1
Cosmology

The science that treats the properties and evolution of the Universe as a whole is *cosmology*. Among the sciences, it is unique in having only a single object of study – there are no other Universes for us to use as controls, nor can we readily run the whole experiment over again. As a consequence, much of the effort in modern cosmology has been to determine the best mathematical description, or 'model', of the Universe we inhabit. As we shall see, that task is not yet complete, despite the rapid advances of the past few decades. The range of possible models is presented later in this chapter. First, though, we need to look at the observational bases of modern cosmology, a set of astronomical observations which have established the Hot Big Bang theory and restricted the range of models we need to consider.

1.1 Astronomical constituents of the Universe

Since cosmology is the study of the Universe as a whole and as a single system, it is only indirectly concerned with subsystems within the Universe. Here, I will mention only two: galaxies and clusters of galaxies. The galaxies are assemblies of 10^8–10^{12} stars; many galaxies also contain appreciable amounts of interstellar gas and dust. Some of the basic physical parameters of galaxies are: radius, typically 10^3–10^4 parsecs*; luminosity, typically 10^7–10^{11} times the luminosity of the sun, or very roughly 10^{34}–10^{38} W; and mass, typically 10^8–10^{12} times the mass of the sun or about 10^{41}–10^{45} g.† The question of the mass of galaxies introduces an important issue in contemporary astronomy – the possible existence of 'dark' non-luminous matter in galaxies (see Kormendy and Knapp, 1987; Trimble, 1987; or Primack *et al.*, 1988). The mass of galaxies, especially galaxies of the spiral form shown in fig. 1.1, can be determined by measuring the speed of their rotation as a function of distance from the center, then applying a generalization of Kepler's Third Law. The masses so determined are in almost every case larger than the sum of the masses of all the stars in the galaxy, often by a factor of 3–10 or so. In addition, the measured rotation speed in the outer parts of such galaxies does not fall off as $r^{-1/2}$, but stays essentially constant (fig. 1.2), suggesting that the bulk of the mass of

* The parsec (1 pc = 3.08×10^{18} cm) is the standard unit of distance used in astronomy, and will be used throughout this book. A further word on units: workers in the field use a jumble of S.I. and c.g.s. units as well as some purely astronomical units like the parsec. In general, I will use the units that have become conventional in the field, giving conversions where necessary to physical units.

† In astronomy, 'solar units' are conventionally used for luminosity and mass. The luminosity of the sun is $L_\odot = 3.9 \times 10^{33}$ erg s^{-1}; the solar mass is $M_\odot = 1.99 \times 10^{33}$ g.

Fig. 1.1 A typical spiral galaxy (M81). Photograph from the Palomar and Mt. Wilson Observatories.

