceased to expand and have fallen back on themselves. Gas would accumulate there, cooling to form clusters of luminous galaxies.

Figure 8.17 combines the information from WMAP in Figure 8.14 with that from the 2dF galaxy survey in Figures 8.3–8.5 to estimate the power spectrum $P(k)$ for matter today. Using a model close to the benchmark cosmology, Dr. Sánchez deduced from the irregularities in the cosmic microwave background what the distribution of WIMPS and baryons must have been at the time $t_{\text{rec}}$. He then calculated how the concentrations of WIMPS became denser according to Equation 8.57, while baryons fell into them. The results agree with $P(k)$ measured from the galaxies of 2dF in the region where they overlap. On these scales, luminous galaxies are distributed in the same way as the dark matter, and both are well described by the model curve. Does this mean that we have now solved all the problems of cosmology? One might hope for a physical understanding of the dark energy, which is now simply inserted as a term in the Friedmann equations. But for the large structures that we have discussed in this chapter, the benchmark cosmology and benchmark initial fluctuations give an excellent account of what we can observe.
The large-scale distribution of galaxies

Fig. 8.17. Data from WMAP (triangles) and the 2dF galaxy survey (dots) are combined to trace the power spectrum $P(k)$. The smooth curve shows the prediction from a flat ($k = 0$) model similar to the benchmark cosmology. The wiggle at $k \approx 0.1$ is an acoustic peak on a scale of $\sim 10$ Mpc, too small to be measured by WMAP – A. Sánchez: model b5 from MNRAS 366, 189 (2006).

8.5.3 How early can galaxies and clusters form?

To find out how long a galaxy or cluster takes to reach its present density, we can use the ‘top-hat’ model, thinking of the overdense protocluster as a uniform sphere. In a homogeneous Universe, the matter beyond that sphere does not exert any forces within it. So we are free to make our sphere more or less dense than its surroundings, and the Friedmann equations still hold. In the following problem, we use Equation 8.26 to examine the collapse of a denser-than-usual region that is destined to become a galaxy or cluster.

Problem 8.30 Suppose that the time $t_0$ refers to a moment when the Universe is matter-dominated, and $\Omega_m > 1$ in our spherical protocluster. By substituting into Equation 8.26, show that the parametric equations

$$\frac{\mathcal{R}(t)}{\mathcal{R}(t_0)} = \frac{\Omega_m}{2(\Omega_m - 1)}(1 - \cos \eta),$$

$$H_0 t = \frac{\Omega_m}{2(\Omega_m - 1)^{3/2}}(\eta - \sin \eta).$$

(8.80)

describe a solution. (This is the same as Equation 4.24 of Section 4.5, for $\epsilon = 0$ – why?) Show that $\mathcal{R}(t)$ is largest when $\eta = \pi$, at the turn-around time $t_a = \pi \Omega_m/[2H_0(\Omega_m - 1)^{3/2}]$, and that the sphere collapses to high density at time $2t_a$. 
At time $t_0$, suppose that this denser region is expanding at the same rate as its surroundings, and that $t_0$ is early enough that we can apply Equation 8.27: $R(t) \propto t^{2/3}$ so that $\rho(t) \propto 1/t^2$, and $t_0 H_0 = 2/3$. Using the result of Problem 8.16, show that, between $t_0$ and $t_a$, the density $\rho_{out}$ outside the sphere drops such that 

$$\rho_{out}(t_a) = \rho_{out}(t_0) \left[ \frac{9\pi^2}{16} \left( \frac{\Omega_m^2}{1 - \Omega_m^2} \right)^3 \right] \quad \text{while inside} \quad \frac{\rho_{in}(t_a)}{\rho_{in}(t_0)} = \left( \frac{\Omega_m - 1 - \Omega_m}{\Omega_m} \right)^3. \quad (8.81)$$

So $\rho_{in}(t_a)/\rho_{out}(t_a) = (3\pi/4)^2$. As it turns around and begins to collapse, this sphere is roughly 5.6 times denser than its surroundings. 

