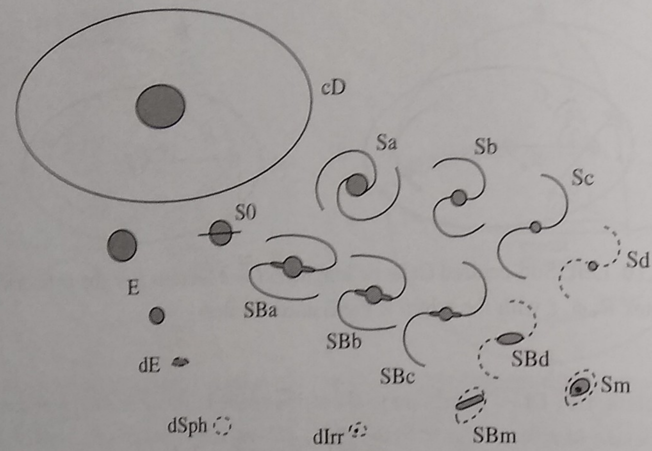

Figure 1.11 Galaxy classification: a modified form of Hubble's scheme.

1.3 Other galaxies

This section introduces the study of galaxies other than our own Milky Way. We discuss how to classify galaxies according to their appearance in optical light, and how to measure the amount of light that they give out. Although big galaxies emit most of the light, the most common type of galaxy is a tiny dim dwarf.

The existence of other galaxies was established only in the 1920s. Before that, they were listed in catalogues of *nebulae*: objects that appeared fuzzy in a telescope and were therefore not stars. Better images revealed stars within some of these 'celestial clouds'. Using the newly opened 100″ telescope on Mount Wilson, Edwin Hubble was able to find variable stars in the Andromeda 'nebula' M31. He showed that their light followed the same pattern of changing brightness as Cepheid variable stars within our Galaxy. Assuming that all these stars were of the same type, with the same luminosities, he could find the relative distances from Equation 1.1. He concluded that the stars of Andromeda were at least 300 kpc from the Milky Way, so the nebula must be a galaxy in its own right. We now know that the Andromeda galaxy is about 800 kpc away.

Hubble set out his scheme for classifying the galaxies in a 1936 book, *The Realm of the Nebulae*. With later additions and modifications, this system is still used today; see Figure 1.11. Hubble recognized three main types of galaxy: ellipticals, lenticulars, and spirals, with a fourth class, the irregulars, for galaxies that would not fit into any of the other categories.

Elliptical galaxies are usually smooth, round, and almost featureless, devoid of such photogenic structures as spiral arms and conspicuous dust lanes. Ellipticals are generally lacking in cool gas and consequently have few young blue stars. Though they all appear approximately elliptical on the sky, detailed study shows

Figure 1.12 Elliptical galaxies. Left, giant elliptical NGC 5128 (Centaurus A), a powerful radio source. The unusual dark lane marks a disk of dusty gas that has probably fallen into the galaxy from outside – NOAO. Right, nearby dwarf elliptical NGC 147 in the *V* band. Individual stars are seen in the outer parts; the brightest stars are in the foreground, and belong to the Milky Way – WIYN telescope.

that large bright ellipticals have rather different structures from their smaller and fainter counterparts.

Ellipticals predominate in rich clusters of galaxies, and the largest of them, the *cD galaxies*, are found in the densest parts of those clusters. Around an elliptical core, the enormous diffuse envelope of a cD galaxy may stretch for hundreds of kiloparsecs; these systems can be up to 100 times more luminous than the Milky Way. Normal or *giant* ellipticals have luminosities a few times that of the Milky Way, with characteristic sizes of tens of kiloparsecs. The stars of these bright ellipticals show little organized motion, such as rotation; their orbits about the galaxy center are oriented in random directions. The left panel of Figure 1.12 shows a giant elliptical, which is also a radio galaxy; see Section 8.1.

In less luminous elliptical galaxies, the stars have more rotation and less random motion. Often there are signs of a disk embedded within the elliptical body. The very faintest ellipticals, with less than ~1/10 of the Milky Way's luminosity, split into two groups. The first comprises the rare compact ellipticals, like the nearby system M32. The other group consists of the faint diffuse *dwarf elliptical* (dE) galaxies, and their even less luminous cousins the *dwarf spheroidal* (dSph) galaxies, which are so diffuse as to be scarcely visible on sky photographs. The right panel of Figure 1.12 shows a dwarf elliptical satellite of M31. The dE and dSph galaxies show almost no ordered rotation.

Lenticular galaxies show a rotating disk in addition to the central elliptical bulge, but the disk lacks any spiral arms or extensive dust lanes. These galaxies

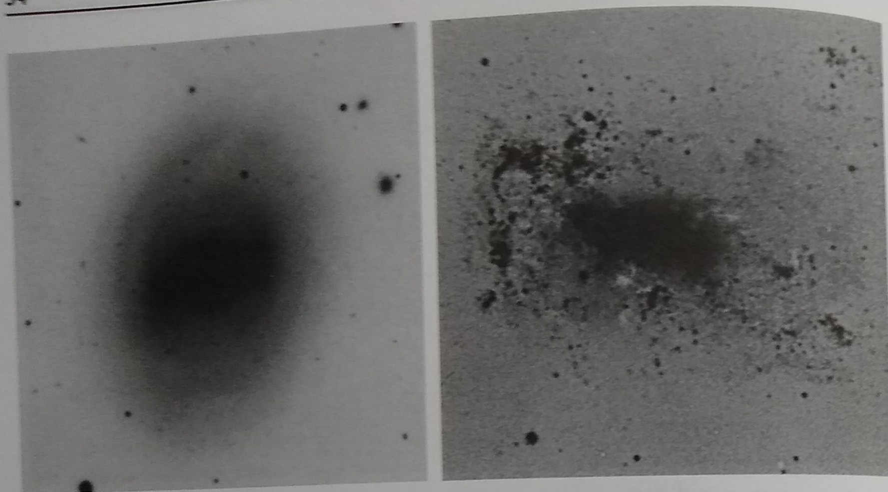

Figure 1.13 Disk galaxies: luminous regions appear darkest in these negative images. Left, NGC 936, a luminous barred S0 with $L \approx 2 \times 10^{10} L_\odot$; the smooth disk has neither dust lanes nor spiral arms. Right, NGC 4449, classified as irregular or SBm; this is a small gas-rich galaxy with $L \approx 4 \times 10^{9} L_\odot$. Bright star-forming knots are strewn about the disk – CFHT.

are labelled S0 (pronounced 'ess-zero'), and they form a transition class between ellipticals and spirals. They resemble ellipticals in lacking extensive gas and dust, and in preferring regions of space that are fairly densely populated with galaxies; but they share with spirals the thin and fast-rotating stellar disk. The left panel of Figure 1.13 shows an SB0 galaxy, with a central linear bar.

