1 Stars and the Universe

1.1 Setting the stage

Stars are not distributed randomly in the universe, but are assembled through gravitational interactions into galaxies. Typical distances between stars in a given galaxy are of the order of 1 parsec (pc) whereas distances between galaxies are typically of the order of 100 kpc–1 Mpc (1 pc is the distance at which the semi-major axis of Earth orbit subtends an angle of 1 arcsecond; this corresponds to $\sim$3.26 light years, where one light year is the distance travelled by light in one year, i.e. 9.4607×10^{17} cm).

There are three basic types of galaxies: spirals, ellipticals and irregulars (see Figure 1.1). Spiral galaxies (our galaxy, the Milky Way, is a spiral galaxy) constitute more than half of the bright galaxies that we observe within $\sim$100 Mpc of the Sun. They generally comprise a faint spherical halo, a bright nucleus (or bulge) and a disk that contains luminous spiral arms; spirals have typical masses of the order of $10^{11} M_{\odot}$ ($1M_{\odot}$ denotes one solar mass, i.e. 1.989×10^{33} g). Spirals are divided into normal and barred spirals, depending on whether the spiral arms emerge from the nucleus or start at the end of a bar springing symmetrically from the nucleus. Dust and young stars are contained in the disk whereas the nucleus and halo are populated by older stars. Elliptical galaxies account for $\sim$10 per cent of the bright galaxies, have an elliptical shape, no sign of a spiral structure nor of dust and young stars, a mass range between $\sim 10^5$ and $\sim 10^{12} M_{\odot}$, and in general resemble the nuclei of spirals. There is no sign of significant rotational motions of the stars within ellipticals, whereas stars in the disks of spirals show ordered rotational motion.

These two broad types of galaxies are bridged morphologically by the so-called lenticular galaxies, which make up about 20 per cent of the galaxies, and look like elongated ellipticals without bars and spiral structure. The third broad group of galaxies are the irregulars, that show no regular structure, no rotational symmetry and are relatively rare and faint.

Evolution of Stars and Stellar Populations Maurizio Salaris and Santi Cassisi

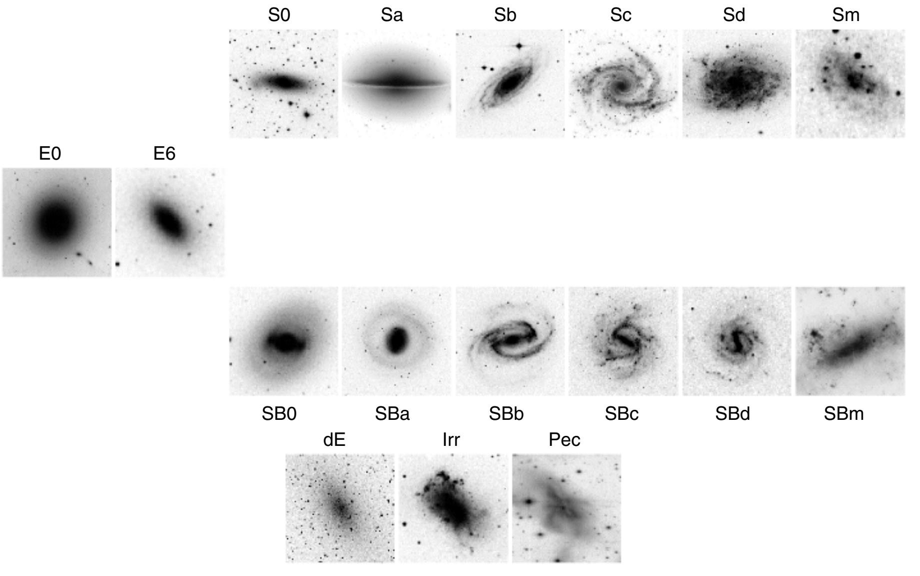

Figure 1.1 The so-called *tuning fork diagram*, i.e. the galaxy morphological classification. Elliptical galaxies are denoted by E (the various subclasses are denoted by the approximate value of the ellipticity) spirals by S, barred spirals by SB; examples of Dwarf elliptical (dE), Irregular (Irr) and Peculiar (Pec) galaxies are also displayed (courtesy of P. James)

