Cosmology is a science studying the structure and evolution of the whole Universe. Research in cosmology tries to answer questions such as: How large and how old is the Universe? How was it born? How is matter distributed? How were the elements formed? What will be the future of the Universe?

After the demise of the Aristotelian world picture, it took hundreds of years of astronomical observations and physical theories to reach a level at which a satisfactory modern scientific picture of the physical universe could be formed. The decisive steps in the development were the clarification of the nature of the galaxies in the 1920's and the general theory of relativity, which gives a basis for theoretical research of the universe, developed by Einstein in the 1910's. Yet, just a few decades ago there was a lot of room for armwaiving and all kinds of speculations. Only recently the observations made with satellites have given reasonably precise values for some parameters and some alternative models could be excluded. The central tenet of modern cosmology is the model of the expanding universe, getting all the time more supporting observational evidence.

20.1 Cosmological Observations

The Olbers Paradox The simplest cosmological observation may be that the sky is dark at night. This fact was first noted by *Johannes Kepler*, who, in 1610, used it as evidence for a finite universe. As the idea of an infinite space filled with stars like the Sun became widespread in consequence of the Copernican revolution, the question of the dark night sky remained a problem. In the 18th and 19th centuries *Edmond Halley*, *Loys de Chéseaux* and *Heinrich Olbers* considered it in their writings. It has become known as the *Olbers paradox* (Fig. 20.1).

The paradox is the following: Let us suppose the Universe is infinite and that the stars are uniformly distributed in space. No matter in what direction one looks, sooner or later the line of sight will encounter the surface of a star. Since the surface brightness does not depend on distance, each point in the sky should appear to be as bright as the surface of the Sun. This clearly is not true. The modern explanation of the paradox is that the stars have only existed for a finite time, so that the light from very distant stars has not yet reached us. Rather than proving the world to be finite in space, the Olbers paradox has shown it to be of a finite age.

Extragalactic Space In 1923 *Edwin Hubble* showed that the Andromeda Galaxy M31 was far outside the Milky Way, thus settling a long-standing controversy concerning the relationship between the nebulae and the Milky Way. The numerous galaxies seen in photographs form an extragalactic space vastly larger than the dimensions of the Milky Way. It is important for cosmology that the distribution and motions of the basic components of extragalactic space, the galaxies and clusters of galaxies, should everywhere be the same as in our local part of the

Fig. 20.1 The Olbers paradox. If the stars were uniformly distributed in an unending, unchanging space, the sky should be as bright as the surface of the Sun, since each line of sight would eventually meet the surface of a star. A two-dimensional analogy can be found in an optically thick pine forest where the line of sight meets a trunk wherever one looks. (Photo M. Poutanen and H. Kart-tunen)

Universe. Galaxies generally occur in various systems, ranging from small groups to clusters of galaxies and even larger superclusters. The largest structures observed are about 100 Mpc in size (see Sect. 19.6). They are thus significantly smaller than the volume of space (a few thousand Mpc in size) in which the distribution of galaxies has been investigated. One way of studying the large-scale homogeneity of the galaxy distribution is to count the number of galaxies brighter than some limiting magnitude m. If the galaxies are uniformly distributed in space, this number should be proportional to $10^{0.6m}$ (see Example 18.1). For example, the galaxy counts made by Hubble in 1934, which included 44,000 galaxies, were consistent with a galaxy distribution independent of position (homogeneity) and of direction (isotropy). Hubble found no "edge" of the Universe, nor have later galaxy counts found one.

Similar counts have been made for extragalactic radio sources. (Instead of magnitudes, flux densities are used. If F is the flux density, then because $m = -2.5 \lg(F/F_0)$, the number count will be proportional to $F^{-3/2}$.) These counts mainly involve very distant radio galaxies and quasars (Fig. 20.2). The results seem to indicate that the radio sources were either much brighter or much more common at earlier epochs than at present (Sect. 19.8). This constitutes evidence in favour of an evolving, expanding universe.

In general the simple geometric relation between brightness and number counts will only hold for objects that are uniformly distributed in space. Local inhomogeneities will cause departures from the expected relationship. For more distant sources the geometry of the Universe as well as cosmic evolution will change the basic $10^{0.6m}$ behaviour.

Hubble's Law (See Fig. 20.3.) In the late 1920's, Hubble discovered that the spectral lines of galaxies were shifted towards the red by an amount proportional to their distances. If the redshift is due to the Doppler effect, this means that the galaxies move away from each other with velocities proportional to their separations, i.e. that the Universe is expanding as a whole.

In terms of the redshift $z = (\lambda - \lambda_0)/\lambda_0$, Hubble's law can be written as

$$z = (H/c)r, \tag{20.1}$$

where c is the speed of light, H is the *Hubble constant* and r the distance of the galaxy. For small velocities ($V \ll c$) the Doppler redshift $z = V/c$, and hence

$$V = Hr, \tag{20.2}$$

which is the most commonly used form of Hubble's law.

For a set of observed "standard candles", i.e. galaxies whose absolute magnitudes are close to some mean M_0, Hubble's law corresponds to

Fig. 20.2 The quasar 3C295 and its spectrum. The quasars are among the most distant cosmological objects. (Photograph Palomar Observatory)
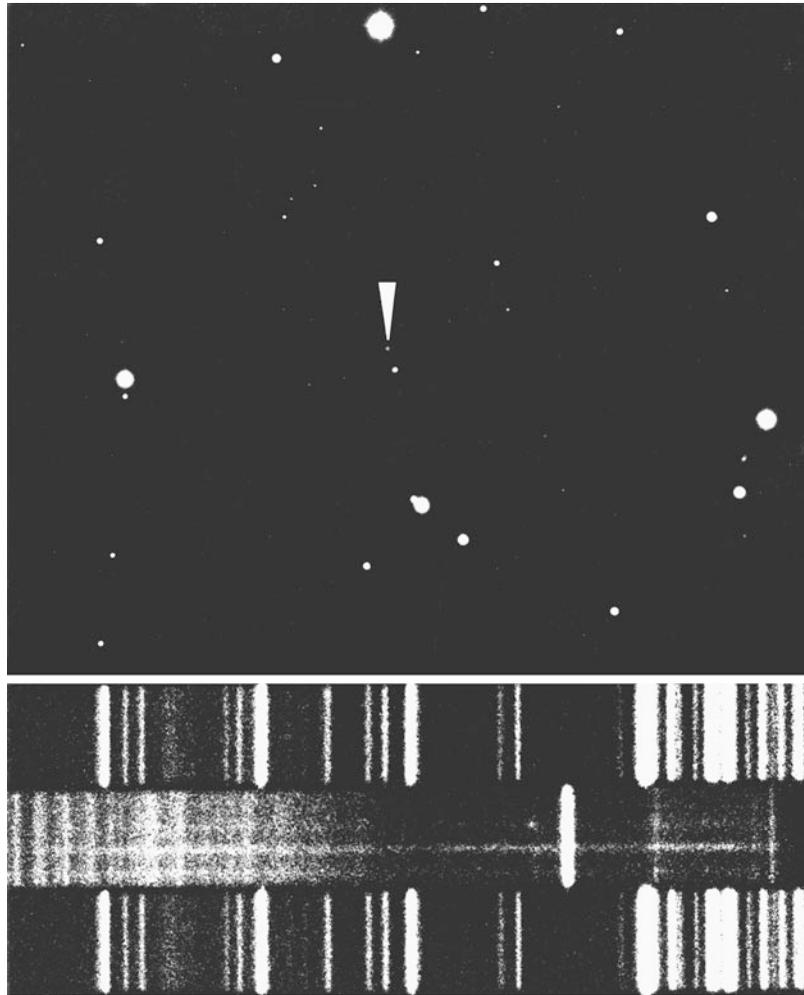
a linear relationship between the apparent magnitude m and the logarithm of the redshift, $\lg z$. This is because a galaxy at distance r has an apparent magnitude $m = M_0 + 5\lg(r/10 \text{ pc})$, and hence Hubble's law yields

$$m = M_0 + 5\lg\left(\frac{cz}{H \times 10 \text{ pc}}\right) = 5\lg z + C,$$
(20.3)

where the constant C depends on H and M_0. Suitable standard candles are e.g. the brightest galaxies in clusters and Sc galaxies of a known luminosity class. Some other methods of distance determination for galaxies were discussed in Sect. 19.2. Most recently, type Ia supernovae (Sect. 14.4) in distant galaxies have been used to determine distances out to the redshift $z = 1.7$,

where departures from Hubble's law are already detectable.

If the Universe is expanding, the galaxies were once much nearer to each other. If the rate of expansion had been unchanging, the inverse of the Hubble constant, $T = H^{-1}$, would represent the age of the Universe. If the expansion is gradually slowing down, the inverse Hubble constant gives an upper limit on the age of the Universe In fact, current indications (discussed later in this chapter) are that the rate of expansion is accelerating at present. In that case the age of the Universe may also be larger. However, H^{-1} will still be an estimate for the age of the Universe.

One reason for the difficulty in determining the value of the Hubble constant is the uncer-

Fig. 20.3 Hubble's law for type Ia supernovae. *The solid curve* represents the "concordance" model, with basic parameters having the values mentioned in Sect. 20.7. Models with a zero cosmological constant cannot explain the observations. (D. Rubin et al., 2013, ApJ 763, 35)
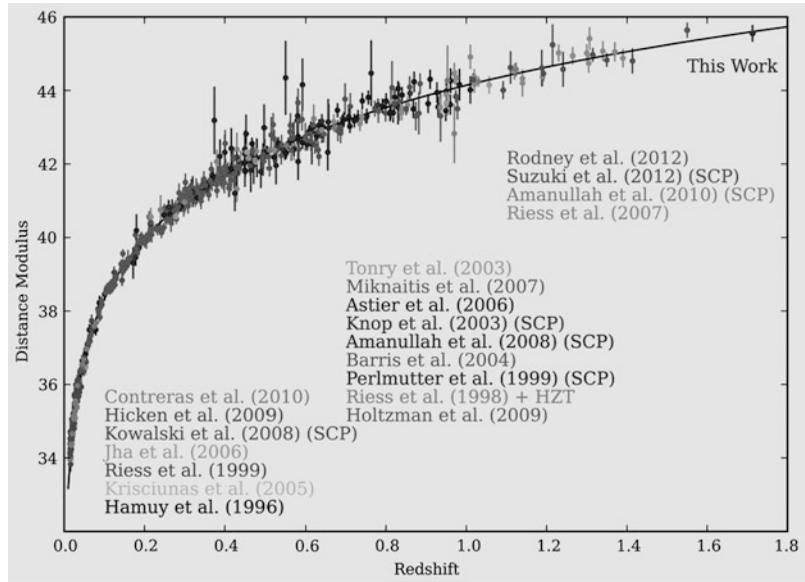
tainty in extragalactic distances. A second problem is that the measured values of the velocity V, corrected to take into account the motion of the Sun within the Local Group, contain a significant component due to the peculiar motions of the galaxies. These peculiar velocities are caused by local mass concentrations like groups and clusters of galaxies. It is possible that the Local Group has a significant velocity towards the centre of the Local Supercluster (the Virgo Cluster). Because the Virgo Cluster is often used to determine the value of H, neglecting this peculiar velocity leads to a large error in H. The size of the peculiar velocity is not yet well known, but it is probably about $250 \, \mathrm{km \, s^{-1}}$.