Just as the free-fall time of Equation 3.23 is the same for all particles in a sphere of uniform density, the collapse time $2t_a$ is the same throughout this sphere. So, in our simple model, all the particles reach the center at the same moment. In the real cosmos, they would have small random motions which prevent this. The dark matter and any stars present will undergo violent relaxation (see Section 6.2) as they settle into virial equilibrium. Gas can lose energy by radiating heat away as it is compressed. Once our protocluster settles into equilibrium, the virial theorem tells us that its energy $E_1 = PE_1 + KE_1 = -PE_1/2$. $E_1$ can be no greater than the total energy $E_0 = P\lambda E_0$ when it was at rest at time $t_a$, poised between expansion and contraction: so we must have $PE_1 < 2P\lambda E_0$.

**Problem 8.31** Use Equation 3.33 for the potential energy $P\lambda E$ of a galaxy of stars to show that, if the distances between stars all shrink by a factor $f$, so the density increases as $1/f^3$, then $P\lambda E$ increases as $1/f$. 

If we make the too-simple approximation that the collapse is homologous, so that all distances between particles shrink by an equal factor, Problem 8.31 tells us that our protocluster’s final radius is no more than half as large as it was at turn-around, and the density is at least eight times greater. Meanwhile, the cosmos continues to expand, and its average density has dropped at least four times since $t_a$ (why?). So at the time that it reaches virial equilibrium, our cluster is $4 \times 8 \times 5.6 \approx 180$ times denser than the critical density for the Universe around it: recall Problem 8.2. In a galaxy cluster, we define the radius $r_{200}$ such that, within it, the average density is 200 times the critical density; $r_{200}$ is sometimes called the virial radius. At larger radii, the cluster cannot yet be relaxed and in virial equilibrium. Even the relaxed core will be disturbed when new galaxies fall through it as they join the cluster.

We can use the ‘top-hat’ model to estimate when the galaxies and clusters could have formed. Within the Sun’s orbit, the Milky Way’s density averages to $10^5\rho_{crit}$. So when it collapsed at time $2t_a$, the average cosmic density was no more than 500 times the present critical density; $\Omega_m = 0.3$, so this is 1700 times
the present average density. The average density varies as \((1 + z)^3\), so the collapse was at \(1 + z \leq (1700)^{1/3} \approx 12\). It could have been later, since the gas can radiate away energy, and become denser than the virial theorem predicts. But it could not have taken place any earlier.

**Problem 8.32** In Problem 7.7 we found that, in the core of the Virgo cluster, luminous galaxies are packed 2500 times more densely than the cosmic average. Assuming that dark and luminous matter are well mixed in the cluster, show that its core could not have assembled before redshift \(z = 1.3\). How early could the central region of NGC 1399, from Problem 6.4, come together?

### 8.5.4 Using galaxies to test model cosmologies

How well does the benchmark cosmology with cold dark matter account for real galaxies? Its huge success is to explain why the cosmic microwave background is so smooth, while the distribution of galaxies is so lumpy. We can even explain the shape of the power spectrum \(P(k)\) in Figure 8.17, which describes the non-uniformity. That power spectrum requires that the smallest lumps of matter are now densest, as we saw in Figure 8.6. Problem 8.30 shows that they are the first to stop expanding and collapse on themselves. So we might expect that all galaxies will contain some very dense regions, which should have made stars early on. Even the smallest galaxies should have some very old stars – as we saw in Section 4.4 for the dwarf galaxies of the Local Group.

Structures that collapsed most recently should be larger and less dense than those that formed earlier. Using Equation 8.21 for the critical density, we can find the mass \(\mathcal{M}_{200}\) measured within the virial radius \(r_{200}\):

\[
\mathcal{M}_{200} = \frac{4}{3} \pi r_{200}^3 \times 200 \rho_{\text{crit}} = \frac{100 r_{200}^3 H^2(t)}{G},
\]

while the speed of a circular orbit at that radius is

\[
V_c^2(r_{200}) = \frac{G \mathcal{M}_{200}}{r_{200}}, \quad \text{so} \quad \mathcal{M}_{200}(t) = \frac{V_c^3(r_{200})}{10GH(t)},
\]

So, if we measure rotational or random speeds near the radius \(r_{200}\) (or if they do not change very much with radius), the mass or luminosity should increase steeply with those measured speeds.