Spiral galaxies are named for their bright spiral arms, especially conspicuous in the blue light that was most easily recorded by early photographic plates (Figure 1.14). The arms are outlined by clumps of bright hot O and B stars, and the compressed dusty gas out of which these stars form. About half of all spiral and lenticular galaxies show a central linear bar: the barred systems SB0, SBa, . . . , SBd form a sequence parallel to that of the unbarred galaxies. Along the sequence from Sa spirals to Sc and Sd, the central bulge becomes less important relative to the rapidly rotating disk, while the spiral arms become more open and the fraction of gas and young stars in the disk increases. Our Milky Way is probably an Sc galaxy, or perhaps an intermediate Sbc type; M31 is an Sb. On average, Sc and Sd galaxies are less luminous than the Sa and Sb systems, but some Sc galaxies are still brighter than a typical Sa spiral.

At the end of the spiral sequence, in the Sd galaxies, the spiral arms become more ragged and less well ordered. The Sm and SBm classes are *Magellanic spirals*, named for their prototype, which is our Large Magellanic Cloud; see Section 4.1. In these, the spiral is often reduced to a single stubby arm. As the galaxy luminosity decreases, so does the speed at which the disk rotates; dimmer

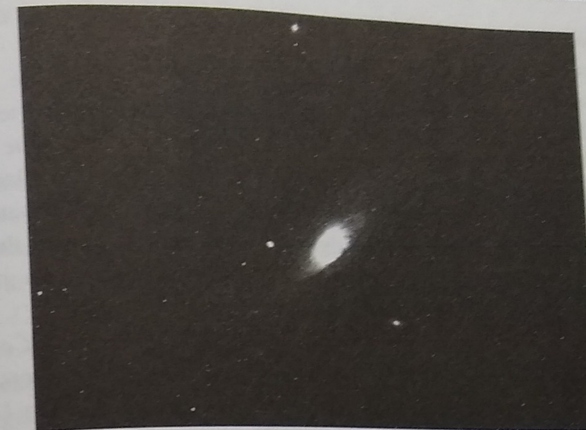

Figure 1.14 Our nearest large neighbor, the Andromeda galaxy M31; north is to the right, east is upward. Note the large central bulge of this Sb galaxy, and dusty spiral arms in the disk. Two satellites are visible: M32 is round and closer to the center, NGC 205 is the elongated object to the west – W. Schoening, Burrell Schmidt telescope, NOAO.

galaxies are less massive. The Large Magellanic Cloud rotates at only 80 km s^{-1}, a third as fast as the Milky Way. Random stellar motions are also diminished in the smaller galaxies, but even so, ordered rotational motion forms a less important part of their total energy. We indicate this in Figure 1.11 by placing these galaxies to the left of the Sd systems.

The terms 'early type' and 'late type' are often used to describe the position of galaxies along the sequence from elliptical galaxies through S0s to Sa, Sb, and Sc spirals. Some astronomers once believed that this progression might describe the life cycle of galaxies, with ellipticals turning into S0s and then spirals. Although this hypothesis has now been discarded, the terms live on. Confusingly, 'early type' galaxies are full of 'late type' stars, and vice versa.

Hubble placed all galaxies that did not fit into his other categories in the *irregular* class. Today, we use that name only for small blue galaxies which lack any organized spiral or other structure (Figure 1.13). The smallest of the irregular galaxies are called *dwarf irregulars*; they differ from the dwarf spheroidals by having gas and young blue stars. It is possible that dwarf spheroidal galaxies are just small dwarf irregulars which have lost or used up all of their gas.

Other galaxies that Hubble would have called irregulars include the *starburst galaxies*; we discuss these in Section 5.5. These systems have formed many stars in the recent past, and their disturbed appearance results in part from gas thrown out by supernova explosions. Interacting galaxies, in which two or more systems have come close to each other, and galaxies that appear to result from the merger of two or more smaller systems, would also have fallen into this class. We have come to realize that galaxies are not 'island universes,' but affect each other's development throughout their lives. Chapters 4, 5, and 6 of this book deal with the structure of nearby galaxies.

We usually refer to galaxies by their numbering in a catalogue. Charles Messier's 1784 catalogue lists 109 objects that look 'fuzzy' in a small telescope; it includes the Andromeda galaxy as M31. The *New General Catalogue* of more than 7000 nonstellar objects includes clusters of stars and gaseous nebulae as well as galaxies. Published by J.L.E. Dreyer in 1888, with additions in 1895 and 1908, it was based largely on the work of William Herschel (who discovered the planet Uranus), his sister Caroline, and son John Herschel. The Andromeda galaxy is included as NGC 224.

Modern catalogues of bright galaxies include the *Third Reference Catalogue of Bright Galaxies*, by G. and A. de Vaucouleurs and their collaborators (1991; Springer, New York), which includes all the NGC galaxies, and the *Uppsala General Catalogue of Galaxies*, by P. Nilson (1973; Uppsala Observatory), with its southern extension, the *ESO/Uppsala Survey of the ESO(B) Atlas*, by A. Lauberts (1982; European Southern Observatory). Galaxies that emit brightly in the radio, X-rays, etc., also appear in catalogues of those sources. Many recent catalogues, such as the current NASA Extragalactic Database (on the World-Wide Web at ned.ipac.caltech.edu), are published electronically.

Further reading: E. Hubble, 1936, *The Realm of the Nebulae* (Yale University Press; reprinted by Dover, New York); for pictures to illustrate Hubble's classification, see A. Sandage, 1961, *The Hubble Atlas of Galaxies* (Carnegie Institute of Washington; Washington, DC). For a modern treatment of galaxy classification on an undergraduate level, see D.M. Elmegreen, 1998, *Galaxies and Galactic Structure* (Prentice-Hall, New Jersey); a graduate text is S. van den Bergh, 1998, *Galaxy Morphology and Classification* (Cambridge University Press, Cambridge, UK).

1.3.1 Galaxy photometry

Unlike stars, galaxies do not appear as points of light; they are extended objects on the sky. Turbulence in the Earth's atmosphere has the effect of blurring galaxy images; this is known as *seeing*. Because of it, a ground-based optical telescope rarely shows details smaller than about $1/3''$. For sharper images, we must use a telescope in space or resort to techniques such as interferometry.