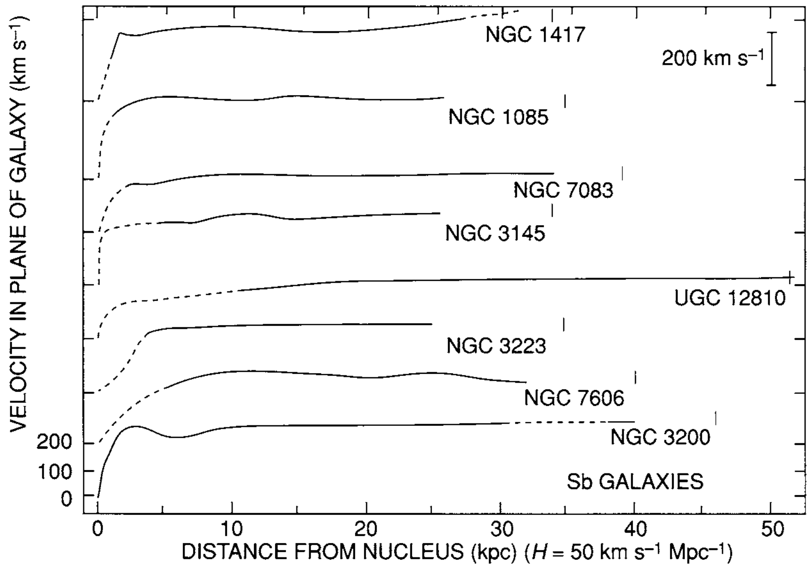

Fig. 1.2 A set of rotation curves (plots of rotation velocity as a function of distance from the center) for several spiral galaxies (from Rubin *et al.*, 1982, with permission). The velocity stays approximately constant well beyond the visible limit of the galaxy, rather than dropping as $r^{-1/2}$, suggesting an extended halo of 'dark matter' in these galaxies.

Fig. 1.3 A cluster of galaxies. A photograph from the Palomar and Mt. Wilson Observatories made with the 200-inch telescope.

galaxies is less concentrated than the luminous matter such as stars or gas. This is the 'dark matter' that we will have occasion to refer to here and in Chapters 7 and 8.

All galaxies emit radio waves as well as optical radiation at some level. In a minority of galaxies, however, the radio luminosity exceeds the optical luminosity; these are the *radio galaxies*, which, together with quasi-stellar objects, make up most of the extra-galactic sources detectable with radio telescopes. We will deal with the radio emission from our own Milky Way Galaxy in Chapter 4, and with radio sources in general in Chapter 7.

Most galaxies are clumped together in small groups (our own Galaxy is a member of the Local Group, as is the Andromeda Galaxy, M31), or larger *clusters* of a few hundred to a few thousand galaxies. An example is shown in fig. 1.3. The clusters contain matter between their constituent galaxies. In some clusters, this matter is directly detected; it is ionized gas at a temperature of about 10^7–10^8 K, which emits detectable X-ray flux (this intergalactic plasma is discussed further in Chapter 8). In other clusters, the evidence for intergalactic matter is indirect. The total gravitational mass of a cluster required to hold it together may be derived by applying the virial theorem to the cluster, assuming that it is in equilibrium. The result is

$$M = \frac{R_c \bar{v}^2}{G}, \tag{1.1}$$

where v^2 is the mean-square velocity of the galaxies within the cluster, and R_c is the radius of the cluster (see Chapter IV of Peebles, 1971). The mass of clusters calculated in this fashion is 10^{47}–10^{48} g, in most cases an order of magnitude larger than the sum of the masses of the individual galaxies. Estimates of the mass of the hot intergalactic gas detected in some clusters show that it cannot account for the discrepancy; it is insufficient to bind clusters gravitationally. Nor do there appear to be enough intergalactic stars to bring the mass of clusters up to the value calculated from eqn. (1.1). Once again, the existence of some form of 'dark' matter is suggested, in this case lying between the galaxies.

1.2 Observational bases of Big Bang cosmology

We now turn to some astronomical observations, which establish properties of the Universe as a whole, and upon which our present cosmological theories are based.

1.2.1 Homogeneity

The presence of clusters of galaxies and more careful analysis of the counts of galaxies (Peebles, 1980) show that, on relatively small cosmological scales, $d \lesssim 30$ Mpc $\equiv 3 \times 10^7$ pc or about 10^{26} cm, the galaxies are inhomogeneously distributed (see figure 8.2). On larger scales $d \gtrsim 300$ Mpc, however, the distribution is approximately isotropic and homogeneous. We thus arrive at the first observational basis of cosmology – on sufficiently large scales, the Universe appears to be homogeneous and isotropic (see Section 8.2, however).

1.2.2 Expansion

One of the landmark discoveries of 20th century science is the recognition that the Universe is an *expanding* system. This expansion was discovered and characterized in the late 1920s by Edwin Hubble, who found that the atomic lines detected in the spectra of distant galaxies almost always appear at wavelengths slightly greater than the rest or laboratory wavelengths of those same atomic lines – that is, they are shifted to longer wavelengths or *redshifted*. The redshift, z, is defined by

$$z + 1 \equiv \lambda_{obs}/\lambda_{rest}, \qquad (1.2)$$

where λ_{obs} is the observed wavelength. Hubble also found that, on the average, the magnitude of redshift observed in the spectrum of a galaxy was proportional to its distance, d, from us.