The most ambitious recent project for determining H used the Hubble Space Telescope in order to measure cepheid distances to a set of nearby galaxies. These distances were then used to calibrate other distance indicators, such as the Tully–Fisher relation and type Ia supernovae. The final result was $H = (72 \pm 8) \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$. The largest remaining source of error in this result is the distance to the Large Magellanic Cloud, used for calibrating the cepheid luminosity.

Nowadays the most accurate value for the Hubble constant is obtained by studying the angular power spectrum of the background radiation (Sect. 20.7). According to the mea-
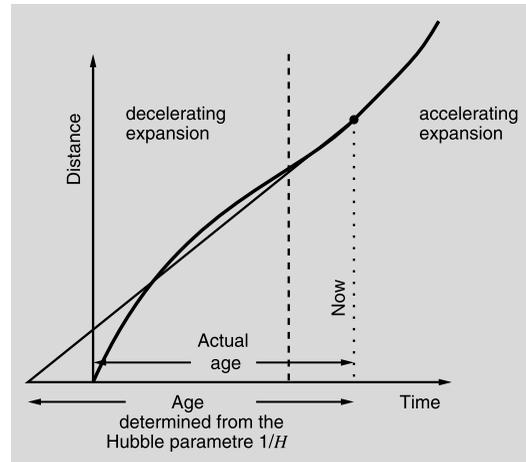

Fig. 20.4 It is nowadays assumed that at the beginning the expansion rate of the Universe was slowing down until about 5 billion years ago the deceleration changed to acceleration. The age of the Universe calculated from the current value of the Hubble constant is bigger than the actual age

surements of the Planck satellite $H = (67.7 \pm 0.5) \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ and the age of the Universe is 13.80 ± 0.02 Ga.

The form of Hubble's law might give the impression that the Milky Way is the centre of the expansion, in apparent contradiction with the Copernican principle. Figure 20.5 shows that, in fact, the same Hubble's law is valid at each point
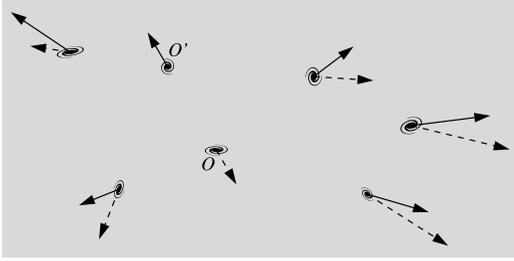
Fig. 20.5 A regular expansion according with Hubble's law does not mean that the Milky Way (O) is the centre of the Universe. Observers at any other galaxy (O') will see the same Hubble flow (*dashed lines*)

Fig. 20.6 Observations of the cosmic microwave background radiation made by the COBE satellite in 1990 are in agreement with a blackbody law at 2.7 K

in a regularly expanding universe. There is no particular centre of expansion.

The Thermal Microwave Background Radiation The most important cosmological discovery since Hubble's law was made in 1965. In that year *Arno Penzias* and *Robert Wilson* discovered that there is a universal microwave radiation, with a spectrum corresponding to that of blackbody radiation (see Sect. 5.6) at a temperature of about 3 K (Fig. 20.7). For their discovery, they received the Nobel prize in physics in 1979.

The existence of a thermal cosmic radiation background had been predicted in the late 1940's by *George Gamow*, who was one of the first to study the initial phases of expansion of the Universe. According to Gamow, the Universe at that time was filled with extremely hot radiation. As it expanded, the radiation cooled, until at present, its temperature would be a few kelvins. After its discovery by Penzias and Wilson, the cosmic background radiation has been studied at wavelengths from 50 cm to 0.5 cm. The first detailed measurements, made from the COBE (Cosmic Background Explorer) satellite showed that it corresponds closely to a Planck spectrum at (2.725 ± 0.002) K (Fig. 20.6). More recently the CMB has been mapped in even greater detail by the WMAP satellite.

The existence of the thermal cosmic microwave background (CMB) gives strong support to the belief that the Universe was extremely hot in its early stages. The background is very nearly isotropic, which supports the isotropic and homogeneous models of the Universe. The COBE
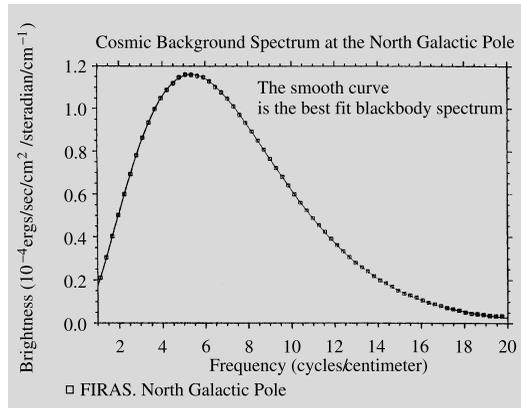
and WMAP satellites have also detected temperature variations of a relative amplitude 6×10^{-6} in the background. These fluctuations are interpreted as a gravitational redshift of the background produced by the mass concentrations that would later give rise to the observed structures in the Universe. They are the direct traces of initial irregularities in the big bang, and provide important constraints for theories of galaxy formation. Perhaps even more importantly, the amplitude of the fluctuations on different angular scales have provided crucial constraints on the cosmological model. We shall return to this question in Sect. 20.7.

The Isotropy of Matter and Radiation Apart from the CMB, several other phenomena confirm the isotropy of the Universe. The distribution of radio sources, the X-ray background, and faint distant galaxies, as well as Hubble's law are all isotropic. The observed isotropy is also evidence that the Universe is homogeneous, since a large-scale inhomogeneity would be seen as an anisotropy.

The Age of the Universe Estimates of the ages of the Earth, the Sun and of star clusters are important cosmological observations that do not depend on specific cosmological models. From the decay of radioactive isotopes, the age of the Earth is estimated to be 4600 million years. The age of the Sun is thought to be slightly larger than this.

Fig. 20.7 The cosmological principle. In the small circle (*A*) about the observer (*O*) the distribution of galaxies does not yet represent the large-scale distribution. In the larger circle (*B*) the distribution is already uniform on the average

20.2   The Cosmological Principle
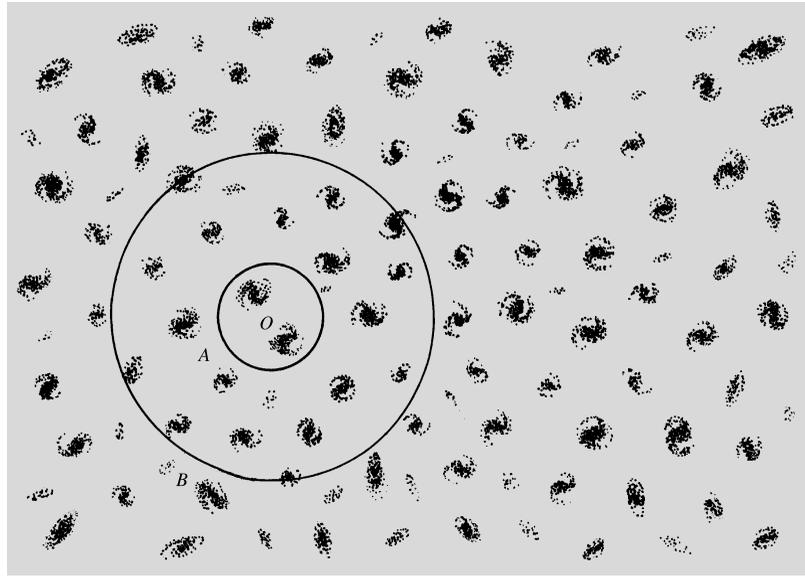
The ages of the oldest star clusters in the Milky Way are 10–15 Ga.

The values thus obtained give a lower limit to the age of the Universe. In an expanding universe, the inverse Hubble constant gives another estimate of that age. It is most remarkable that the directly determined ages of cosmic objects are so close to the age given by the Hubble constant. This is strong evidence that Hubble's law really is due to the expansion of the Universe. It also shows that the oldest star clusters formed very early in the history of the Universe.

The Relative Helium Abundance   A cosmological theory should also give an acceptable account of the origin and abundances of the elements. Even the abundance of the elementary particles and the lack of antimatter are cosmological problems that have begun to be investigated in the context of theories of the early Universe.

Observations show that the oldest objects contain about 25 % by mass of helium, the most abundant element after hydrogen. The amount of helium produced is sensitive to the temperature of the Universe, which is related to that of the background radiation. The computations made for the standard models of the expanding Universe (the Friedmann models) yield a helium abundance of exactly the right size.

One hopes that as ever larger volumes of the Universe are observed, its average properties will become simple and well defined. Figure 20.7 attempts to show this. It shows a distribution of galaxies on a plane. As the circle surrounding the observer *O* becomes larger, the mean density inside the circle becomes practically independent of its size. The same behaviour occurs, regardless of the position of the centre of *O*: at close distances, the density varies randomly (Fig. 20.8), but in a large enough volume, the average density is constant. This is an example of the *cosmological principle:* apart from local irregularities, the Universe looks the same from all positions in space.