Although the classification of galaxies is still based on their appearance in optical images, most work on galaxies is quantitative, measuring how much light, at what wavelengths, is emitted by the different regions. The *surface brightness* of a galaxy $I(\mathbf{x})$ is the amount of light per square arcsecond on the sky at a particular point $\mathbf{x}$ in the image. Consider a small square patch of side D in a galaxy that we view from a distance d, so that it subtends an angle $\alpha = D/d$ on the sky. If the combined luminosity of all the stars in this region is L, its apparent brightness F

is given by Equation 1.1; then the surface brightness is

$$I(\mathbf{x}) \equiv \frac{F}{\alpha^2} = \frac{L/4\pi d^2}{D^2/d^2} = \frac{L}{4\pi D^2}. \tag{1.17}$$

The units for I are mag arcsec^{-2}: the apparent magnitude of a star that appears as bright as one square arcsecond of the galaxy's image. The surface brightness at any point does not depend on distance unless d is so large that the expansion of the Universe has the effect of reducing $I(\mathbf{x})$; we discuss this further in Section 8.3. Contours of constant surface brightness on a galaxy image are called *isophotes*. Equation 1.17 shows that the position of an isophote within the galaxy is independent of the observer's distance.

We generally measure surface brightness in a fixed wavelength band, just as for stellar photometry. The centers of galaxies reach only $I_B \approx 18$ mag arcsec^{-2} or $I_R \approx 16$ mag arcsec^{-2}, and the stellar disks are much fainter. Galaxies do not have sharp edges, so we often measure their sizes within a fixed isophote. One popular choice is the 25th-magnitude isophote in the blue B band, denoted R_{25}. This is about 1% of the sky level on an average night; before CCD photometry (see Section 5.1), it was close to the limit of what could be measured reliably. Another option is the *Holmberg radius* at $I_B(\mathbf{x}) = 26.5$ mag arcsec^{-2}. To find the luminosity of the whole galaxy, we measure how the amount of light coming from within a given radius grows as that radius is moved outward, and we extrapolate to reach the total.

Problem 1.10: In a galaxy at a distance of d Mpc, what would be the apparent B magnitude of a star like our Sun? In this galaxy, show that $1''$ on the sky corresponds to $5d$ pc, and hence that the surface brightness $I_B = 27$ mag arcsec^{-2} is equivalent to $1L_\odot$ pc^{-2}.

Table 1.6 gives the surface brightness of the *night sky* measured in the band-passes of Figure 1.7. These are approximate average values, since the sky brightness depends on the solar activity (sunspot cycle), the observatory's location on Earth, and the direction in the sky. Typically, the sky is brighter than all but the inner core of a galaxy, and on a moonlit night even the center can disappear into the bright sky. From Earth's surface, optical observations of galaxies must generally be made during the dark of the moon.

If our eyes could perceive colors at such low light levels, we would see the sky glowing red with emission in the bands of atmospheric molecules. In the near-infrared at 2 μm, the sky is over a thousand times brighter than it would be in space. Figure 1.15 shows how steeply its emission rises at longer wavelengths, in the *thermal infrared*. Standard infrared filters are chosen to lie where the atmosphere is most transparent. Between these regions, we see a blackbody spectrum corresponding to the temperature of the opaque layers.

Table 1.6 Sky brightness in UV, optical and infrared

Bandpass	*Wavelength*	*From space*	*Dark sky*	*Full moon*
	1500 Å	25.0	—	—
	2000 Å	26.0	—	—
	2500 Å	25.6	—	—
U	3700 Å	23.2	22.0	—
B	4400 Å	23.4	22.7	19.4
V	5500 Å	22.7	21.8	19.7
R	6400 Å	22.2	20.9	19.9
I	8000 Å	22.2	19.9	19.2
J	1.2 μm	20.7	15.0	15.0
H	1.6 μm	20.9	13.7	13.7
K	2.2 μm	21.3	12.5	12.5
K′	2.2 μm	21.3	13.7	13.7

Note: Sky brightness units are magnitudes per square arcsecond.

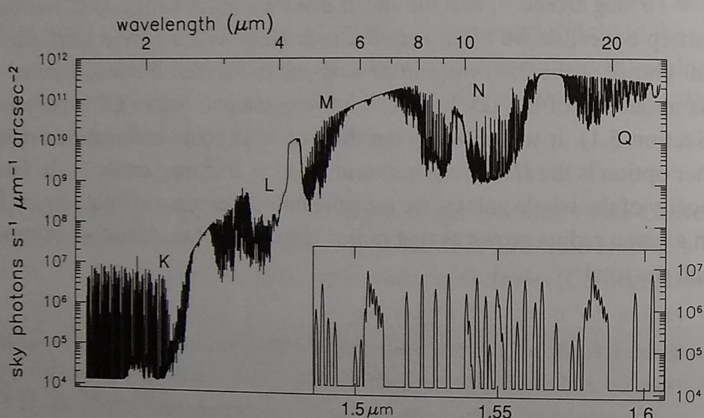

Figure 1.15 Sky emission on Mauna Kea, Hawaii, at 4000 m elevation; standard infrared bandpasses are indicated. Inset shows that the sky background consists mainly of closely spaced emission lines – Gemini telescope project.

We can cut down the sky light by designing our filters to exclude some of the strongest lines; using the K' filter instead of K blocks out about two-thirds of the emission. However, Table 1.6 makes clear that when we observe from the ground, the infrared sky is *always* brighter than the galaxy. To find the surface brightness accurately, we must measure the brightness of a patch of blank sky as it changes throughout the night just as accurately as we measure the galaxy-plus-sky; the small difference between the two gives $I(\mathbf{x})$. Using a telescope in space gives us a much darker sky at red and infrared wavelengths; we can also observe in the near-ultraviolet, where the sky brightness is yet lower.