Hubble interpreted the redshifts he observed as instances of the Doppler effect; for recession velocities $v \ll c$, eqn. (1.2) gives

$$z = v/c \propto d.$$

In this interpretation, recession velocity is proportional to distance. This linear relation is just what one would expect for uniform expansion of the Universe. The measured constant of proportionality in the relation between v and d is now known as Hubble's constant, H_0, and is evidently a measure of the *rate* of expansion of the Universe:

$$v = H_0 d, \quad \text{or} \quad z = H_0 d/c. \qquad (1.3)$$

Astronomical measurements of the redshift and d show that H_0 lies in the range $(1.3\text{–}3.2) \times 10^{-18}$ s^{-1}, or in more conventional astronomical units, 40–100 km s^{-1} per megaparsec. The factor of two uncertainty arises primarily from the difficulty of making reliable measurements of the distance of extragalactic objects (see Rowan-Robinson, 1985). To account for the uncertainty in H_0, we will generally write it as $100h$ km s^{-1} per megaparsec, with $0.4 \lesssim h \lesssim 1.0$.

The strictly linear relationship between redshift and distance breaks down for larger distances and higher velocities (see Weinberg, 1972). Since the redshift is a more easily measured quantity than distance itself, it is commonly used by cosmologists to parameterize the distance to a galaxy or other source, and is so used in this book.

While Hubble interpreted the redshift as a Doppler shift induced by motion of galaxies, the modern interpretation, based on ideas introduced in General Relativity, is somewhat different. In modern cosmological theory, the galaxies are taken as more or less fixed* in a geometry that is itself expanding. The apparent relative recessional velocity of an observer and a distant galaxy is then explained by the expansion of space between the two. The expansion is specified through a quantity R known as the *scale factor*, which is time dependent and increasing. The distance between any two objects in the Universe at time t may thus be written as

$$d_{12}(t) = \frac{R(t)}{R(t_0)} d_{12}(t_0),$$

where $d_{12}(t_0)$ is the distance between those two objects at present (denoted throughout as t_0), and $R(t_0)$ is the present value of the dimensionless scale factor. $R(t_0)$ is often set equal to 1, and we will follow that convention. Since $\dot{R} > 0$, it follows that all lengths and distances measured in this expanding space were shorter in the past. That statement is true of the wavelengths of freely propagating photons as well (Weinberg, 1972). It thus follows that $\lambda_{obs} = R^{-1}(t)\, \lambda_{rest}$, for a photon emitted at some earlier time t. Hence

$$R(t) = [z(t) + 1]^{-1}, \tag{1.4}$$

establishing the connection between the scale factor and redshift.

Likewise, if $R(t_0) \equiv 1$, it may easily be shown that $H_0 = \dot{R}(t_0)$, where the subscript '0' is used to show explicitly that we are concerned with the present value of both the scale factor and Hubble's 'constant,' since both may be functions of time.

1.2.3 Age of the Universe

If there are no forces to slow down the expansion of the Universe, $\dot{R}$ will remain constant. Under these conditions, a backward extrapolation of the present expansion reveals that $R = 0$ at some finite time in the past. As the scale factor R goes to zero, so do all distances. Hence the density goes to infinity and we cannot sensibly extrapolate further into the past. This moment of infinite (or at least very high) density is the Big Bang origin of the Universe. Again, assuming a constant value for $\dot{R}$, it is easy to show that the time elapsed since the Big Bang is H_0^{-1}. This interval is the present age of the Universe, t_0. For $\dot{R}$ = constant, t_0 lies in the range about $(3–7) \times 10^{17}$ s or 10–20 billion years, depending on the value assumed for H_0. As we shall see, for more realistic cosmological assumptions, this result is in fact an upper limit on t_0. Support for the Big Bang theory is provided by independent geophysical and astronomical measurements of the age of various constituent parts of the Universe. The Earth–Moon system, for instance, and by inference the solar system, is known to be 4.6 billion years old. The age of certain long-lived radio-isotopes found in meteoritic material is 11–12 billion years (see

* Galaxies may have small random or even systematic velocities relative to this background geometry. These peculiar velocities, as they are called, are typically a few hundred kilometers per second and are discussed further in Chapter 8.