The cosmological principle is a far-reaching assumption, which has been invoked in order to put constraints on the large variety of possible cosmological theories. If in addition to the cosmological principle one also assumes that the Universe is isotropic, then the only possible cosmic flow is a global expansion. In that case, the local velocity difference V between two nearby points has to be directly proportional to their separation ($V = Hr$); i.e. Hubble's law must apply.

The plane universe of Fig. 20.7 is homogeneous and isotropic, apart from local irregular-

Fig. 20.8 The galaxies seem to be distributed in a "foamlike" way. Dense strings and shells are surrounded by relatively empty regions. (Seldner, M. et al. (1977): Astron. J. **82**, 249)
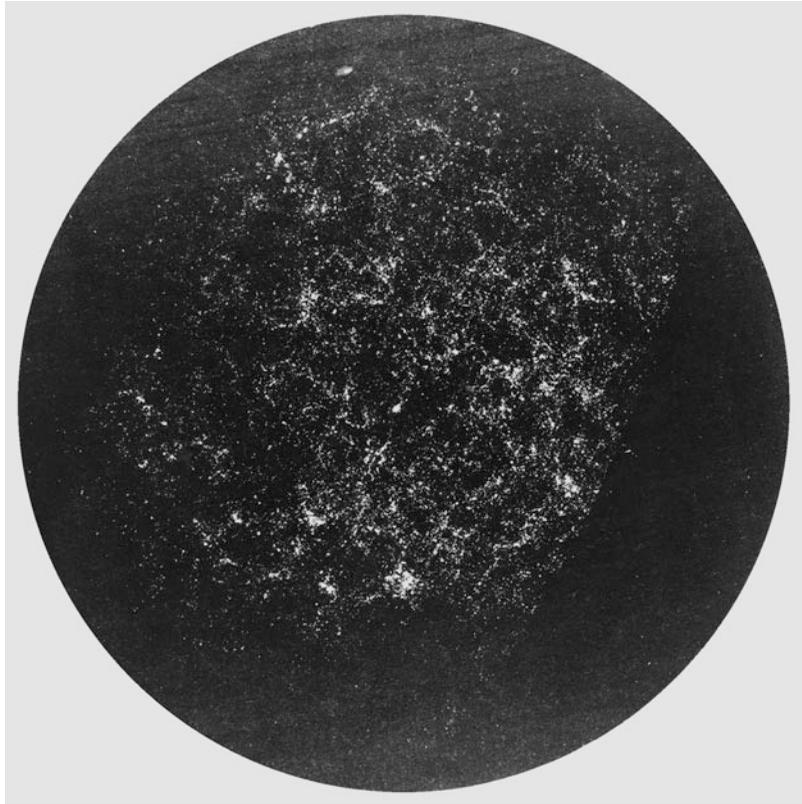
ities. Isotropy at each point implies homogeneity, but homogeneity does not require isotropy. An example of an anisotropic, homogeneous universe would be a model containing a constant magnetic field: because the field has a fixed direction, space cannot be isotropic.

We have already seen that astronomical observations support the homogeneity and isotropy of our observable neighbourhood, the *metagalaxy*. On the grounds of the cosmological principle, these properties may be extended to the whole of the Universe.

The cosmological principle is closely related to the Copernican principle that our position in the Universe is in no way special. From this principle, it is only a short step to assume that on a large enough scale, the local properties of the metagalaxy are the same as the global properties of the Universe.

Homogeneity and isotropy are important simplifying assumptions when trying to construct *cosmological models* which can be compared with local observations. They may therefore reasonably be adopted, at least as a preliminary hypothesis.

20.3 Homogeneous and Isotropic Universes

Under general conditions, space and time coordinates in a universe may be chosen so that the values of the space coordinates of observers moving with the matter are constant. It can be shown that in a homogeneous and isotropic universe, the line element (Appendix B) then takes the form

$$\mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + R^2(t)$$

$$\times \left[\frac{\mathrm{d}r^2}{1 - kr^2} + r^2\big(\mathrm{d}\theta^2 + \cos^2\theta\,\mathrm{d}\phi^2\big) \right],$$

(20.4)

known as the *Robertson–Walker line element*. (The radial coordinate r is defined to be dimensionless.) $R(t)$ is a time-dependent quantity representing the scale of the Universe. If R increases
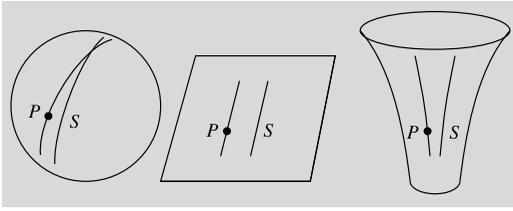
Fig. 20.9 The two-dimensional analogues of the Friedmann models: a spherical surface, a plane and a saddle surface

Fig. 20.10 When space expands, all galaxy separations grow with the scale factor R: $r' = [R(t'_0)/R(t_0)]r$

with time, all distances, including those between galaxies, will grow. The coefficient k may be $+1$, 0 or -1, corresponding to the three possible geometries of space, the *elliptic* or *closed*, the *parabolic* and the *hyperbolic* or *open* model.

The space described by these models need not be Euclidean, but can have positive or negative curvature. Depending on the curvature, the volume of the universe may be finite or infinite. In neither case does it have a visible edge.

The two-dimensional analogy to elliptical ($k = +1$) geometry is the surface of a sphere (Fig. 20.9): its surface area is finite, but has no edge. The scale factor $R(t)$ represents the size of the sphere. When R changes, the distances between points on the surface change in the same way. Similarly, a three-dimensional "spherical surface", or the space of elliptical geometry, has a finite volume, but no edge. Starting off in an arbitrary direction and going on for long enough, one always returns to the initial point (unless the distance is increased too much by the expansion).

When $k = 0$, space is flat or Euclidean, and the expression for the line element (20.4) is almost the same as in the Minkowski space. The only difference is the scale factor $R(t)$. All distances in a Euclidean space change with time. The two-dimensional analogue of this space is a plane.

The volume of space in the hyperbolic geometry ($k = -1$) is also infinite. A two-dimensional idea of the geometry in this case is given by a saddle surface or a horn becoming infinitely wider.

In a homogeneous and isotropic universe, many physical quantities will depend on time through the scale factor $R(t)$. For example, from the form of the line element, it is evident that all distances will be proportional to R (Fig. 20.10).
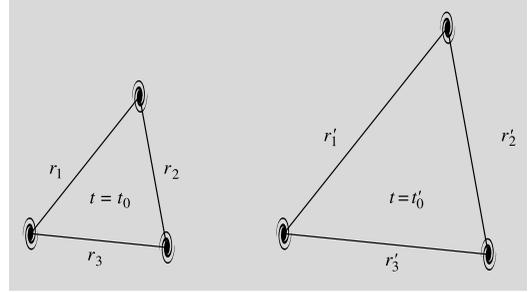
Thus, if the distance to a galaxy is r at time t, then at time t_0 (in cosmology, the subscript 0 refers to the present value) it will be

$$\frac{R(t_0)}{R(t)} r. \tag{20.5}$$

Similarly, all volumes will be proportional to R^3. From this it follows that the density of any conserved quantity (e.g. mass) will behave as R^{-3}.

It can be shown that the wavelength of radiation in an expanding universe is proportional to R, like all other lengths. If the wavelength at the time of emission, corresponding to the scale factor R, is λ, then it will be λ_0 when the scale factor has increased to R_0:

$$\frac{\lambda_0}{\lambda} = \frac{R_0}{R}. \tag{20.6}$$

The redshift is $z = (\lambda_0 - \lambda)/\lambda$, and hence

$$1 + z = \frac{R_0}{R}; \tag{20.7}$$

i.e. the redshift of a galaxy expresses how much the scale factor has changed since the light was emitted. For example, the light from a quasar with $z = 1$ was emitted at a time when all distances were half their present values.

For small values of the redshift, (20.7) approaches the usual form of Hubble's law. This can be seen as follows. When z is small, the change in R during the propagation of a light signal will also be small and proportional to the light travel time t. Because $t = r/c$ approximately, where r is the distance of the source, the redshift will be

proportional to r. If the constant of proportionality is denoted by H/c, one has

$$z = Hr/c. \qquad (20.8)$$

This is formally identical to Hubble's law (20.1). However, the redshift is now interpreted in the sense of (20.7).

As the universe expands, the photons in the background radiation will also be redshifted. The energy of each photon is inversely proportional to its wavelength, and will therefore behave as R^{-1}. It can be shown that the number of photons will be conserved, and thus their number density will behave as R^{-3}. Combining these two results, one finds that the energy density of the background radiation is proportional to R^{-4}. The energy density of blackbody radiation is proportional to T^4, where T is the temperature. Thus the temperature of cosmic background radiation will vary as R^{-1}.

20.4 The Friedmann Models

The results of the preceding section are valid in any homogeneous and isotropic universe. In order to determine the precise time-dependence of the scale factor $R(t)$ a theory of gravity is required.

In 1917 *Albert Einstein* presented a model of the Universe based on his general theory of relativity. It described a geometrically symmetric (spherical) space with finite volume but no boundary. In accordance with the cosmological principle, the model was homogeneous and isotropic. It was also *static*: the volume of space did not change.

In order to obtain a static model, Einstein had to introduce a new repulsive force, the *cosmological term*, in his equations. The size of this cosmological term is given by the *cosmological constant* Λ. Einstein presented his model before the redshifts of the galaxies were known, and taking the Universe to be static was then reasonable. When the expansion of the Universe was discovered, this argument in favour of the cosmological constant vanished. Einstein himself later called it the biggest blunder of his life. Nevertheless, the most recent observations now seem to indicate that a non-zero cosmological constant has to be present.

The St. Petersburg physicist *Alexander Friedmann* and later, independently, the Belgian *Georges Lemaître* studied the cosmological solutions of Einstein's equations. If $\Lambda = 0$, only evolving, expanding or contracting models of the Universe are possible. From the Friedmann models exact formulas for the redshift and Hubble's law may be derived.

The general relativistic derivation of the law of expansion for the Friedmann models will not be given here. It is interesting that the existence of three types of models and their law of expansion can be derived from purely Newtonian considerations, with results in complete agreement with the relativistic treatment. The detailed derivation is given on Box 20.1, but the essential character of the motion can be obtained from a simple energy argument.