Many galaxies show various types of non-thermal emission over a large wavelength range, from radio to X-ray, and are called active galaxies. These active galaxies display a large range of properties that can probably be explained invoking one single mechanism (possibly related to accretion of matter onto a black hole); the difference in their properties is most likely due to the fact that we are observing the same kind of object at different angles, and therefore we see radiation from different regions within the galaxy. Examples of active galaxies are the Seyfert galaxies, radio galaxies, BL Lac objects and quasars. There are also so-called starburst galaxies, e.g. galaxies displaying a mild form of activity, and showing a strong burst of star formation.

For many years it was believed that galaxies extend as far as they are visible. However, starting from the 1970s, the orbits of neutral hydrogen clouds circling around individual spiral galaxies provided rotation curves (e.g. rotational velocity as a function of the distance d from the galactic centre) that, instead of dropping as $\sqrt{d}$ beyond the edge of the visible matter distribution (as expected from Keplerian orbits after the limit of the mass distribution is reached) show a flat profile over large distances well beyond this limit. This can be explained only by a steady increase with distance of the galaxy mass, beyond the edge of the visible mass distribution. This dark matter reveals its presence only through its gravitational pull, since it does not produce any kind of detectable radiation.

Even galaxies are not distributed randomly in the universe, but are aggregated in pairs or groups, which in turn are often gathered into larger clusters of galaxies. Our galaxy (often referred to as the Galaxy) belongs to the so-called Local Group of galaxies, that includes about 20 objects (mainly small) among them the Large Magellanic Cloud (LMC) the Small Magellanic Cloud (SMC) and Andromeda (M31). The nearest cluster of galaxies is the Virgo cluster (at a distance of about 20 Mpc). Further away are other galaxy clusters, among them the Coma cluster, located at a distance of about 100 Mpc, that contains thousands of objects. Deep galaxy surveys (e.g. the APM, COSMOS, 2dF and SDSS surveys) have studied and are still probing the distribution of galaxies in the universe, and have revealed even more complex structures, like filaments, sheets and superclusters, that are groupings of clusters of galaxies.

Dark matter is also found within clusters of galaxies. This can be inferred studying the X-ray emission of the hot ionized intracluster gas that is accelerated by the gravitational field of the cluster. A rough comparison of visible and dark matter contribution to the total matter density of the universe tells us that about 90 per cent of the matter contained in the universe is dark.

It is evident from this brief description that overall the universe appears to be clumpy, but the averaged properties in volumes of space of the order of 100 Mpc are smoother, and the local inhomogeneities can be treated as perturbations to the general homogeneity of the universe.

The dynamical status of the universe is revealed by spectroscopic observations of galaxies. The observed redshift of their spectral lines shows that overall galaxies are receding from us (in the generally accepted assumption that the observed redshift is due to the Doppler effect) with a velocity v that increases linearly with their distance D, so

$$v = H_0 \times D,$$

as first discovered by Hubble and Humason during the 1920s (hence the name of Hubble law for this relationship). The constant H_0 is called the Hubble constant. Taken at face value this relationship seems to locate us in a privileged point, from where all galaxies are escaping. However, if one considers the overall homogeneity of the universe, the same Hubble law has to apply to any other location and the phenomenon of the recession of the galaxies might be looked upon as an expansion of the universe as a whole; a useful and widely used analogy is that of the two-dimensional surface of a balloon that is being inflated. If the galaxies are points drawn on the surface of the balloon, they will appear to be receding from each other in the same way as the Hubble law, irrespective of their location.