We see this pattern in the Tully–Fisher relation for disk galaxies (Figure 5.23), the fundamental plane for elliptical galaxies (Figure 6.13), and the relation between temperature and X-ray luminosity for gas in galaxy clusters (Figure 7.12). In the past the Hubble parameter \(H(t)\) was larger, so we expect that temperatures
and speeds were higher for a given mass or luminosity. Figure 6.13 shows this effect, but in Figure 7.12 galaxy clusters at $z \sim 1$ follow the same relation as local objects. Adding gas to simulations like that of Figure 8.16 does result in model galaxies that follow Tully and Fisher’s dependence of mass on rotation speed. But the ‘galaxies’ fail to gather enough gas from large distances, so the disk has too little angular momentum and its radius is too small.

The slope of $P(k)$ in Figure 8.17 means that small objects will be far more numerous than large ones. The smallest have roughly the solar mass, since this is the mass $M_H(WIMP)$ within the horizon when the random motions of the WIMPs drop below near-light speeds. The halo of a galaxy like the Milky Way is made by merging many thousands of smaller objects, most of which are torn apart. Those that fall in relatively late survive today as distinct objects: satellite dark halos. In models such as that in Figure 8.16 a Milky-Way-sized dark halo will have $\sim 300$ dark satellites massive enough to have $V_c > 10$ km s$^{-1}$. But the real Milky Way only has ten or so luminous satellites. Choosing ‘warm’ dark matter, for which the random motions remained relativistic until the mass within the horizon $M_H(WIMP) \sim 10^9 M_\odot$, would erase all but the largest satellites. Some theories of particle physics include a ‘sterile neutrino’ with mass $\sim 1$ keV, which would have this property. Other possibilities are that their first few stars blew all the remaining gas out of most of these dark halos, or that fierce ultraviolet and X-ray radiation from the first galaxies heated it so far that it could not cool to make new stars. The Milky Way would then have 10 luminous satellites and 290 dark ones.

Another facet of the same difficulty is that we see large objects ‘too early’ in cosmic history. In Section 9.4 we will find that some massive galaxies have formed more than $10^{11} M_\odot$ of stars, corresponding to more than $10^{12} M_\odot$ of dark matter, less than 2 Gyr after the Big Bang at $z \sim 3$. The benchmark model with cold dark matter predicts that such early ‘monster’ galaxies should be extremely rare; it is not clear whether it already conflicts with the observations.

If the dark matter is cold, then all galaxies should have very dense cores. Figure 8.6 shows that $\sigma_R$, the variation in density on lengthscale $R$, rises at small $R$ or large $k$. So the first regions to collapse will be the smallest and also the densest. Equation 8.83 shows that the velocities of particles within them will also be low, because $H(t)$ is large. So the positions and velocities of these first objects are tightly grouped: the density in phase space is high. Simulations like Figure 8.16 show that, as the galaxy is built, such objects fall together into a very dense center: the density of WIMPs follows $\rho_w \propto r^{-\alpha}$, where $1 \lesssim \alpha \lesssim 1.5$. The Navarro–Frenk–White model of Equation 3.24 was developed to describe this central cusp. The prevalence of dwarfs in galaxy clusters (see Figure 7.8) shows that they must indeed be robust, and hence dense, to avoid being torn apart by tidal forces. But the stars we observe are never as concentrated as this model requires the WIMPS to be, and the rotation curves of spiral galaxies in Figure 5.21 also seem to rise more gently than this form of $\rho_w$ would allow.
The large-scale distribution of galaxies

However, we know that galaxy centers contain mostly normal baryonic matter, and that gas physics is complex – we do not even know how to predict the masses of stars formed locally in our own Milky Way. So it is no surprise that we cannot yet use basic physics to calculate exactly how the galaxies should form. In the next chapter we turn to observations of the distant Universe, and to what we can learn by viewing galaxies and protogalactic gas as they were 8–10 Gyr ago as the Milky Way began to form its disk, and even earlier when our oldest stars were born.