Let us consider a small expanding spherical region in the Universe. In a spherical distribution of matter, the gravitational force on a given spherical shell depends only on the mass inside that shell. We shall here assume $\Lambda = 0$.

We can now consider the motion of a galaxy of mass m at the edge of our spherical region. According to Hubble's law, its velocity will be $V = Hr$ and the corresponding kinetic energy,

$$T = mV^2/2. \qquad (20.9)$$

The potential energy at the edge of a sphere of mass M is $U = -GMm/r$. Thus the total energy is

$$E = T + U = mV^2/2 - GMm/r, \qquad (20.10)$$

which has to be constant. If the mean density of the Universe is ρ, the mass is $M = (4\pi r^3/3)\rho$. The value of ρ corresponding to $E = 0$ is called the *critical density*, ρ_c. We have

$$
\begin{aligned}
E &= \frac{1}{2}mH^2r^2 - \frac{GMm}{r} \\
&= \frac{1}{2}mH^2r^2 - Gm\frac{4\pi}{3}\frac{r^3\rho_c}{r} \\
&= mr^2\left(\frac{1}{2}H^2 - \frac{4}{3}\pi G\rho_c\right) = 0, \qquad (20.11)
\end{aligned}
$$

whence

$$\rho_{\rm c} = \frac{3H^2}{8\pi G}. \qquad (20.12)$$

The expansion of the Universe can be compared to the motion of a mass launched vertically from the surface of a celestial body. The form of the orbit depends on the initial energy. In order to compute the complete orbit, the mass M of the main body and the initial velocity have to be known. In cosmology, the corresponding parameters are the mean density and the Hubble constant.

The $E = 0$ model corresponds to the Euclidean Friedmann model, the *Einstein–de Sitter* model. If the density exceeds the critical density, the expansion of any spherical region will turn to a contraction and it will collapse to a point. This corresponds to the closed Friedmann model. Finally, if $\rho < \rho_{\rm c}$, the ever expanding hyperbolic model is obtained. The behaviour of the scale factor in these three cases is shown in Fig. 20.11.

These three models of the universe are called the *standard models*. They are the simplest relativistic cosmological models for $\Lambda = 0$. Models with $\Lambda \neq 0$ are mathematically more complicated, but show the same general behaviour.

The simple Newtonian treatment of the expansion problem is possible because Newtonian mechanics is approximately valid in small regions of the Universe. However, although the resulting equations are formally similar, the interpretation
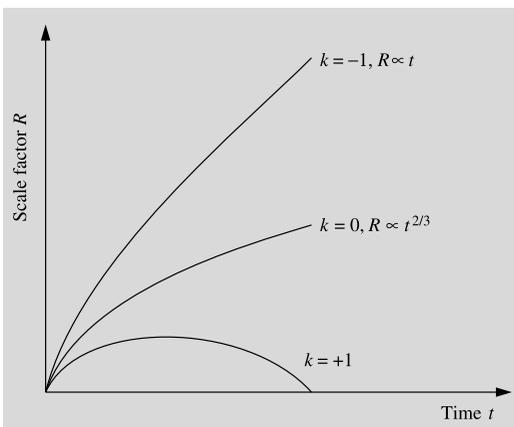

Fig. 20.11 The time dependence of the scale factor for different values of k. The cosmological constant $\Lambda = 0$

of the quantities involved (e.g. the parameter k) is not the same as in the relativistic context. The global geometry of the Friedmann models can only be understood within the general theory of relativity.

Consider two points at a separation r at the time t. Let their relative velocity be V. If the distance at t_0 is r_0 then

$$r = \frac{R(t)}{R(t_0)}\, r_0 \quad \text{and} \quad V = \dot r = \frac{\dot R(t)}{R(t_0)}\, r_0, \qquad (20.13)$$

and thus the Hubble constant is

$$H = \frac{V}{r} = \frac{\dot R(t)}{R(t)}. \qquad (20.14)$$

The deceleration of the expansion is described by the *deceleration parameter q*, defined as

$$q = -R\ddot R / \dot R^2. \qquad (20.15)$$

The deceleration parameter describes the change of the rate of the expansion $\dot R$. The additional factors have been included in order to make it dimensionless, i.e. independent of the choice of units of length and time.

The value of the deceleration parameter can be expressed in terms of the Hubble constant and density. Substitute the definition of H and the expression of $\ddot R$ (from Box 20.1) to the definition of q. We get

$$q = \frac{4\pi G}{3} \frac{\rho_0 R_0^3}{R^3 H^2}. \qquad (20.16)$$

The density is usually expressed in terms of the *density parameter Ω* defined as $\Omega = \rho/\rho_{\rm c}$, so that $\Omega = 1$ corresponds to the Einstein–de Sitter model. Thus

$$\Omega = \frac{8\pi G}{3} \frac{\rho_0 R_0^3}{R^3 H^2}. \qquad (20.17)$$

Hence there is a simple relation between Ω and q:

$$\Omega = 2q. \qquad (20.18)$$

The value $q = 1/2$ of the deceleration parameter corresponds to the critical density $\Omega = 1$. Both quantities are in common use in cosmology. It

should be noted that the density and the deceleration can be observed independently. The validity of (20.18) is thus a test for the correctness of general relativity with $\Lambda = 0$.

20.5 Cosmological Tests

A central cosmological problem is the question of which Friedmann model best represents the real Universe. Different models make different observational predictions. Recently there has been considerable progress in the determination of the cosmological parameters, and for the first time there is now a set of parameters that appear capable of accounting for all observations. In the following, some possible tests will be considered. These tests are related to the average properties of the Universe. Further cosmological constraints can be obtained from the observed structures. These will be discussed in Sect. 20.7.

The Critical Density   If the average density ρ is larger than the critical density ρ_c, the Universe is closed. For the Hubble constant $H = 100$ km s^{-1} Mpc^{-1}, the value of $\rho_c = 1.9 \times 10^{-26}$ kg m^{-3}, corresponding to roughly ten hydrogen atoms per cubic metre. Mass determinations for individual galaxies lead to smaller values for the density, favouring an open model. However, the density determined in this way is a lower limit, since there may be significant amounts of invisible mass outside the main parts of the galaxies.

If most of the mass of clusters of galaxies is dark and invisible, it will increase the mean density nearer the critical value. Using the virial masses of X-ray clusters of galaxies (Sect. 19.2), one finds $\Omega_0 = 0.3$. Considerations of the observed velocities of clusters of galaxies indicate that the relative amount of dark matter does not increase further on even larger scales.

Neutrinos have a small mass (about 10^{-4} electron mass). A large neutrino background should have been produced in the big bang. In spite of the small suggested mass of the neutrinos, they can still form a part of the mass of the Universe. Nowadays it is known that the mass of the neutrino is too small to explain the dark matter problem. Also, they move at relativistic speeds and are

too "hot" to form the observed structures. Instead much recent work in cosmology has been based on the hypothesis of *Cold Dark Matter* (CDM), i.e. the idea that a significant part of the mass of the Universe is in the form of non-relativistic particles of an unknown kind.

It is understood that the mean density of the University cannot be reliably determined from the observations of the current constituents of the Universe. Instead, a very precise value for the mean density is found by studying the details of the cosmic background radiation, since that gives a single picture of the contents of the whole Universe. The measurements show that the density is exactly the critical density, which means that the space is flat. The most recent result of Planck obtained in 2015 gives a curvature of 0.001 ± 0.004, meaning that the deviation from zero is at most 0.5 %.

The Magnitude-Redshift Test   Although for small redshifts, all models predict the Hubble relationship $m = 5 \lg z + C$ for standard candles, for larger redshifts there are differences, depending on the deceleration parameter q. This provides a way of measuring q.

The models predict that galaxies at a given redshift look brighter in the closed models than in the open ones (see Fig. 20.4). Measurements of type Ia supernovae out to redshifts $z = 1.7$ using the Hubble Space Telescope have now shown that the observed q is inconsistent with models having $\Lambda = 0$. Assuming $\Omega_0 = 0.3$ these observations require $\Omega_\Lambda = 0.7$, where Ω_Λ is defined below in Box 20.1.

The Angular Diameter-Redshift Test   Along with the magnitude-redshift test, the relation between angular diameter and redshift has been used as a cosmological test. Let us first consider how the angular diameter θ of a standard object varies with distance in static models with different geometries. In a Euclidean geometry, the angular diameter is inversely proportional to the distance. In an elliptical geometry, θ decreases more slowly with distance, and even begins to increase beyond a certain point. The reason for this can be understood by thinking of the surface of a sphere.

For an observer at the pole, the angular diameter is the angle between two meridians marking the edges of his standard object. This angle is smallest when the object is at the equator, and grows towards the opposite pole. In a hyperbolic geometry, the angle θ decreases more rapidly with distance than in the Euclidean case.

In an expanding closed universe the angular diameter should begin to increase at a redshift of about one. This effect has been looked for in the diameters of radio galaxies and quasars. No turnover has been observed, but this may also be due to evolution of the radio sources or to the selection of the observational data. At smaller redshifts, the use of the diameters of clusters of galaxies has yielded equally inconclusive results.

Evolutionary effects are the most important factors for the uncertainty. We know that the sizes of large scale structures are changing while the Universe is evolving. The evolution of the structures has been investigated using cosmological simulations (e.g. Millennium, Illustris Fig. 20.14). The simulations give estimates for the object sizes at different redshift, and these values can be compared with the actual observations. The results are consistent with the cosmological ΛCDM concordance model.

Basically the same idea can be applied to the angular scale of the strongest fluctuation in the cosmic microwave background. The linear size of these depends only weakly on the cosmological model and can therefore be treated as a standard measuring rod. Their redshift is determined by the decoupling of matter and radiation (see Sect. 20.6). Observations of their angular size have provided strong evidence that $\Omega_0 + \Omega_\Lambda = 1$, i.e. the Universe is flat.

Primordial Nucleosynthesis The standard model predicts that 25 % of the mass of the Universe turned into helium in the "big bang". This amount is not sensitive to the density and thus does not provide a strong cosmological test. However, the amount of deuterium left over from helium synthesis does depend strongly on the density. Almost all deuterons formed in the big bang unite into helium nuclei. For a larger density the collisions destroying deuterium were more frequent.