There are many more small dim galaxies than large bright ones. Figure 1.16 shows the number of galaxies measured at each R-band absolute magnitude M_R,

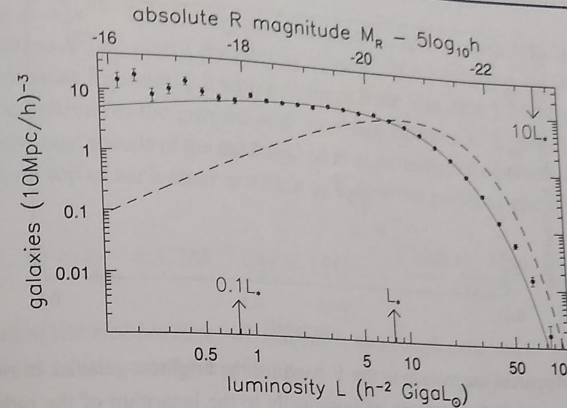

Figure 1.16 Number of galaxies $\Phi(M)$ per 10 Mpc cube between absolute magnitude M_R and $M_R + 1$; vertical bars indicate errors. The solid line shows the luminosity function of Equation 1.18; the dashed line is $\Phi(M) \times L/L_\star$, the luminosity from galaxies in each interval of absolute magnitude – H. Lin.

in a survey from the Las Campanas observatory in Chile. The solid curve shows what is expected if the number of galaxies $\Phi(L)\Delta L$ per Mpc^3 between luminosity L and $L + \Delta L$ is given by

$$\Phi(L)\Delta L = n_\star \left(\frac{L}{L_\star}\right)^\alpha \exp\left(-\frac{L}{L_\star}\right) \frac{\Delta L}{L_\star}; \tag{1.18}$$

this is the *Schechter function*. According to this formula, the number of galaxies brighter than the luminosity $L_\star$ drops very rapidly; we often use the criterion $L \gtrsim 0.1 L_\star$ to define a 'bright' or 'giant' galaxy, as opposed to a dwarf. The solid curve is for $L_\star \approx 8 \times 10^9 h^{-2} L_\odot$, corresponding to $M_{R,\star} = -20.3 + 5\log_{10} h$; as explained in the next section, the parameter h measures the rate at which the Universe expands. Taking $h = 0.75$, we find $L_\star \approx 2 \times 10^{10} L_\odot$, roughly the Milky Way's luminosity.

The number of galaxies in each unit interval in absolute magnitude is almost constant when $L < L_\star$; the curve is drawn for $n_\star = 0.019 h^3$ Mpc^{-3} and for $\alpha = -0.7$. The Schechter formula overestimates the density of very faint galaxies; for $\alpha \leq -1$, it even predicts that the total number of galaxies $\int_L^\infty \Phi(L)\mathrm{d}L$ should increase without limit as $L \to 0$. But the dashed line shows that most of the light comes from galaxies close to $L_\star$. Integrating Equation 1.18, we estimate the total luminosity density to be

$$\int_0^\infty \Phi(L) L \,\mathrm{d}L = n_\star L_\star \Gamma(\alpha + 2) \approx 1.4 \times 10^8 h L_\odot \,\mathrm{Mpc}^{-3}. \tag{1.19}$$

[Here Γ is the gamma function; $\Gamma(j+1) = j!$ when j is an integer.]

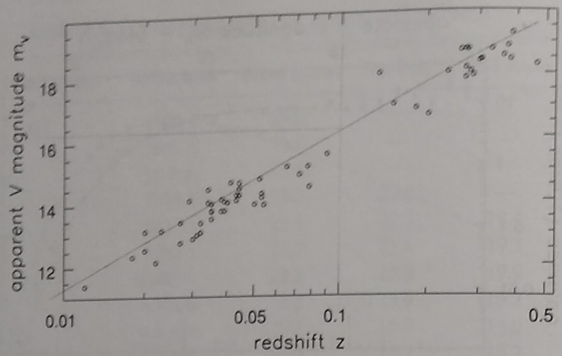

Figure 1.17 Apparent magnitude in the V band for the brightest galaxies in rich galaxy clusters. The magnitude increases proportionally to the logarithm of the redshift z, as we expect if the galaxy's distance is proportional to its recession speed cz – data from J.E. Gunn and J.B. Oke, ApJ 195, 255; 1975.

1.4 Galaxies in the expanding Universe

The Universe is expanding; the galaxies are rushing away from us. The recession speed, as measured by the Doppler shift of a galaxy's spectral lines, is larger for more distant galaxies. We can extrapolate this motion back into the past to estimate when the Universe had its beginning in the *Big Bang*. Doing this, we link the recession speed or *redshift* that we measure for a galaxy with the time after the Big Bang at which its light was given out; the redshift becomes a measure of the galaxy's age when it emitted that light.

In 1929, based on only 22 measurements of radial velocities for nearby galaxies, and some distance estimates which turned out to be wrong by about a factor of ten, Hubble claimed that the galaxies are moving away from us with speeds V_r proportional to their distance d:

$$V_r \approx H_0 d. \tag{1.20}$$

Subsequent work proved him right, and this relation is now known as *Hubble's law*. Current estimates for the parameter H_0, the *Hubble constant*, lie between 40 and $80\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Figure 1.17 shows that galaxies that recede faster are indeed fainter, as expected if they all have roughly the same luminosity, but are progressively more distant.

We often use Hubble's law to estimate the distances of galaxies from their measured velocities. It is common to indicate the uncertainty in the Hubble constant explicitly, by writing h for the value of H_0 in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Then Equation 1.20 implies

$$d = h^{-1}[V_r(\mathrm{km\,s^{-1}})/100]\,\mathrm{Mpc}. \tag{1.21}$$

When the distance of a galaxy is found from its radial velocity V_r, the derived

luminosity $L \propto h^{-2}$. This is why the parameter $L_\star$ of Equation 1.18 has a value proportional to h^2; similarly, the density $n_\star \propto h^3$. If we estimate the mass $\mathcal{M}$ of a galaxy by using Equation 1.2 with a distance from Equation 1.21, together with Newton's equation for the gravitational force (see Section 3.1), then $\mathcal{M} \propto h^{-1}$.

If the average speeds of the galaxies had always remained constant, they would have been on top of each other at a time t_H before the present, where

$$t_H = \frac{1}{H_0} = 9.78h^{-1}\,\mathrm{Gyr} = 15\,\mathrm{Gyr} \times \frac{67\,\mathrm{km\,s^{-1}\,Mpc^{-1}}}{H_0}. \tag{1.22}$$

This is called the *Hubble time*; we can use it as a rough estimate of the age of the Universe, the time since the Big Bang.

Problem 1.11: If a galaxy has absolute magnitude M, use Equations 1.1 and 1.21 to show that its apparent magnitude m is related to the redshift $z = V_r/c$ of Equation 1.16 by $m = M + 5\log_{10} z + C$, where C is a constant, the same for all objects. Draw an approximate straight line through the points in Figure 1.17; check that its slope is roughly what you would expect if the brightest galaxy in a rich cluster always had the same luminosity.