Fowler, 1987; Weinberg, 1972; or Narlikar, 1983) in reasonable agreement with H_0^{-1}. So too is the calculated age of the oldest stars in our Galaxy (see summary by Tayler, 1986). It is important to note that no objects within the Universe have yet been found with ages clearly in excess of H_0^{-1} – thus age measurements are consistent with the Big Bang theory.

Finally, the age of the Universe establishes a very rough limit to its extent. In a Universe of age t_0, photons can have traveled at most a distance about ct_0, and hence an astronomer cannot 'see' further than about ct_0. This distance, very roughly 5×10^9 pc, is the effective radius of the Universe. It is important to note that this same argument implies that the Universe was smaller in the past, since it was younger. (The role of particle horizons will reappear in Chapter 8; see also standard cosmology texts.)

1.2.4 Evidence for a Hot Big Bang

The discovery of the cosmic microwave background radiation (henceforth abbreviated CBR) established that the early Universe was hot as well as dense. The key to this argument is the blackbody or thermal spectrum of the radiation (the observations are presented in Chapter 4). Let us ask what happens to a blackbody radiation field if we extrapolate backwards in time to an epoch when the scale factor R was smaller, so $z > 0$. The wavelength of all photons is decreased proportionally to R or $(z + 1)^{-1}$. The Planck function, however, depends only on the product λT. It follows (see Chapter 5) that the spectrum of the radiation was also blackbody in the past, but the temperature was higher by a factor $z + 1$ (see Weinberg, 1972, Section 15.5):

$$T(t) = T_0(z + 1), \tag{1.5}$$

where T_0 is the present temperature of the CBR, approximately 3 K. Knowing the present value of the temperature, we can calculate the temperature at any earlier epoch using eqn. (1.5). For instance, for redshifts greater than 1000, the temperature was > 3000 K, sufficient to ionize the major atomic constituent of the Universe, hydrogen. At still larger redshifts, corresponding to earlier times in the history of the expanding Universe, the temperature was even greater. Note, however, that the strict linear dependence of $z + 1$ and T breaks down at higher temperatures, where the number of light particle species goes up (see Kolb and Turner, 1990).

One earlier epoch is of particular interest. A few minutes after the Big Bang origin of the Universe, the temperature dropped to about 10^9 K, low enough to permit fusion of neutrons and protons present in the hot primordial plasma (see Section 1.6.4 below). The nuclei of light elements, primarily ^{4}He, were produced. This process of primordial nucleosynthesis has been extensively studied (Peebles, 1966; Wagoner, Fowler and Hoyle, 1967; Schramm and Wagoner, 1979; Audouze, 1987), and detailed predictions have been made of the abundances of the light nuclei produced in the Hot Big Bang. These predicted abundances (fig. 1.4) agree well with astronomical determinations of the abundances of these same nuclei in the oldest stars and other matter in our Galaxy (Boesgaard and Steigman, 1985; Walker *et al.*, 1991), providing additional strong support for the Hot Big Bang model.