Thus a small present deuterium abundance indicates a high cosmological density. Similar arguments apply to the amounts of ^{3}He and ^{7}Li produced in the big bang. The interpretation of the observed abundances is difficult, since they have been changed by later nuclear processes. Still, present results for the abundances of these nuclei are consistent with each other and with a density corresponding to Ω_0 about 0.04. Note that this number only refers to the mass in the form of baryons, i.e. protons and neutrons. Since the virial masses of clusters of galaxies indicate that Ω_0 is about 0.3, this has stimulated models such as the CDM model, where most of the mass is not in the form of normal baryons.

Ages The ages of different Friedmann models can be compared with known ages of various systems. The age t_0 of a Friedmann model with given Ω_0 and Ω_Λ is obtained by integrating equation (8) of Box 20.1. This gives

$$t_0 = H_0^{-1} \int_0^1 \mathrm{d}a \big(\Omega_0 a^{-1} + \Omega_\Lambda a^2 + 1$$
$$- \Omega_0 - \Omega_\Lambda \big)^{-1/2}. \quad (20.19)$$

This age is required to be larger than the ages of the oldest known astronomical objects.

If the density is critical and $\Lambda = 0$, $t_0 H_0 = 2/3$. Thus if $H_0 = 75 \ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, the age is $t_0 = 9$ Ga. Larger values of Ω_0 give smaller ages, whereas positive values of Λ lead to larger ages. It has been a source of embarrassment that the best values of H have tended to give an age for the Universe only marginally consistent with the ages of the oldest astronomical objects. With the introduction of a positive cosmological constant and a slight downward revision of stellar ages this problem has disappeared. In 2003 the WMAP results gave an age of 13.7 Ga. A more recent value by the Planck satellite is 13.80 ± 0.02 Ga.

The best current parameter values give 13–14 Ga for the age of the Universe.

The "Concordance" Model In summary there has been a remarkable recent convergence between different cosmological tests. The resulting model has a positive cosmological constant, and

most of the matter is cold and dark. It is thus referred to as the ΛCDM model. The best parameter values are $H = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.7$, $\Omega_0 = 0.3$, with cold dark matter making up 85 % of the total density.

The concordance model is by no means definitive. In particular the reason for the cosmological constant is a major puzzle. In order to allow for the possibility that Λ is variable it has become customary to refer to it as *dark energy*, saving the term cosmological constant for the case of a constant Λ. Even if some alternative mechanism can produce the same effect as a non-zero Λ, finding at least one set of acceptable parameters, thanks to the WMAP and Planck satellites, is an important step forward.

An additional set of constraints on the cosmological model comes from the large-scale structures of the Universe. These constraints, which contribute further support for the concordance model, will be considered in Sect. 20.7.

20.6 History of the Universe

We have seen how the density of matter and of radiation energy and temperature can be computed as functions of the scale factor R. Since the scale factor is known as a function of time, it is possible to calculate how these quantities change with time.

When we back in time R is decreasing. Since $\rho \propto R^{-3}$, $T \propto R^{-1}$, densities and temperatures at the beginning were so immense that all theories about the physical processes taking place are highly conjectural. Nevertheless attempts have been made at understanding the most fundamental properties of the Universe on the basis of modern theories of particle physics. For example, no indications of significant amounts of antimatter in the Universe have been discovered. Thus, for some reason, the number of matter particles must have exceeded that of antimatter particles by a factor of 1.000000001. Because of this *symmetry breaking*, when 99.9999999 % of the hadrons were annihilated, 10^{-7} % was left later to form galaxies and everything else. It has been speculated that the broken symmetry originated in particle processes about 10^{-35} s after the initial time.
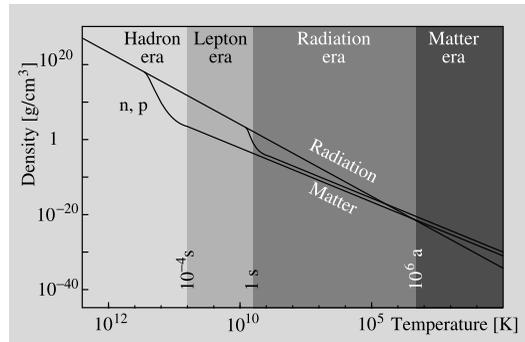

Fig. 20.12 The energy densities of matter and radiation decrease as the Universe expands. Nucleon–antinucleon pairs annihilate at 10^{-4} s; electron–positron pairs at 1 s

The breaking of fundamental symmetries in the early Universe may lead to what is known as *inflation* of the Universe. In consequence of symmetry breaking, the dominant energy density may be the zero-point energy of a quantum field. This energy density will lead to inflation, a strongly accelerated expansion, which will dilute irregularities and drive the density very close to the critical value. One may thus understand how the present homogeneity, isotropy and flatness of the Universe have come about. In the inflationary picture the Universe has to be very nearly flat, $\Omega_0 + \Omega_\Lambda = 1$. The inflationary models also make specific predictions for the form of the irregularities in the CMB. These predictions are in general agreement with what has been observed.

As the Universe expanded, the density and temperature decreased (Fig. 20.12) and conditions became such that known physical principles can be applied. During the hot early stages, photons and massive particles were continually changing into each other: high-energy photons collided to produce particle-antiparticle pairs, which then were annihilated and produced photons. As the Universe cooled, the photon energies became too small to maintain this equilibrium. There is a *threshold temperature* below which particles of a given type are no longer produced. For example, the threshold temperature for *hadrons* (protons, neutrons and mesons) is $T = 10^{12}$ K, reached at the time $t = 10^{-4}$ s. Thus the present building blocks of the atomic nuclei, protons and neutrons, are relics from the time 10^{-8}–10^{-4} s, known as the *hadron era*.

Even before the hadron era, when the age of the Universe was about 10^{-12} s, the dark matter particles forming about 5/6 of all matter were born. They were decoupled from radiation around 10^{-9} s and started to form structures to which the ordinary matter started to condense hundred million years later.

The Lepton Era   In the time period 10^{-4}–1 s, the *lepton era*, the photon energies were large enough to produce light particles, such as electron–positron pairs. Because of matter-antimatter symmetry breaking, some of the electrons were left over to produce present astronomical bodies. During the lepton era *neutrino decoupling* took place. Previously the neutrinos had been kept in equilibrium with other particles by fast particle reactions. As the density and temperature decreased, so did the reaction rates, and finally they could no longer keep the neutrinos in equilibrium. The neutrinos decoupled from other matter and were left to propagate through space without any appreciable interactions. It has been calculated that there are at present 600 such cosmological neutrinos per cubic centimetre, but their negligible interactions make them extremely difficult to observe. After the neutrino decoupling the amount of free neutrons kept decreasing as they decayed to protons and electrons.

The Radiation Era   After the end of the lepton era, about 1 s after the initial time, the most important form of energy was electromagnetic radiation. This stage is called the *radiation era*. At its beginning the temperature was about 10^{10} K and at its end, about one million years later, when the radiation energy density had dropped to that of the particles, it had fallen to about 40,000 degrees. At the very beginning of the radiation era within a few hundred seconds helium was produced.

Just before the epoch of helium synthesis, the number ratio of free protons and neutrons was changing because of the decay of the free neutrons. After about 100 s the temperature had dropped to about 10^9 K, which is low enough for deuterons to be formed. All remaining neutrons were then incorporated in deuterons; these,

in turn, were almost entirely consumed to produce helium nuclei. Thus the amount of helium synthesised was determined by the number ratio of protons and neutrons at the time of deuterium production $t = 100$ s. Calculations show that this ratio was about 14 : 2. Thus, out of 16 nucleons, 2 protons and 2 neutrons were incorporated in a helium nucleus. Consequently $4/16 = 25$ % of the mass turned into helium. This is remarkably close to the measured primordial helium abundance.

Only the isotopes ^{2}H, ^{3}He, ^{4}He and ^{7}Li were produced in appreciable numbers by nuclear processes in the big bang. The heavier elements have formed later in stellar interiors, in supernova explosions and perhaps in energetic events in galactic nuclei.

Radiation Decoupling. The Matter Era   As we have seen, the mass density of radiation (obtained from the formula $E = mc^2$) behaves as R^{-4}, whereas that of ordinary matter behaves as R^{-3}. Thus the radiation mass density decreases more rapidly. At the end of the radiation era it became smaller than the ordinary mass density. The *matter era* began, bringing with it the formation of galaxies, stars, planets and human life. At present, the mass density of radiation is much smaller than that of matter. Therefore the dynamics of the Universe is completely determined by the density of massive particles.

Soon after the end of the radiation era, radiation decoupled from matter. This happened when the temperature had dropped to a few thousand degrees, and the protons and electrons combined to form hydrogen atoms. It was the beginning of the *"dark ages"* at redshifts $z = 1000$–100, before stars and galaxies had formed, when the Universe only contained dark matter, blackbody radiation, and slowly cooling neutral gas.

At present, light can propagate freely through space. The world is transparent to radiation: the light from distant galaxies is weakened only by the r^{-2} law and by the redshift. Since there is no certain detection of absorption by neutral gas, there must have been a *reionisation* of the Universe. It is thought that this occurred around $z = 5$–10.

20.7   The Formation of Structure

As we go backward in time from the present, the distances between galaxies and clusters of galaxies become smaller. For example, the typical separation between galaxies is 100 times their diameter. At the redshift $z = 101$ ($R(\text{now})/R(t) = 100$) most galaxies must have been practically in contact. For this reason galaxies in their present form cannot have existed much earlier than about $z = 100$. Since the stars were presumably formed after the galaxies, all present astronomical systems must have formed later than this.

It is thought that all observed structures in the Universe have arisen by the gravitational collapse of small overdensities. Whereas the presently observable galaxies have undergone considerable evolution, which makes it difficult to deduce their initial state, on larger scales the density variations should still be small and easier to study. These are the structures considered in the present section. The later evolution of galaxies is discussed in Sect. 19.8, and the formation of the Milky Way in Sect. 18.5.

The Statistical Description of Large-Scale Structure   The departures from strict homogeneity in the Universe are random in character, and must therefore be described using statistical methods. Perhaps the most straightforward way of doing this is to take regions of a given size, specified in terms of their mass, and give the probability distribution for relative density variations on that scale.

A second method is to consider the spatial separations between individual objects such as galaxies or clusters. The distribution of these separations is used to define the *correlation function*, which is a measure of the clustering of the objects in question.