Superimposed on the general recession of galaxies are local peculiar velocities due to the gravitational pull generated by the local clumpiness of the universe. For example, the Milky Way and M31 are moving towards each other at a speed of about $120\,\mathrm{km\,s^{-1}}$, and the Local Group, is approaching the Virgo Cluster at a speed of $\sim 170\,\mathrm{km\,s^{-1}}$. On a larger scale, the Local Group, Virgo Cluster and thousands of other galaxies are streaming at a speed of about $600\,\mathrm{km\,s^{-1}}$ towards the so called Great Attractor, a concentration of mass in the Centaurus constellation, located at a distance of the order of 70 Mpc. These peculiar velocities become negligible with respect to the general recession of the galaxies (Hubble flow) when considering increasingly distant objects, for which the recession velocity predicted by the Hubble law is increasingly high.

Another discovery of fundamental importance for our understanding of the universe was made serendipitously in 1965 by Penzias and Wilson. Observations of electromagnetic radiation in a generic frequency interval reveal peaks associated with discrete sources – i.e. stars or galaxies – located at specific directions; when these peaks are eliminated there remains a dominant residual radiation in the microwave frequency range. The spectrum of this cosmic microwave background (CMB) radiation is extremely well approximated by that of a black body with a temperature of 2.725 K. After removing the effect of the local motion of the Sun and of our galaxy, the CMB temperature is to a first approximation constant when looking at different points in the sky, suggesting a remarkable isotropy which is hard to explain in terms of residual emission by discrete sources. From the CMB temperature one easily obtains the energy density associated to the CMB, ϵ_{CMB}, given that $\rho_{\mathrm{CMB}} = \epsilon_{\mathrm{CMB}}/c^2 \sim 4.64 \times 10^{-34}\,\mathrm{g\,cm^3}$

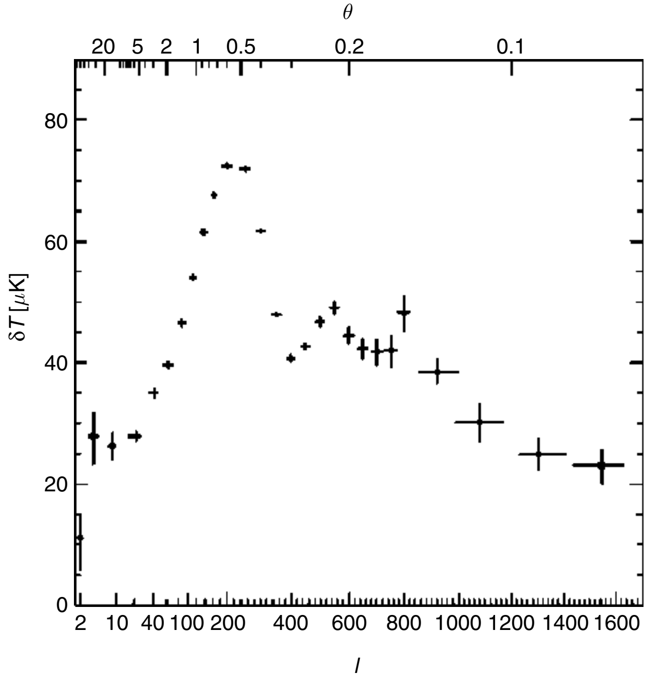

Figure 1.2 Plot of the CMB temperature fluctuations (in units of 10^{-6} K) as a function of the angular scale in degrees (upper horizontal axis) and the so-called wave number $l \sim \pi/\theta$ (lower horizontal axis); this is also called the power spectrum of the CMB fluctuations

(c denotes the speed of light, $2.998 \times 10^{10}\,\mathrm{cm\,s^{-1}}$). This CMB photon density is the dominant component of the present radiation density in the universe; a rough comparison of ρ_{CMB} with the present matter density ρ shows that at the present time the density associated with the photons is about three orders of magnitude lower than the matter density, including the dark matter contribution.

In 1992 the COBE satellite first discovered tiny variations δT of the CMB temperature, of the order of $\delta T/T \sim 10^{-5}$ (where T is the global mean of the CMB temperature) when looking at different points in the sky. By computing the average over the sky of the ratio $\delta T/T$ (temperature fluctuation) measured from any two points separated by an angle θ, one obtains what is called the angular power spectrum of the CMB temperature anisotropies, displayed in Figure 1.2. This power spectrum shows the existence of a series of peaks located at specific angular scales.