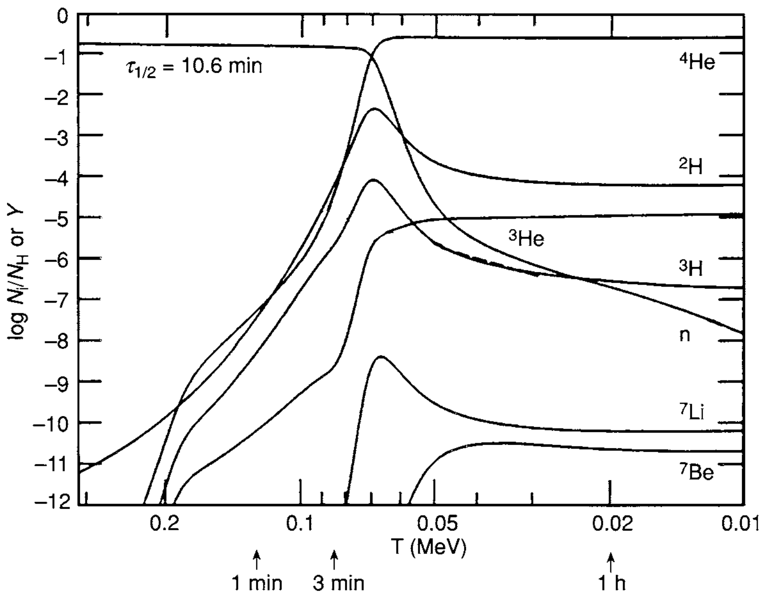

Fig. 1.4 Predicted abundances for the light nuclei produced in the first few minutes of a Hot Big Bang. Note the initial rapid rise in the abundance of deuterium (^{2}H) as the Universe cooled. Between t = 1 and 4 min, the deuterium was incorporated into other nuclei, especially ^{4}He. Adapted from Wagoner (1973); $n_b/n_\gamma = 3 \times 10^{-10}$ was assumed.

1.3 Cosmological models

We will now incorporate these observations into a general mathematical description for the properties of the Universe as a whole. Such descriptions are called *cosmological models.* Note the plural; as we will soon see, many mathematical models are consistent with the observational evidence now available. Of course, only a single model can *best* describe the Universe, and much of the effort in modern cosmology has been devoted to testing the models observationally, with the hope of reducing the range of possible models.

Like most models in physics, cosmological models ignore some of the details (e.g., inhomogeneities in the Universe). Most are based on the *cosmological principle*, the notion that the Universe is isotropic and homogeneous on a large scale or, more descriptively, 'the Universe is the same everywhere.'

1.3.1 The Robertson–Walker metric

If the Universe is isotropic and homogeneous on a large scale, the underlying geometry of the Universe must also be isotropic (exceptions are discussed in Section 8.3). The space–time geometry of the Universe may be completely specified by giving its metric tensor $g_{\mu\nu}$ – see texts on General Relativity; Peebles (1971), Weinberg (1972) or Narlikar (1983), for instance.

For a general set of four space–time coordinates, x^μ, the invariant interval ds^2 is given in terms of the metric tensor as

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu.$$

For instance, ordinary Minkowski space of Special Relativity has

$$g_{00} = 1, \quad g_{11} = g_{22} = g_{33} = -1,$$

and all other, off-diagonal, elements are 0; here x^0 is chosen to be ct.

The presence of mass (and hence gravity) in the Universe precludes use of the Minkowski metric in cosmology. Instead, the appropriate metric for an expanding isotropic Universe is the Robertson–Walker metric (see Robertson and Noonan, 1968; Peebles, 1971; or Weinberg, 1972, for a derivation and further details). In spherical coordinates, the metric is

$$ds^2 = c^2 \, dt^2 - R^2(t)\left\{\frac{dr^2}{1-kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2\right\}. \quad (1.6)$$

The quantities r, θ and ϕ are coordinates fixed in the expanding geometry and are called *comoving coordinates*. As we have noted, the galaxies are approximately at rest in comoving coordinates, and expansion is accounted for by the scale factor, $R(t)$. The spatial part of the Robertson–Walker metric can have three global curvatures, depending on the value of the quantity k. For $k = 0$, the spatial geometry of the Universe is flat, i.e. Euclidean, so that comoving distances are given by the usual relation, $d^2 = x^2 + y^2 + z^2$. The geometry may also be positively or negatively curved (with $k \lessgtr 0$), however, in which case $d^2 \gtrless x^2 + y^2 + z^2$, respectively. A positively curved Robertson–Walker metric is a closed geometry, limited in volume but without edges, just as the two-dimensional surface of a sphere is closed, finite and without boundaries. The negatively curved case, like the flat case, is an open, infinite geometry.