Using Hubble's law to find approximate distances for galaxies, we can examine their distribution in space. Figure 1.18 shows nearby galaxies in about 700 square degrees of the southern sky. They are not spread uniformly through space, but concentrated into groups and clusters. Rich clusters of galaxies, like the Hydra cluster at longitude 265°, and $cz \approx 3500\,\mathrm{km\,s^{-1}}$, are easy to spot; the galaxies seem to lie in a dense line pointing directly at us. This linear appearance is deceptive; it is caused by motions within the cluster. A galaxy's measured radial velocity V_r has two components: the cosmic expansion, and a *peculiar velocity* V_{pec}. Equation 1.20 should be modified to read

$$V_r = H_0 d + V_{\mathrm{pec}}. \tag{1.23}$$

Within rich clusters, the galaxies' orbits give them peculiar velocities up to $1500\,\mathrm{km\,s^{-1}}$; if we use Equation 1.21 to find their positions, they will appear to be closer or more distant than they really are.

Between the clusters, individual galaxies and small groups lie along filaments or in large sheets. In Figure 1.18, the wall-like structure that extends from $l \sim 320°$, $cz \approx 12\,000\,\mathrm{km\,s^{-1}}$ to $l \sim 290°$, $cz \approx 7000\,\mathrm{km\,s^{-1}}$ is $70h^{-1}$ Mpc long. The groups and associations of galaxies within it are less rich than clusters, but more numerous. Our Milky Way and its neighbor Andromeda form part of the *Local Group*, which includes a few dozen smaller systems within a radius of 1–2 Mpc. Between the sheets and filaments are vast nearly empty regions, such as the one at $l \sim 280°$ and $5000\,\mathrm{km\,s^{-1}} \lesssim cz \lesssim 9000\,\mathrm{km\,s^{-1}}$, which is $40h^{-1}$ Mpc across. In th[illegible] few isolated galaxies are found.

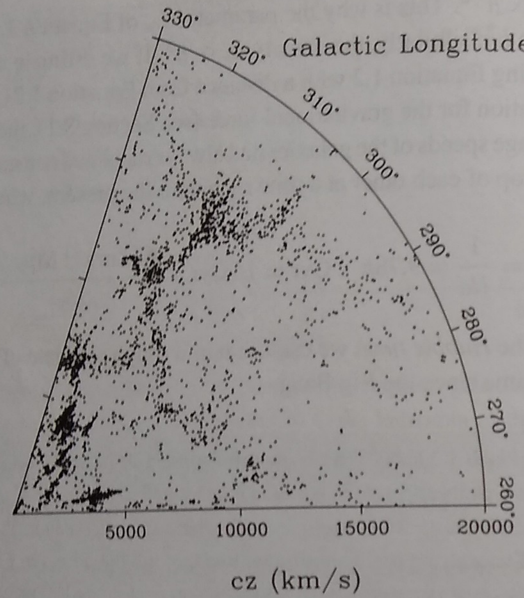

Figure 1.18 Redshifts of galaxies within a slice 70° by 10° of the southern sky, centred at $\alpha \approx 12^h, \delta \approx -15°$. The nearer clusters with $cz \lesssim 8000\,\mathrm{km\,s^{-1}}$ form the Hydra-Centaurus supercluster; most of the remaining galaxies belong to the Shapley supercluster, at $11\,000\,\mathrm{km\,s^{-1}} \lesssim cz \lesssim 16\,000\,\mathrm{km\,s^{-1}}$ – R. Hola.

1.4.1 Densities and ages

If the average density of the Universe is now greater than the *critical density*, the expansion will eventually reverse to a contraction; if it is less, the galaxies will continue to recede from each other forever. In Section 7.2 we will calculate this density as

$$\rho_{\mathrm{crit}}(\mathrm{now}) = \frac{3H_0^2}{8\pi G} = 1.9 \times 10^{-29} h^2\,\mathrm{g\,cm^{-3}}$$
$$= 2.8 \times 10^{11} h^2 \mathcal{M}_\odot\,\mathrm{Mpc^{-3}}. \quad (1.24)$$

For $H_0 = 67\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, the critical density is equivalent to a good-sized galaxy in each megaparsec cube, or about five hydrogen atoms per cubic meter. If the Universe has exactly this density, we will see in Section 7.2 that the time t_0 from the Big Bang to the present day is

$$t_0 = \frac{2}{3H_0} \approx 10\,\mathrm{Gyr} \times \left(\frac{67\,\mathrm{km\,s^{-1}\,Mpc^{-1}}}{H_0}\right). \quad (1.25)$$

If the average density exceeds ρ_{crit}, the Universe is younger than this, while if the density is less, it is older. We shall see that the density is unlikely to be greater than the critical value, so the time since the Big Bang is at least that given by Equation 1.25. But unless the equations of General Relativity are modified by including a *cosmological constant* Λ, which represents a force pushing the galaxies away from each other, the present age t_0 can be no larger than t_H.

Problem 1.12: Use Equation 1.19 to show that for the Universe to be at the critical density, the average ratio of mass to luminosity $\mathcal{M}/L$ would have to be approximately $2000h\,\mathcal{M}_\odot/L_\odot$.

Why is the history of the Universe relevant to our study of galaxies? First, as we will see in Section 2.2, the Hubble time t_H is very close to the ages that we estimate for the oldest stars in our own Galaxy and others. The galaxies, and the stars in them, can be no older than the Universe. If we had confidence in our estimates of the stellar ages, and an accurate measurement of H_0, we could set limits on the present density of the Universe as a whole.

Then, to understand how galaxies came into existence, we must know how much time it took to form the earliest stars, and to build up the elements heavier than helium. The atmospheres of old low-mass stars in galaxies are fossils from the early Universe, preserving a record of the abundances of the various elements in the gas out of which they formed. The theory of stellar evolution provides us with a clock measuring in gigayears how long ago these stars began their lives on the main sequence. The redshifts of distant galaxies tell the time by a different clock, giving information on how long after the Big Bang their light set off on its journey to us. To relate times measured by these two clocks, we must know how the scale of the Universe has changed with time. In Section 7.2 we will see how to calculate the *scale length* $\mathcal{R}(t)$, which grows proportionally to the distance between the galaxies; the Hubble constant H_0 is given by $\dot{\mathcal{R}}(t_0)/\mathcal{R}(t_0)$. For the simplest models, $\mathcal{R}(t)$ depends only on the value of H_0 and the present density $\rho(t_0)$.