A comprehensive theory for the structure and evolution of the universe must be able to explain the basic observations outlined above in terms of evolutionary processes rooted in accepted physics theories. The following sections introduce briefly the Hot Big Bang theory, which is the presently widely accepted cosmological theory. Detailed presentations of cosmology at various levels of complexity can be found in [11], [57], [118] and [142].

1.2 Cosmic kinematics

A cornerstone of the Big Bang theory is the so-called cosmological principle: it states that the large-scale structure of the universe is homogeneous and isotropic. Homogeneity means that the physical properties of the universe are invariant by translation; isotropy means that they are also rotationally invariant. Both these properties can be applied only considering average properties of large volumes of space, where the local structures (galaxies, clusters of galaxies) are smeared out over the averaging volumes.

As discussed before, the adequacy of the cosmological principle can be empirically verified by studying the distribution of clusters of galaxies on scales of the order of 100 Mpc and by the isotropy of the CMB. Locally the universe is clumpy, but this clumpiness disappears when averaging the matter density over large enough volumes. In this way the local clumpiness is treated as a perturbation to the general smoothness of the universe. The universe is then treated as a fluid whose particles are galaxies, moving according to the Hubble law; within this picture of a cosmic fluid the cosmological principle implies that every co-moving observer (i.e. moving with the Hubble flow) in the cosmic fluid has the same history.

A first step when discussing events happening in the universe is to set up an appropriate coordinate system. For the time coordinate a natural choice is to use standard clocks co-moving with the cosmic fluid, that will define a cosmic time t; an operational way to synchronize t for co-moving observers at different locations is to set t to the same value when each observer sees that a property of the cosmic fluid, i.e. the average local density of matter ρ, has reached a certain agreed value. After synchronization, by virtue of the cosmological principle, the two observers must be able to measure exactly the same value (possibly different from the one at the time of synchronization) of that property whenever their clocks show the same time.

As for the three spatial coordinates, the cosmological principle greatly restricts the possible geometries. The assumption of homogeneity and isotropy requires that the tridimensional space has a single curvature, i.e. it must have the same value at all positions, but can in principle depend on time. The space–time interval ds between two events in an homogeneous and isotropic static space can be written as follows

$$ds^2 = c^2 dt^2 - \left(\frac{dr^2}{1 - Kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right)$$

where K is the spatial curvature, dt the cosmic time separation, r the radial coordinate and θ and ϕ the polar and azimuthal angles in spherical coordinates, respectively. The expansion (or contraction) of the universe can be accounted for by redefining the radial coordinate r as $r \equiv R(t)\chi - \chi$ being dimensionless – and the curvature K as $K(t) \equiv k/R(t)^2$. The constant k and coordinate χ are defined in a way that $k = +1$ for a positive spatial curvature, $k = 0$ for a flat space and $k = -1$ for a negative curvature. $R(t)$ is the so-called cosmic scale factor, that has the dimension of a distance and is dependent on the cosmic time t. With these substitutions one obtains the so-called Friedmann–Robertson–Walker (FRW) metric:

$$ds^2 = c^2 dt^2 - R(t)^2 \left(\frac{d\chi^2}{1 - k\chi^2} + \chi^2 d\theta^2 + \chi^2 \sin^2 \theta d\phi^2 \right) \tag{1.1}$$

The values of the three spatial coordinates χ, θ and ϕ are constant for an observer at rest with respect to the expansion of the cosmic fluid. One can easily see that the factor $R(t)$ in Equation (1.1) allows a scaling of the spatial surfaces that depends only on time, thus preserving the homogeneity and isotropy dictated by the cosmological principle. It is important to stress that it is only by virtue of the cosmological principle that we can uniquely define a four-dimensional coordinate system co-moving with the cosmic flow. As an example, the definition of cosmic time would be impossible in a universe without homogeneity and isotropy, because we could not synchronize the various clocks using mean properties (that would not be the same everywhere at a given time t) of the cosmic flow.