1.3.2 Density and curvature

These three possible curvatures are directly linked by General Relativity to the amount of matter in the Universe. A high density produces positive curvature; in the low density case, the curvature is negative. The particular density corresponding to a flat geometry, the *critical density*, is denoted ρ_c; in the next section we will evaluate it numerically.

1.3.3 Dynamics

We know that the Universe is expanding now, so that $R(t)$ increases with time. If there were no forces* to alter the expansion, $\dot{R}$ would remain constant, as shown by curve a in fig. 1.5. Since the density of the Universe is nonzero, however, we know that at least one long-range attractive force is acting – gravity. It acts to slow the expansion. The slowing of the expansion may be represented schematically by curvature in fig. 1.5 – see curves b and c. One evident consequence of the presence of matter in the Universe is that the present age of the Universe (defined as the time since $R = 0$ at the Big Bang) is less than H_0^{-1}.

The relation governing expansion of the Universe – that is, the function $R(t)$ – may be found by solving the field equations of General Relativity (see Weinberg, 1972). Here,

* I recognize that, for pedagogical purposes, I am mixing Newtonian concepts like force with General Relativistic ones like space curvature.

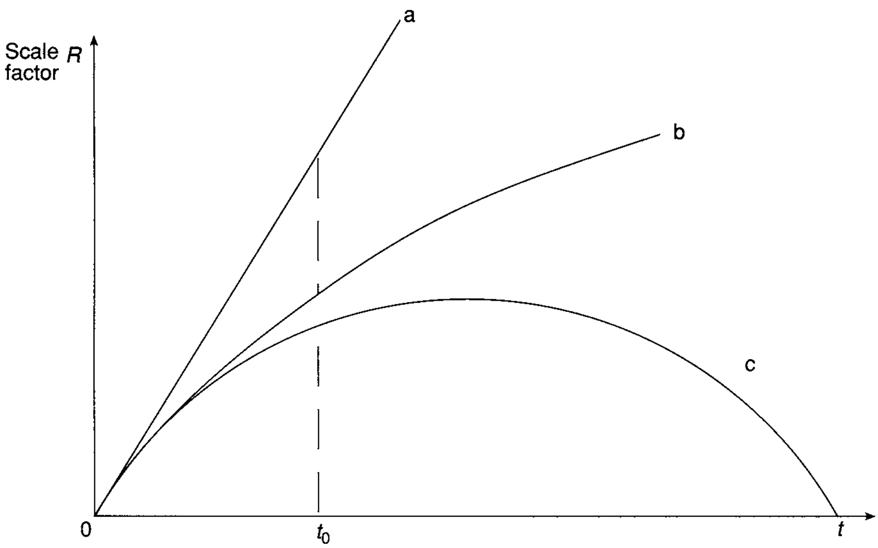

Fig. 1.5 The scale factor R as a function of time for various cosmological models. The different models are specified by their space curvature, k, and by the mean value of present density, ρ_0. The slope of the curves at the present epoch, t_0, is fixed by measurements of Hubble's constant, H_0. From the figure, it may be seen that the age of the Universe (the time since $R = 0$) is less than H_0^{-1} for models with $\rho_0 > 0$.

following McCrea and Milne (1934) and Callan, Dicke and Peebles (1965), I will take a simpler, but quite valid, approach using Newtonian physics.