Finally, the expansion of the Universe affects the light that we receive from galaxies. Consider two galaxies separated by a distance d, separating at speed $V_r = H_0 d$ according to Equation 1.20. If one of these emits light of wavelength λ_e, an observer in the other galaxy will receive it at a time $\Delta t = d/c$ later, with a longer wavelength $\lambda_{\mathrm{obs}} = \lambda_e + \Delta\lambda$. If the galaxies are fairly close, so that $V_r \ll c$, we can use the Doppler formula of Equation 1.16 to show that the ratio of the wavelengths $\lambda_{\mathrm{obs}}/\lambda_e$ is

$$1 + \frac{\Delta\lambda}{\lambda_e} \approx 1 + \frac{H_0 d}{c} = 1 + H_0 \Delta t = 1 + \left[\frac{1}{\mathcal{R}(t)}\frac{d\mathcal{R}(t)}{dt}\right]_{t_0} \Delta t. \quad (1.26)$$

We can rewrite this as an equation for λ as a function of time:

$$\frac{1}{\lambda}\frac{d\lambda}{dt} = \frac{1}{\mathcal{R}(t)}\frac{d\mathcal{R}(t)}{dt}. \tag{1.27}$$

Integrating gives the formula for the *cosmological redshift z:*

$$1 + z \equiv \frac{\lambda_{obs}}{\lambda_e} = \frac{\mathcal{R}(t_0)}{\mathcal{R}(t_e)}, \tag{1.28}$$

which holds for large redshifts as well as small. Since the wavelength of light expands proportionally to $\mathcal{R}(t)$, its frequency decreases by a factor of $1 + z$. All processes in a distant galaxy appear stretched in time by this same factor; when we observe the distant Universe, we see events taking place in slow motion.

All the topics of this section will be treated in greater depth in Chapters 7 and 8.

1.5 The pregalactic era: a brief history of matter

Here, we sketch what we know of the history of matter in the Universe before the galaxies formed. When a gas is compressed, as in filling a bicycle tire, it heats up; when it is allowed to expand, as in using a pressurized spray can, its temperature drops. The gas of the early Universe was extremely hot and dense, and it has been cooling off during its expansion. This is the *Big Bang* model for the origin of the Universe: the cosmos came into existence with matter at a very high temperature, and expanding rapidly. The physics that we have explored in laboratories on Earth then predicts how this fireball developed into the cosmos that we know today. Two aspects of the early hot phase are especially important for our study of galaxies.

First, the abundance of the lightest elements, hydrogen, deuterium (heavy hydrogen), helium, and lithium, was largely determined by conditions in the first half-hour after the Big Bang. The observed abundance of helium is, amazingly, quite closely what is predicted by the Big Bang model. The measured fraction of deuterium, ^{3}He, and lithium can tell us how much matter the Universe contains. In later chapters, we will compare this figure with the masses that we measure in and around galaxies.

Second, the cosmic microwave background radiation, a relic of the pregalactic Universe, allows us to find our motion relative to the rest of the cosmos. The Milky Way's speed through the cosmic microwave background is our *peculiar velocity*, as defined in Equation 1.23. It turns out to be surprisingly large, indicating that huge concentrations of distant matter have exerted a strong pull on our Local Group.

Further reading: For an undergraduate-level introduction, see M. Lachièze-Rey, *Cosmology: a First Course* (English translation, 1995; Cambridge University Press, Cambridge, UK).

1.5.1 The hot early Universe

For at least the first hundred thousand years after the Big Bang, most of the energy in the Universe was that of the blackbody radiation emitted from the hot matter, and of *relativistic* particles: those moving close to the speed of light, so rapidly that they behave much like photons. During the expansion, Equation 1.28 tells us that wavelengths grow proportionally to the scale length $\mathcal{R}(t)$. By Equation 1.5, the radiation temperature T varies inversely as the wavelength λ_{max} where most light is emitted, and the temperature drops as $T \propto 1/\mathcal{R}(t)$; see the problem below.

Problem 1.13: If photons now fill the cosmos uniformly with number density $n(t_0)$, show that at time t, the density $n(t) = n(t_0)\mathcal{R}^3(t_0)/\mathcal{R}^3(t)$. Use Equation 1.28 to show that the energy density of radiation decreases as $1/\mathcal{R}^4(t)$. For blackbody radiation at temperature T, the number density of photons with energy between ν and $\nu + \Delta\nu$ is

$$n(\nu)\Delta\nu = \frac{2\nu^2}{c^3}\frac{\Delta\nu}{\exp(h\nu/k_B T) - 1}. \tag{1.29}$$

Show that if the present spectrum is that of blackbody radiation at temperature T_0, then at time t the expansion transforms this exactly into blackbody radiation at temperature $T(t) = T_0\mathcal{R}(t_0)/\mathcal{R}(t)$.

In the first 3 minutes of its life, the Universe was full of energetic gamma rays, which would smash any atomic nuclei apart into their consituent particles. When the temperature of the radiation field is high enough, pairs of particles and their antiparticles can be created out of the vacuum. Because photons are never at rest, two of them are required to produce a particle pair. A typical photon of radiation at temperature T carries energy $\mathcal{E} = 4k_BT$, where k_B is Boltzmann's constant; so proton-antiproton pairs could be produced when $k_BT \gtrsim m_pc^2$, where m_p is the proton mass. We usually measure these energies in units of an *electron volt*, the energy that an electron gains by moving through a potential difference of 1 volt: $1\ \text{eV} = 1.6 \times 10^{-19}$ J or 1.6×10^{-12} erg. In these units, $m_pc^2 \approx 10^9$ eV or 1 GeV, so pairs of protons and antiprotons were created freely in the first 10^{-4} s, when

$$k_BT \gg m_pc^2, \quad \text{or} \quad T \gg 10^{13}\ \text{K}. \tag{1.30}$$

As the expansion continued, the temperature fell, and the photons had too littl[...] [...]oton pair; almost all the antiprotons met with a proton and annihilated to leave a pair of γ-rays. We do not understand why this was so, but in the early Universe there were slightly more protons: about $10^9 + 1$ protons for every 10^9 antiprotons. The small excess of matter over antimatter was left over to form the galaxies. The photons produced in the annihilation are seen today as the cosmic microwave background.