The geometrical properties of the three-dimensional space determined by the value of k can be briefly illustrated as follows. Let us consider at a cosmic instant t a sphere with centre at an arbitrary origin where $\chi = 0$, and surface located at a fixed value χ; the difference between the coordinates of the centre of the sphere and the surface is equal to $r = R(t)\chi$. The area A of the spherical surface of coordinate radius $r = R(t)\chi$ is, by definition, $A = 4\pi r^2 = 4\pi R(t)^2 \chi^2$. The physical radius R_p of the spherical surface is the distance between the centre and surface of the sphere measured with a standard rod at the same cosmic time t. This means that one has to determine the interval Δs^2 between the two events assuming $dt = 0$, so that $R_p = \sqrt{-\Delta s^2}$. From the FRW metric one obtains

$$R_p = R(t) \int_0^\chi \frac{d\chi}{\sqrt{1 - k\chi^2}} \tag{1.2}$$

R_p is equal to $R(t) \sin^{-1} \chi$, $R(t)\chi$ and $R(t) \sinh^{-1} \chi$ when $k = 1, 0$ and -1, respectively. When $k = 0$ one has $\chi = R_p/R(t)$, and $A = 4\pi R_p^2$, i.e. r is equal to R_p and the area A increases as R_p^2, as in Euclidean geometry. When $k = +1$ one has $r = R(t) \sin(R_p/R(t))$ and $A = 4\pi R(t)^2 \sin^2(R_p/R(t))$, which reaches a maximum value $A = 4\pi R(t)^2$ when $R_p = (\pi/2)R(t)$, then becomes zero when $R_p = \pi R(t)$ and has in general a periodic behaviour. This means that in the case of $k = +1$ space is closed and the periodicity corresponds to different circumnavigations. In the case of $k = -1$ then $A = 4\pi R(t)^2 \sinh^2(R_p/R(t))$, which increases with R_p faster than in the case of a Euclidean space.

It is easy to see how simply $R(t)$ describes the observed expansion of the universe. Let us set $\chi = 0$ at the location of our own galaxy, that is approximately co-moving with the local cosmic fluid (hence its spatial coordinate does not change with time) and consider another galaxy – also at rest with respect to the expansion of the universe – whose position is specified by a value χ of the radial coordinate (the angles θ and ϕ are assumed to be equal to zero for both galaxies). Its proper distance (defined in the same way as for the proper radius R_p discussed before) D at a given cosmic time t is given by:

$$D = R(t) \int_0^\chi \frac{d\chi}{\sqrt{1 - k\chi^2}}$$

As in the case of Equation (1.2) D is equal to $R(t) \sin^{-1} \chi$, $R(t)\chi$ and $R(t) \sinh^{-1} \chi$ when $k = 1, 0$ and -1, respectively. The velocity v of the recession of the galaxy due to the expansion of the universe is

$$v = \frac{dD}{dt} = \frac{dR(t)}{dt} \int_0^\chi \frac{d\chi}{\sqrt{1 - k\chi^2}} = \frac{dR(t)}{dt} \frac{1}{R(t)} D$$

This looks exactly like the Hubble law; in fact, by writing

$$H(t) = \frac{dR(t)}{dt} \frac{1}{R(t)} \tag{1.3}$$

we obtain

$$v = H(t) \times D$$

$H(t)$ corresponds to the Hubble constant and one can notice that its value can change with cosmic time. The value of $H(t)$ determined at the present time is denoted as H_0.

This result is clearly independent of the location of the origin for the radial coordinate χ, since any position in the universe is equivalent according to the cosmological principle. It is important to notice that locally, e.g. within the solar system or within a given galaxy, one cannot see any effect of the cosmic expansion, since the local gravitational effects dominate. For distances large enough ($D > c/H(t)$) the last equation predicts recession velocities larger than the speed of light, an occurrence that seems to go against special relativity. The contradiction is, however, only apparent, given that galaxies recede from us faster than the speed of light (superluminal recession) because of the expansion of space; locally, they are at rest or moving in their local inertial reference frame with peculiar velocities $\ll c$.

In the following section we will briefly describe the observational counterpart of $v = H(t) \times D$ and show how it probes the evolution of the kinematic status of the universe.