Consider a sphere centered at an arbitrary point O in the expanding Universe. Let its radius at a particular time be $R(t)$; we will assume that $R(t)$ is large enough that the sphere represents a fair sample of the Universe, yet small enough that the curvature of space can be neglected. Now consider the acceleration of a unit mass on the sphere's surface towards O; it is

$$\ddot{R}(t) = -\frac{GM}{R^2(t)}, \tag{1.7}$$

where M is the mass inside the sphere.* In turn, in this Newtonian model,

$$M = \frac{4}{3}\pi R^3(t)\rho(t). \tag{1.8}$$

Here $\rho(t)$ is the density at time t.

We now make use of the conservation of mass. If the density of the Universe includes only material particles, which interact only gravitationally and exert no pressure (we specifically *exclude* radiation), then $V(t)\,\rho(t) = V(t_0)\,\rho(t_0)$, which leads to

$$\rho(t) = \frac{R^3(t_0)}{R^3(t)}\rho(t_0).$$

If, as above, we set $R(t_0) = 1$, and write the present density as ρ_0 for simplicity, we have for this simple case excluding radiation,

* The rest of the isotropic, homogeneous Universe outside the sphere exerts no net force on the unit mass.

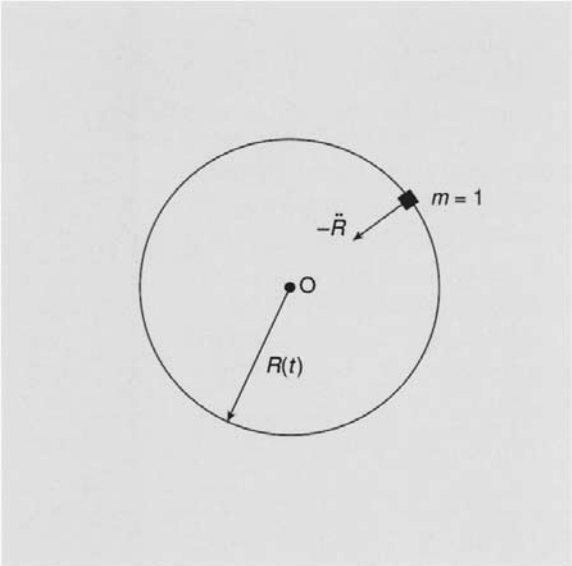

Fig. 1.6 A sphere of radius R about an arbitrary point O in a homogeneous Universe; $-\ddot{R}$ is the magnitude of the inward acceleration of the test mass, m.

$$\rho(t) = R^{-3}(t)\,\rho_0 \,. \tag{1.9}$$

Combining (1.7), (1.8) and (1.9),

$$\ddot{R}(t) = -\frac{4}{3}\pi G\rho_0 R^{-2}(t). \tag{1.10}$$

The first integral of eqn. (1.10) may be found by multiplying both sides by $\dot{R}$, then noting that

$$\ddot{R}\dot{R} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{R}^2}{2}\right), \quad \dot{R}R^{-2} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{-1}{R}\right).$$

Thus, after integration, we find

$$\dot{R}^2(t) = \frac{8}{3}\pi G\rho_0 R^{-1}(t) + \text{constant}. \tag{1.11}$$

In this Newtonian calculation, it may easily be shown that the constant of integration is related to the total energy per unit mass. In the full, General Relativistic solution, the connection between dynamics, density and space curvature becomes manifest; the constant of integration is found to be $-kc^2$. Hence, finally,

$$\dot{R}^2(t) = \frac{8}{3}\pi G\rho_0 R^{-1}(t) - kc^2, \tag{1.12}$$

where $k > 0$, $= 0$ or < 0 for positively curved, flat or negatively curved space, respectively. Although we have used Newtonian concepts in this derivation, the result is the same as is found (e.g. Weinberg, 1972) from a fully relativistic calculation.

Solutions to eqn. (1.12), which is the basic equation of mathematical cosmology, are presented in texts such as those by Bondi (1960), Peebles (1971), Weinberg (1972),