Electrons are about 2000 times less massive than protons; their rest energy $m_e c^2$ is only 0.5 MeV. So the radiation still produced pairs of electrons and anti-electrons (*positrons*, e^+), until the temperature dropped a thousandfold, to $T \sim 10^{10}$ K. Before this time, the reaction

$$e^- + e^+ \longleftrightarrow \nu_e + \bar{\nu}_e$$

could produce *electron neutrinos* ν_e and their antiparticles $\bar{\nu}_e$. The great abundance of electrons, positrons, and neutrinos allowed neutrons to turn into protons, and vice versa, through reactions such as

$$e^- + p \longleftrightarrow n + \nu_e, \quad \bar{\nu}_e + p \longleftrightarrow n + e^+, \quad n \longleftrightarrow p + e^- + \bar{\nu}_e.$$

In equilibrium at temperature T, there would have been slightly fewer neutrons than protons, since the neutron mass m_n is larger. The ratio of neutrons to protons was given by

$$n/p = e^{-Q/k_BT}, \quad \text{where } Q = (m_n - m_p)c^2 = 1.293\,\text{MeV}. \tag{1.31}$$

Neutrinos are very weakly interacting particles; from the Sun, 10^{15} of them fly harmlessly through each square meter of the Earth's surface every second. Only in the extremely hot material of supernova cores, or in the early Universe, do they have an appreciable chance of reacting with other particles. While electron-positron pairs were still numerous, the density of neutrinos was high enough to keep the balance between neutrons and protons at this equilibrium level. But later, once $k_BT \lesssim 0.8$ MeV or $t \gtrsim 1$ s, expansion had cooled the matter and neutrinos so much that a neutron or proton was very unlikely to interact with a neutrino. The neutrons *froze out*, with $n/p \approx 1/5$.

1.5.2 Making the elements

Neutrons can survive if they are bound up in the nuclei of atoms, but free neutrons are not stable; they decay into a proton, an electron, and an antineutrino $\bar{\nu}_e$. The half-life, the time taken for half of them to decay, is measured as 887 ± 2 s. Very few neutrons would now be left, if they had not combined with protons to form *deuterium*, a nucleus of 'heavy hydrogen' containing a neutron and a proton, by the reaction

$$n + p \to D + \gamma;$$

here γ represents a photon, a γ-ray carrying away the 2.2 MeV of energy set free in the reaction. This reaction also took place at earlier times, but any deuterium that managed to form was immediately torn apart by photons in the blackbody radiation.

After the electron-positron pairs were gone, at $T \lesssim 3 \times 10^9$ K, the energy density in the Universe was almost entirely due to blackbody radiation. The general theory of relativity tells us that the temperature fell according to

$$t = \left(\frac{3c^2}{32\pi G a_B T^4}\right)^{1/2} \approx 230\,\text{s}\left(\frac{10^9\,\text{K}}{T}\right)^2; \tag{1.32}$$

here $a_B = 7.56 \times 10^{-16}\,\text{J m}^{-3}\,\text{K}^{-4}$ is the blackbody constant. About a quarter of the neutrons had decayed before the temperature fell to about 10^9 K, when they could be locked into deuterium; this left about one neutron for every seven protons. The excess protons, which became the nuclei of hydrogen atoms, accounted for about 75% of the total mass.

Deuterium easily combines with other particles to form ^{4}He, a helium nucleus with two protons and two neutrons. Essentially all the neutrons, and so about 25% of the total mass of neutrons and protons, ended up in ^{4}He. Only a little deuterium and some ^{3}He (with two protons and one neutron) remained. Traces of boron and lithium were also formed, but the Universe expanded too rapidly to build up heavier nuclei. The amount of helium produced depends on the half-life of the neutron, but hardly at all on the density of matter at that time; almost every neutron could find a proton and form deuterium, and almost every deuterium nucleus reacted to make helium. The observed abundance of helium is between 22% and 24%, in rough accord with this calculation. If, for example, we had found the Sun to contain 10% of helium by weight, that observation would have been very hard to explain in the Big Bang cosmology.

Problem 1.14: Use Equation 1.32 to show that $k_BT = 100$ MeV is reached at $t \approx 450$ s, at which time about 1/4 of the free neutrons have decayed. The neutron half-life is hard to measure; until recently, laboratory values varied from 700 s to 1400 s. If the half-life had been 700 s, show that the predicted fraction of helium would be about 1% lower, while if it had been 1100 s, we would expect to find close to 1% more helium.

By contrast, the small fraction of deuterium left over is very much dependent on the density of neutrons and protons, collectively known as *baryons*. If there had been very little matter, many of the deuterium nuclei would have missed the chance to collide with other particles, before reactions ceased as the Universe became too dilute. If the present number of baryons had been as low as $n_B = 10^{-8}\,\text{cm}^{-3}$, then as many as 1% of the deuterium nuclei would remain. If the density was now as high as $n_B = 2 \times 10^{-6}\,\mathrm{cm}^{-3}$, we would expect to find less than one deuterium nucleus for each 10^9 atoms of hydrogen.

Deuterium also burns readily to helium inside stars. Thus to measure how much was made in the Big Bang, we must look for old metal-poor stars that have not burned the deuterium in their outer layers, or at intergalactic clouds of gas that have not yet formed many stars. Our best measurements show one deuterium nucleus for every 20 000 or 30 000 atoms of hydrogen. Along with measurements of ^{3}He and lithium, these show that the combined density of neutrons and protons today is

$$n_B = (1.4 \pm 0.3) \times 10^{-7}\,\mathrm{cm}^{-3}, \quad \text{or} \quad \rho_B = (3.5 \pm 0.8) \times 10^9 \mathcal{M}_\odot\,\mathrm{Mpc}^{-3}. \tag{1.33}$$

This is much less than the critical density of Equation 1.24: the ratio is

$$0.009h^{-2} \lesssim \rho_B/\rho_{\mathrm{crit}} \lesssim 0.015h^{-2}, \tag{1.34}$$

where h is Hubble's constant H_0 in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Observations seem to require $h \gtrsim 0.4$, so neutrons and protons cannot make up more than about 10% of the critical density. Either the Universe will continue to expand forever, or some other form of matter contributes substantial mass. Perhaps the cosmos is filled with particles that, like neutrinos, are weakly interacting – or we should have seen them – but are massive enough to make up the balance of ρ_{crit}: collectively, these are known as weakly interacting massive particles, or WIMPs. Since $h \lesssim 0.8$, baryons account for no less than 2% of the critical density. We will find in Section 5.3 that this is more mass than we can see as the gas and luminous stars of galaxies; at least part of their 'dark stuff' must consist of normal matter.

Problem 1.15: The Universe must contain at least as much matter as that of the neutrons and protons: use Equations 1.33 and 1.19 to show that the average mass-to-light ratio must exceed $(20\text{–}30)h^{-1}\mathcal{M}_\odot/L_\odot$.

1.5.3 Recombination: light and matter uncoupled

The next few hundred thousand years of the Universe's history were rather boring. Its density had dropped too low for nuclear reactions, and the background radiation was energetic enough to ionize hydrogen and disrupt other atoms. The cosmos was filled with glowing gas, like the inside of a fluorescent light. Photons could not pass freely through this hot plasma, since they were scattered by free electrons. Matter would not collapse under its own gravity to form stars or other dense objects, because the pressure of the radiation trapped inside was too high.