A third method to describe large-scale fluctuations is by means of the *power spectrum*. Here the density variations (in space or in projection on the sky) are represented as sum of waves. The power spectrum is the squared amplitude of these waves as a function of wavelength.

All three methods are representations of the density variations in the Universe, and they are
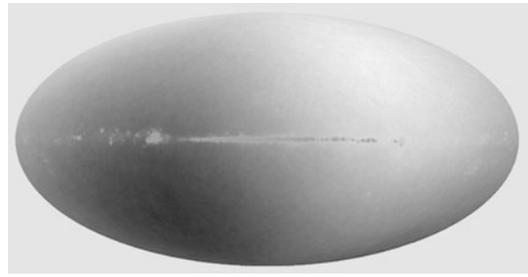

Fig. 20.13 The motion of the Milky Way in relation to the microwave background can be seen in the measurements of WMAP. One side of the sky is darker (colder), and the other side is lighter (warmer). *The horizontal stripe* is the densest part of the Milky Way. (Photo NASA)

theoretically closely related. However, in practice they are observed in different ways, and therefore, which representation is most suitable depends on what kind of observations are being analysed. The density variations are usually described by means of a spectral index n and an amplitude σ_8 to be introduced below.

The Growth of Perturbations   In order to describe the growth of structures in the Universe, consider a given region containing the mass M. If its density is slightly larger than the mean density, its expansion will be slightly slower than that of the rest of the Universe, and its relative overdensity will grow. The rate of growth as a function of mass depends on the relative importance of the material components of the Universe, dark matter, radiation, and ordinary baryonic matter.

It is assumed that there is an initial distribution of perturbations where the fluctuations with mass M have an amplitude that is proportional to $M^{-(n+3)/6}$. The spectral index n is a cosmological parameter to be determined from observations.

The first step in structure formation is when a given mass comes within the *horizon*, i.e. when there has been enough time since the big bang for light signals to cross the given region. During the radiation era the horizon mass grows proportionally to $t^{3/2}$, and thus the time at which the mass M comes within the horizon will be proportional to $M^{2/3}$. Any perturbation will initially be larger than the horizon mass, and while this is the case it grows in proportion to t. Once a matter

perturbation comes inside the horizon its amplitude will remain constant. This constant amplitude will behave as $t M^{-(n+3)/6}$, which is proportional to $M^{-(n-1)/6}$. If $n = 1$, the perturbations enter the horizon with an amplitude that is independent of mass. We shall see that the observed value of n is in fact very close to 1.

Perturbations of both dark and baryonic matter density will behave as described above during the radiation era. At the end of the radiation era at time t_{EQ}, when the mass densities of radiation and (non-relativistic) matter become equal, the amplitude of the perturbations will be given by the horizon mass $M_{EQ} \approx 10^{16} M_{\odot}$ for $M \ll M_{EQ}$, and will be proportional to $M^{-(n+3)/6}$ for $M \gg M_{EQ}$. After equality the dark matter perturbations will be free to start growing again as $t^{2/3}$, independent of mass.

Unlike dark matter, ordinary baryonic matter perturbations cannot grow as long as the Universe remains ionised. Instead there is a minimum mass of a collapsing gas cloud given by the *Jeans mass* M_J:

$$M_J \approx \frac{P^{3/2}}{G^{3/2} \rho^2}, \qquad (20.20)$$

where ρ and P are the density and pressure in the cloud (see Sect. 6.11). The value of M_J before decoupling was

$$M_J = 10^{18} M_{\odot} \qquad (20.21)$$

and after decoupling

$$M_J = 10^5 M_{\odot}. \qquad (20.22)$$

The reason for the large difference is that before decoupling, matter feels the large radiation pressure ($P = u/3$, see Box 11.1). After decoupling, this pressure no longer affects the gas.

The large Jeans mass before decoupling means that overdense regions of normal gas cannot start growing before $z = 1000$. Rather than growing they oscillate like sound waves. After decoupling a large range of masses become Jeans unstable. By then density perturbations of dark matter have had time to grow, and the gas will therefore rapidly fall into their potential wells. The first stars will start forming in the collapsing regions, and will reionise the Universe.

Fig. 20.14 Cosmological simulations are used to study the evolution of the structures of the Universe. The international Illustris simulation shows how the network of dark matter evolves from redshift $z = 4$ (*bottom frame*) to $z = 0$ (*topmost frame*). (Illustris Collaboration, www. illustris-project.org)
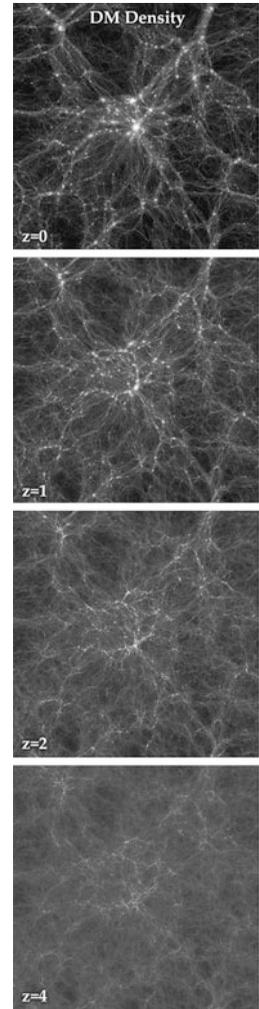
Because the expansion of the Universe works against the collapse, the density of Jeans unstable regions grows rather slowly. In order to produce the observed systems, the density perturbations at decoupling cannot be too small. In models without dark matter the variations in the CMB predicted on this basis tended to be too large. In the CDM model the predicted variations are of the expected amplitude.

In the CDM model the amplitude of fluctuations on scales about M_{EQ} and smaller depends only weakly on mass. This is why the CDM model leads to an hierarchical description of structure formation. In this picture, systems of all masses above $10^5 M_{\odot}$ begin forming after decoupling. Because smaller systems will collapse

more rapidly, they are the first to form, at redshifts about 20. Once the first sources of light, starbursts or AGNs, had formed, they could reionise the gas. This marked the ending of the dark ages at redshifts $z = 10$–5.

The redshift of reionisation is still not well known, and is therefore treated as a parameter to be determined in tests based on large-scale structures. It is usually expressed by means of τ, the optical depth to electron scattering of the background radiation. A larger value of τ corresponds to a higher electron density, implying reionisation at a higher redshift. Since galaxies can be seen at redshifts larger than 6, the corresponding value of $\tau = 0.03$ represents a minimum. The final results (2012) of the observations with WMAP gave $\tau = 0.081 \pm 0.012$, which would have corresponded to a redshift 10.1 ± 1.0. The final results (2015) of Planck the gave 8.8 ± 0.1 as the mean redshift of reionisation.

One finally has to ask where the initial perturbations came from. An attractive feature of the inflationary model is that it makes specific predictions for these initial perturbations, deriving them from quantum effects at very early times. In this way the observed properties of the largest astronomical systems contain information about the earliest stages of our Universe.

Fluctuations of the Cosmic Microwave Background One important way of studying the large-scale structure of the early Universe is by means of the irregularities of the cosmic microwave background. The overdensities that were later to give rise to observed structures should also give rise to temperature variations of the CMB.

The temperature variations in the microwave background have been mapped, first by the COBE satellite, then by WMAP (Fig. 20.13) and Planck. A map of the CMB according to Planck is shown in Fig. 20.15. The observed variations are in qualitative agreement with the scenario for structure formation described above.

A more quantitative view of the observations is provided by the power spectrum of the observations shown in Fig. 20.16. This shows the amplitude of the temperature variations as a function of angular scale on the sky.
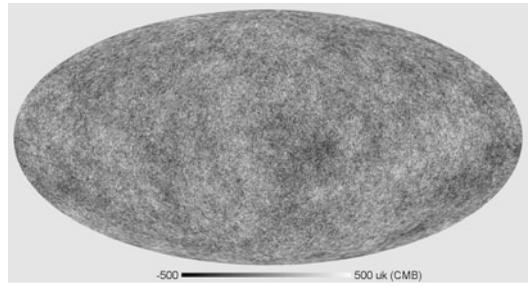

Fig. 20.15 Temperature distribution of the background radiation according to the measurements of the Planck satellite. *Dark areas* are colder, *light areas* warmer and more dense. (Planck Collaboration)

The physical processes we have described give rise to the features in the power spectrum. Thus the first peak is produced by perturbations that have just had time to collapse to a maximum density since equality of matter and radiation, before bouncing back. The linear size of this peak does not depend strongly on the exact parameters of the model, and its position (angular scale) can therefore be used as a standard measuring rod in the diameter—redshift test. The second and third peaks in the power spectrum, the *"acoustic peaks"*, correspond to perturbations that have bounced back and collapsed again, respectively. The positions and amplitudes of these features in the power spectrum depend on the cosmological parameters. Thus the WMAP observations of the CMB can be fitted using a model containing six free parameters: the Hubble constant, the densities of dark matter and baryons, the optical depth τ, and the amplitude σ_8 and shape n of the initial perturbations. The curvature of the Universe, $1 - \Omega_0 - \Omega_\Lambda$ can be set equal to zero, which determines the value of Ω_Λ. The results of WMAP and Planck have given values for these parameters in full agreement with the concordance model.

Very Large Scale Structure After decoupling structures on different scales are free to grow. The important dividing line is whether a given overdensity is still expanding with the rest of the Universe or whether it has had time to recollapse. Systems that have recollapsed will fairly rapidly

Fig. 20.16 Angular power spectrum of the temperature variations of the cosmic microwave background from the results of the Planck satellite. This gives the square of the amplitude of the temperature variations on different angular scales. (Planck Collaboration)
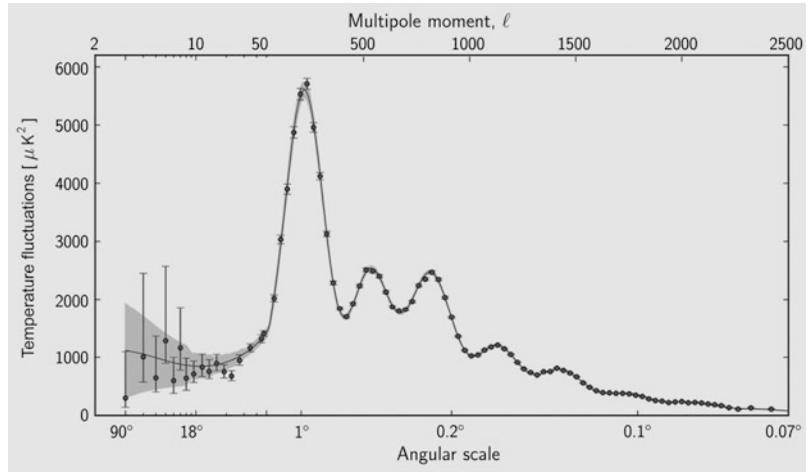
virialise, i.e. settle down into a stationary state. Such systems can be considered real astronomical systems.

The largest astronomical systems are clusters of galaxies. For some time there was controversy about whether even larger structures, superclusters, existed. The controversy was settled in the 1970's when it was realised that this largely depended on what was meant by superclusters. Regions of higher density existed on scales larger than clusters of galaxies (a few Mpc), all the way up to 100 Mpc, but they did not really form individual structures in approximate equilibrium like the clusters of galaxies.

Between these two alternatives are structures that are only now turning round and beginning to recollapse. One way of specifying the amplitude of the initial fluctuations is to give the scale of this transition point. This scale corresponds roughly to a present linear size of 8 Mpc, and for this reason the fluctuation amplitude is commonly given by means of σ_8, the amplitude of the fluctuations at present scale 8 Mpc. Because of the way it has been chosen σ_8 should be close to 1.

Smaller systems will already have collapsed. There are many ways of statistically describing their large-scale distribution. One of the most studied descriptors of clustering is the *correlation function*, $\xi(r)$. Consider two infinitesimal volumes dV_1 and dV_2 separated by the distance r. If there were no clustering, the probability of finding galaxies in these volumes would

be $N^2 \, dV_1 \, dV_2$, where N is the number density of galaxies. Because of clustering this probability is actually $N^2(1 + \xi(r)) \, dV_1 \, dV_2$. The correlation function thus measures the higher probability of finding galaxies near each other.

Although there are ways of estimating the correlation function from the distribution on the sky, a more reliable estimate can be made by mapping the distribution of galaxies in three dimensions, using redshifts to determine their distances. The distribution from one such survey (SDSS, Sloan Digital Sky Survey) is shown in Fig. 20.17.

As expected, the general form of the correlation function is a power law (close pairs are abundant, distant pairs rarer). It has been known since the 1970's that the correlation function is approximately proportional to r^γ. The constant of proportionality is related to the amplitude σ_8 and the exponent γ is related to the index n of the initial perturbations. The observed value of γ is about 1.8, depending to some extent on the sample of objects studied.

At large separations the curve shows an unexpected maximum, observed for the first time in 2005. This is the same highest acoustic peak seen in the angular power spectrum of Planck. The observation confirms the ΛCDM cosmology and shows that the acoustic waves born around $z = 1000$ have remained till the present time and are seen in the distribution of galaxies. The same peak has been seen later in the larger SDSS survey.
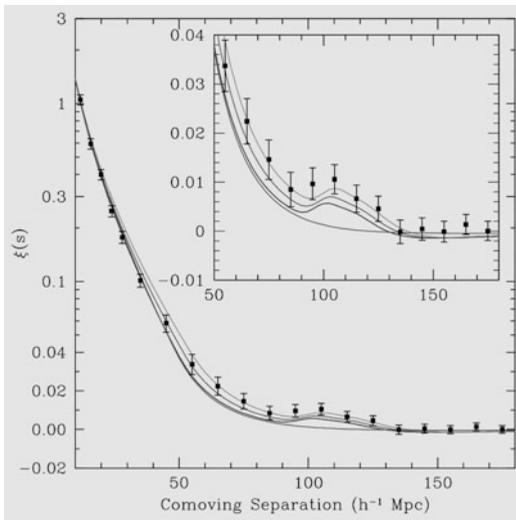
Fig. 20.17 Correlation function for the distribution of 47,000 bright red SDSS galaxies. *The inset* shows the found "bump" in more detail. The value of h is the Hubble constant divided by 100, or about 0.7. Thus a difference $100h^{-1}$ corresponds to a distance of about 150 Mpc (about 6 degrees on the sky). The bump tells about acoustic oscillations of baryons, i.e. large scale acoustic waves of the visible matter. (Eisenstein et al., 2005, ApJ 633, 560)

In order to determine the cosmological parameters the (three-dimensional) power spectrum of the galaxy distribution is usually used. It is then compared to theoretical power spectra that depend on the parameters in order to find the best-fitting model. The optical depth parameter τ does not affect these tests. The results have been completely consistent with the concordance model.

The Concordance Model Again   We have discussed the use of the cosmic microwave background and the large-scale distribution of galaxies in order to determine the cosmological parameters. Both methods independently give consistent results. Combining them produces even more (formally) accurate values. Furthermore, these values are in agreement with the ones obtained from the traditional cosmological tests discussed in Sect. 20.5.

There are eight basic parameters describing the concordance model. According the final results of Planck (2015) they are:

Hubble constant H_0 $6.7 \pm 0.5\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$,
baryon density parameter Ω_b 0.049 ± 0.001,

mass density parameter Ω_0 0.259 ± 0.006,
redshift of reionisation z_{ri} 8.8 ± 1.1,
spectral index n 0.967 ± 0.004,
fluctuation amplitude σ_8 0.82 ± 0.01.

Other, less comprehensive tests, such as the gravitational lensing by cosmic structures, the velocities induced by mass concentrations, and the number of clusters of galaxies, have added support to the concordance model. This remarkable agreement between many independent determinations of the cosmological parameters have earned the model its name. Once the parameters necessary for a general description of the Universe are reasonably well known, other components, such as gravitational waves or neutrinos, can be included in the cosmological model. This new development has been called the era of precision cosmology. However, it should still not be forgotten that in the model one has to assume the presence of both dark matter and dark energy, neither of which is based on any evidence apart from their role in cosmology.

20.8   The Future of the Universe

The standard models allow two alternative prospects for the future development of the Universe. Expansion may either go on forever, or it may reverse to a contraction, where everything is eventually squeezed back to a point. In the final squeeze, the early history of the Universe would be repeated backwards: in turn, galaxies, stars, atoms and nucleons would be broken up. Finally the state of the Universe could no longer be treated by present-day physics. It is not known whether the collapse would continue to a single point or would the Universe avoid the singularity and start a new expansion.

However, according to the current observations, this is not the fate of the Universe. Instead, the expansion will continue forever, even at an accelerating pace.

In the open models the future is quite different. The evolution of the stars may lead to one of four end results: a white dwarf, a neutron star or a black hole may be formed, or the star may be completely disrupted. After about 10^{11} years, all

present stars will have used up their nuclear fuel and reached one of these four final states.

Some of the stars will be ejected from their galaxies; others will form a dense cluster at the centre. In about 10^{27} years the central star clusters will become so dense that a black hole is formed. Similarly the galaxies in large clusters will collide and form very massive black holes.

Not even black holes last forever. By a quantum mechanical tunnelling process, mass can cross the event horizon and escape to infinity—the black hole is said to "evaporate". The rate of this phenomenon, known as the *Hawking process*, is inversely proportional to the mass of the hole. For a galactic-mass black hole, the evaporation time is roughly 10^{98} years. After this time, almost all black holes will have disappeared.

The ever expanding space now contains black dwarfs, neutron stars and planet-size bodies (unless the predictions of a finite proton lifetime of about 10^{31} years are confirmed; in that case all these systems will have been destroyed by proton decay). The temperature of the cosmic background radiation will have dropped to 10^{-20} K.

Even further in the future, other quantum phenomena come into play. By a tunnelling process, black dwarfs can change into neutron stars and these, in turn, into black holes. In this way, all stars are turned into black holes, which will then evaporate. The time required has been estimated to be $10^{10^{26}}$ years! At the end, only radiation cooling towards absolute zero will remain (Fig. 20.18).
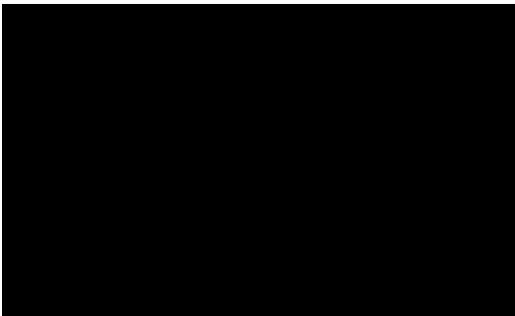

Fig. 20.18 According to current results the future of the Universe looks like this. After a very, very long time all matter will have changed to radiation. The wavelength of the radiation will increase infinitely in an ever expanding and cooling space

It is of course highly doubtful whether our current cosmological theories really are secure enough to allow such far-reaching predictions. New theories and observations may completely change our present cosmological ideas.