The density of radiation decreases with the scale length $\mathcal{R}(t)$ as $T^4 \propto \mathcal{R}^{-4}(t)$. So after some time it must drop below that of matter, which falls only as $\mathcal{R}^{-3}(t)$. By

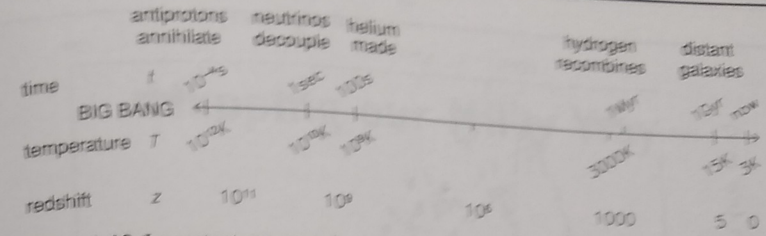

Figure 1.19 Important moments in the history of the Universe.

coincidence, this time of *matter-radiation equality* is close to the time when the radiation cooled to $T \approx 4000$ K. As measured by $\mathcal{R}(t)$, the Universe was then about 1/1500 of its present size. Later on, photons of the blackbody radiation lacked the energy to remove the electron from a hydrogen atom. During its subsequent expansion, hydrogen atoms *recombined*, the gas becoming neutral and transparent as it is today. By the time that $\mathcal{R}(t)/\mathcal{R}(t_0) \approx 1/1100$, photons of the background radiation were able to escape from the matter. Their outward pressure no longer prevented the collapse of matter into the galaxies and clusters that we now observe.

The most distant galaxies so far observed are at redshifts $z \sim 5$; when their light left them, the Universe was only about 1 Gyr old. Figure 1.19 presents a brief summary of cosmic history up to that time.

The radiation coming to us from the period of recombination has been redshifted according to Equation 1.28; it now has a much longer wavelength. Its temperature $T = 2.728 \pm 0.002$ K, so it is known as the *cosmic microwave background*. There are about 420 of these photons in each cm^3 of space, so according to Equation 1.34, we have $2 - 4 \times 10^9$ photons for every neutron or proton. The energy density of the background radiation is about equal to that of starlight in the outer reaches of the Milky Way. It is given by $a_B T^4 = 4.2 \times 10^{-14}\,\mathrm{J\,m^{-3}}$; so from each steradian of the sky we receive $c a_B T^4/4\pi \approx 10^{-6}\,\mathrm{W\,m^{-2}}$.

Problem 1.16: Using Equation 1.19, show that even if we ignore the energy loss that goes along with the redshift, it would take more than 100 Gyr for all the galaxies, at their present luminosity, to emit as much energy as is in the microwave background today.

Figure 1.20 shows the extragalactic background radiation, estimated by observing from our position in the Milky Way and attempting to subtract local contributions. The energy of the cosmic background is far larger than that in the infrared, visible, and ultraviolet spectral regions. It would be very difficult to explain such enormous energy as coming from any other source than the Big Bang. Radiation from the submillimeter region through to the ultraviolet at $\lesssim 0.1$ keV comes from stars and active galactic nuclei, either directly or after reradiation by heated dust. The high-energy 'tail' in X-rays and γ-rays is mainly from active nuclei. Since photons lose energy in an expanding Universe, almost all this

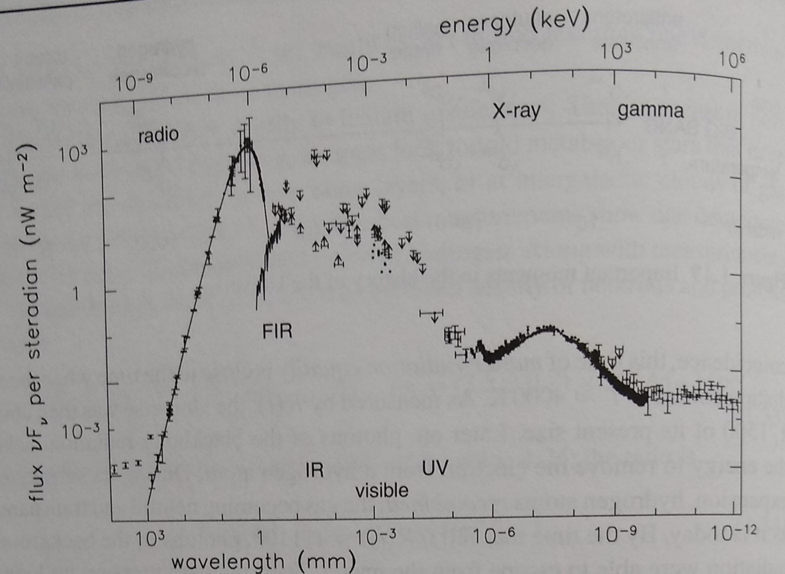

Figure 1.20 Extragalactic background radiation: vertical *logarithmic* scale shows energy density per decade in frequency or wavelength. Arrows show upper and lower limits. The curve peaking at $\lambda \sim 1$ mm is the cosmic microwave background; the far-infrared background is the light of stars and active galactic nuclei, re-radiated by dust – T. Ressell, D. Scott.

radiation, aside from that in the microwave background, must have been emitted over the past ~10 Gyr, at times corresponding to redshifts $z \lesssim 3$.

The microwave background is now very close to a blackbody spectrum; it is also extremely uniform. Between different parts of the sky, we see small irregularities in its temperature that are just a few parts in 100 000 – with only one exception. In the direction $l = 265°$, $b = 48°$ the peak wavelength is shorter than average, and the temperature higher, by a little more than 0.1%. In the opposite direction, the temperature is lower by the same amount. This difference reflects the Sun's motion through the background radiation. If T_0 is the temperature measured by an observer at rest relative to the backgound radiation, then an observer moving with relative speed $V \ll c$ would measure a temperature $T(\theta)$ at an angle θ to the direction of motion, given by

$$T(\theta) \approx T_0(1 + V \cos\theta/c). \tag{1.35}$$

For the Sun, $V = 370\,\mathrm{km\,s^{-1}}$. Taking into account the Sun's orbit about the Milky Way, and the Milky Way's motion relative to nearby galaxies, we find that our Local Group has a peculiar motion of $V_{pec} \approx 600\,\mathrm{km\,s^{-1}}$ relative to the background radiation and to the Universe as a whole. The Local Group's motion is unexpectedly and troublingly large: we discuss it further in Chapter 7.