Box 20.1 (Newtonian Derivation of a Differential Equation for the Scale Factor $R(t)$) Let us consider a galaxy at the edge of a massive sphere (see figure). It will be affected by a central force due to gravity and the cosmological force

$$m\ddot{r} = -\frac{4\pi G}{3}\frac{r^3 \rho m}{r^2} + \frac{1}{3}m\Lambda r,$$

or

$$\ddot{r} = -\frac{4\pi}{3}G\rho r + \frac{1}{3}\Lambda r. \qquad (1)$$
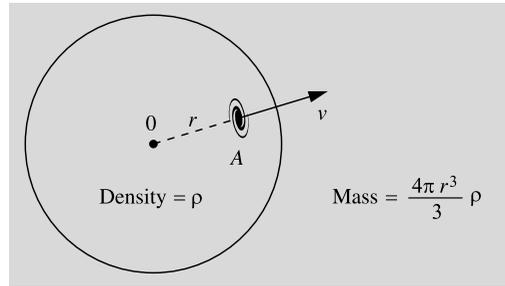
In these equations, the radius r and the density ρ are changing with time. They may be expressed in terms of the scale factor R:

$$r = (R/R_0)r_0, \qquad (2)$$

where R is defined to be R_0 when the radius $r = r_0$.

$$\rho = (R_0/R)^3 \rho_0, \qquad (3)$$

where the density $\rho = \rho_0$ when $R = R_0$. Introducing (2) and (3) in (1), one obtains

$$\ddot{R} = -\frac{a}{R^2} + \frac{1}{3}\Lambda R, \qquad (4)$$

where $a = 4\pi G R_0^3 \rho_0/3$. If (4) is multiplied on both sides by $\dot{R}$, the left-hand side yields

$$\dot{R}\ddot{R} = \frac{1}{2}\frac{d(\dot{R}^2)}{dt},$$

and thus (4) takes the form

$$d(\dot{R}^2) = -\frac{2a}{R^2}\,dR + \frac{2}{3}\Lambda R\,dR. \qquad (5)$$

Let us define $R_0 = R(t_0)$. Integrating (5) from t_0 to t gives

$$\dot{R}^2 - \dot{R}_0^2 = 2a\left(\frac{1}{R} - \frac{1}{R_0}\right)$$
$$+ \frac{1}{3}\Lambda\left(R^2 - R_0^2\right). \qquad (6)$$

The constants $\dot{R}_0$ and a can be eliminated in favour of the Hubble constant H_0 and the density parameter Ω_0. Because $\rho_c = 3H_0^2/8\pi G$,

$$2a = 8\pi G R_0^3 \rho_0/3$$
$$= H_0^2 R_0^3 \rho_0/\rho_c = H_0^2 R_0^3 \Omega_0, \qquad (7)$$

where $\Omega_0 = \rho_0/\rho_c$. Using expression (6) and $\dot{R}_0 = H_0 R_0$ in (7), and defining $\Omega_\Lambda = \Lambda/(3H_0^2)$, one obtains

$$\frac{\dot{R}^2}{H_0^2 R_0^2} = \Omega_0\frac{R_0}{R} + \Omega_\Lambda\left(\frac{R}{R_0}\right)^2$$
$$+ 1 - \Omega_0 - \Omega_\Lambda \qquad (8)$$

as the basic differential equation governing $R(t)$.

For simplicity we now set $\Omega_\Lambda = 0$. Then the time behaviour of the scale factor R depends on the value of the density parameter Ω_0. Because $\dot{R}^2 > 0$ always, according to (8)

$$\Omega_0\frac{R_0}{R} - \Omega_0 + 1 \geq 0,$$

or

$$\frac{R_0}{R} \geq \frac{\Omega_0 - 1}{\Omega_0}. \qquad (9)$$

If $\Omega_0 > 1$, this means that

$$R \leq R_0\frac{\Omega_0}{\Omega_0 - 1} \equiv R_{\max}.$$

When the scale factor reaches its maximum value $R_{\max}$, then according to (8), $\dot{R} = 0$, and the expansion turns into contraction. If $\Omega_0 < 1$,

the right-hand side of (8) is always positive and the expansion continues forever.

The equation for the time dependence of the scale factor in the general theory of relativity contains the constant k which determines the geometry of space:

$$\dot{R}^2 = \frac{8\pi G R_0^3 \rho_0}{3R} - kc^2. \qquad (10)$$

Equations (10) and (6) (or (8)) can be made identical if one chooses

$$H_0^2 R_0^2(\Omega_0 - 1) = kc^2.$$

Thus, complete agreement between the Newtonian and the relativistic equation for R is obtained. The values of the geometrical constant $k = +1, 0, -1$ correspond respectively to $\Omega_0 > 1, = 1$ and < 1. More generally, the condition for a flat model, $k = 0$, corresponds to $\Omega_0 + \Omega_\Lambda = 1$.

When $k = 0$, the time dependence of the expansion is very simple. Setting $\Omega_0 = 1$ and using (10) and (7), one obtains

$$\dot{R}^2 = \frac{H_0^2 R_0^3}{R}.$$

The solution of this equation is

$$R = \left(\frac{3H_0 t}{2}\right)^{2/3} R_0. \qquad (11)$$

It is also easy to calculate the time from the beginning of the expansion: $R = R_0$ at the time

$$t_0 = \frac{2}{3}\frac{1}{H_0}.$$

This is the age of the Universe in the Einstein–de Sitter model.

Box 20.2 (Three Redshifts) The redshift of a distant galaxy is the result of three different mechanisms acting together. The first one is the peculiar velocity of the observer with respect to the mean expansion: the Earth moves about the Sun, the Sun about the centre of the Milky Way, and the Milky Way and the Local Group

of galaxies is falling towards the Virgo Cluster. The apparatus measuring the light from a distant galaxy is not at rest; the velocity of the instrument gives rise to a Doppler shift that has to be corrected for. Usually the velocities are much smaller than the speed of light. The Doppler shift is then

$$z_D = v/c. \qquad (1)$$

For large velocities the relativistic formula has to be used:

$$z_D = \sqrt{\frac{c+v}{c-v}} - 1. \qquad (2)$$

The redshift appearing in Hubble's law is the *cosmological redshift* z_c. It only depends on the values of the scale factor at the times of emission and detection of the radiation (R and R_0) according to

$$z_c = R_0/R - 1. \qquad (3)$$

The third type of redshift is the *gravitational redshift* z_g. According to general relativity, light will be redshifted by a gravitational field. For example, the redshift of radiation from the surface of a star of radius R and mass M will be

$$z_g = \frac{1}{\sqrt{1 - R_S/R}} - 1, \qquad (4)$$

where $R_S = 2GM/c^2$ is the Schwarzschild radius of the star. The gravitational redshift of the radiation from galaxies is normally insignificant.

The combined effect of the redshifts can be calculated as follows. If the rest wavelength λ_0 is redshifted by the amounts z_1 and z_2 by two different processes, so that

$$z = \frac{\lambda_2 - \lambda_0}{\lambda_0} = \frac{\lambda_2}{\lambda_0} - 1 = \frac{\lambda_2}{\lambda_1}\frac{\lambda_1}{\lambda_0} - 1,$$

or

$$(1 + z) = (1 + z_1)(1 + z_2).$$

Similarly, the three redshifts z_D, z_c and z_g will combine to give an observed redshift z, according to

$$1 + z = (1 + z_D)(1 + z_c)(1 + z_g). \qquad (5)$$

20.9 Examples

Example 20.1 (a) In a forest there are n trees per hectare, evenly spaced. The thickness of each trunk is D. What is the distance of the wood not seen for the trees? (Find the probability that the line of sight will hit a trunk within a distance x.) (b) How is this related to the Olbers paradox?

(a) Imagine a circle with radius x around the observer. A fraction $s(x)$, $0 \le s(x) \le 1$, is covered by trees. Then we'll move a distance dx outward, and draw another circle. There are $2\pi n x\, dx$ trees growing in the annulus limited by these two circles. They hide a distance $2\pi x n D\, dx$ or a fraction $n D\, dx$ of the perimeter of the circle. Since a fraction $s(x)$ was already hidden, the contribution is only $(1 - s(x))n D\, dx$. We get

$$s(x + dx) = s(x) + \big(1 - s(x)\big)n D\, dx,$$

which gives a differential equation for s:

$$\frac{ds(x)}{dx} = \big(1 - s(x)\big)n D.$$

This is a separable equation, which can be integrated:

$$\int_0^s \frac{ds}{1-s} = \int_0^x n D\, dx.$$

This yields the solution

$$s(x) = 1 - e^{-nDx}.$$

This is the probability that in a random direction we can see at most to a distance x. This function s is a cumulative probability distribution. The corresponding probability density is its derivative ds/dx. The mean free path λ is the expectation of this distribution:

$$\lambda = \int_0^\infty x \left(\frac{ds(x)}{dx} \right) dx = \frac{1}{nD}.$$

For example, if there are 2000 trees per hectare, and each trunk is 10 cm thick, we can see to a distance of 50 m, on the average.

(b) The result can easily be generalised into three dimensions. Assume there are n stars per unit volume, and each has a diameter D and surface $A = \pi D^2$ perpendicular to the line of sight. Then we have

$$s(x) = 1 - e^{-nAx},$$

where

$$\lambda = 1/nA.$$

For example, if there were one sun per cubic parsec, the mean free path would be 1.6×10^4 parsecs. If the universe were infinitely old and infinite in size, the line of sight would eventually meet a stellar surface in any direction, although we could see very far indeed.

Example 20.2   Find the photon density of the 2.7 K background radiation.

The intensity of the radiation is

$$B_\nu = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/(kT)} - 1}$$

and the energy density

$$u_\nu = \frac{4\pi}{c} B_\nu = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/(kT)} - 1}.$$

The number of photons per unit volume is found by dividing the energy density by the energy of a single photon, and integrating over all frequencies:

$$N = \int_0^\infty \frac{u_\nu \, d\nu}{h\nu} = \frac{8\pi}{c^3} \int_0^\infty \nu^2 \frac{d\nu}{e^{h\nu/(kT)} - 1}.$$

We substitute $h\nu/kT = x$ and $d\nu = (kT/h) \, dx$:

$$N = 8\pi \left(\frac{kT}{hc}\right)^3 \int_0^\infty \frac{x^2 \, dx}{e^x - 1}.$$

The integral cannot be expressed in terms of elementary functions (however, it can be expressed as an infinite sum $2\sum_{n=0}^\infty (1/n^3)$), but it can be evaluated numerically. Its value is 2.4041. Thus the photon density at 2.7 K is

N

$$= 16\pi \left(\frac{1.3805 \times 10^{-23} \, \text{J K}^{-1} \times 2.7 \, \text{K}}{6.6256 \times 10^{-34} \, \text{J s} \times 2.9979 \times 10^8 \, \text{m s}^{-1}}\right)^3$$

$$\times 1.20206$$

$$= 3.99 \times 10^8 \, \text{m}^{-3} \approx 400 \, \text{cm}^{-3}.$$

20.10 Exercises

Exercise 20.1 The apparent diameter of the galaxy NGC 3159 is 1.3′, apparent magnitude 14.4, and radial velocity with respect to the Milky Way 6940 km s^{-1}. Find the distance, diameter and absolute magnitude of the galaxy. What potential sources of error can you think of?

Exercise 20.2 The radial velocity of NGC 772 is 2562 km s^{-1}. Compute the distance obtained from this information and compare the result with Exercise 18.1.

Exercise 20.3 If the neutrinos have nonzero mass, the universe can be closed. What is the minimum mass needed for this? Assume that $\Lambda = 0$, the density of neutrinos is 600 cm^{-3}, and the density of other matter is one tenth of the critical density.