1.2.1 Cosmological redshifts and distances

What we measure to estimate the recession velocity of galaxies is a redshift z, that can be related to the change of $R(t)$ with time. Consider light reaching us (located at $\chi = 0$) from a galaxy at a radial coordinate χ. Two consecutive maxima of the electromagnetic wave are emitted at times t_e and $t_e + \delta t_e$ and received at times t_0 and $t_0 + \delta t_0$; if $\delta t_e = \delta t_0$ we would not observe any redshift since the wavelength of the electromagnetic wave is given by the spatial distance between the two consecutive maxima, i.e. the observed wavelength is $\lambda_0 = c\delta t_0$, and the emitted one is $\lambda_e = c\delta t_e$. We will now find the relationship between δt_e and δt_0. Since $ds = 0$ for light, we have

$$\int_{t_e}^{t_0} \frac{dt}{R(t)} = \frac{1}{c}\int_0^{\chi} \frac{d\chi}{\sqrt{1-k\chi^2}}$$

$$\int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{R(t)} = \frac{1}{c}\int_0^{\chi} \frac{d\chi}{\sqrt{1-k\chi^2}}$$

for the first and second maximum, respectively. The right-hand side of both equations is the same, therefore we can write

$$\int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{R(t)} - \int_{t_e}^{t_0} \frac{dt}{R(t)} = 0$$

The first term on the left-hand side of the previous equation can be rewritten as

$$\int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{R(t)} = \int_{t_e}^{t_0} \frac{dt}{R(t)} + \int_{t_0}^{t_0+\delta t_0} \frac{dt}{R(t)} - \int_{t_e}^{t_e+\delta t_e} \frac{dt}{R(t)}$$

and therefore

$$\int_{t_0}^{t_0+\delta t_0} \frac{dt}{R(t)} - \int_{t_e}^{t_e+\delta t_e} \frac{dt}{R(t)} = 0$$

The intervals δt_e and δt_0 are negligible compared with the expansion timescale of the universe, and therefore $R(t)$ is to a good approximation constant during these two time intervals; inserting this condition into the previous equation provides

$$\frac{\delta t_0}{R(t_0)} = \frac{\delta t_e}{R(t_e)}$$

The redshift $z = (\lambda_0 - \lambda_e)/\lambda_e$ is therefore given by

$$z = \frac{\delta t_0}{\delta t_e} - 1 = \frac{R(t_0)}{R(t_e)} - 1 \tag{1.4}$$

In an expanding universe $z > 0$ (since $R(t_0) > R(t_e)$) as observed. If the redshift is small enough, i.e. t_e is close to t_0 in cosmological terms, we can expand $R(t_e)$ about t_0 using the Taylor formula, and retain only the terms up to the second order:

$$R(t_e) = R(t_0) + (t_e - t_0)\frac{dR(t_0)}{dt} + \frac{1}{2}(t_e - t_0)^2 \frac{d^2R(t_0)}{dt^2}$$

We can now define H_0 as

$$H_0 \equiv H(t_0) = \frac{dR(t_0)}{dt}\frac{1}{R(t_0)}$$

i.e. the present value of the Hubble constant, and the so-called deceleration parameter

$$q_0 \equiv -\frac{d^2R(t_0)}{dt^2}\frac{1}{R(t_0)H_0^2} \tag{1.5}$$

Both H_0 and q_0 are related to the present rate of expansion of the universe. H_0 measures the actual expansion rate, whilst q_0 is positive if the expansion is slowing down (hence the name deceleration parameter) or negative if the opposite is true. With these definitions the second-order expansion of $R(t_e)$ can be rewritten as

$$R(t_e) = R(t_0)\left[1 + H_0(t_e - t_0) - \frac{1}{2}q_0 H_0^2 (t_e - t_0)^2\right]$$

and after additional manipulations one obtains the following useful results:

$$z = H_0(t_0 - t_e) + H_0^2(t_0 - t_e)^2\left(1 + \frac{1}{2}q_0\right) \tag{1.6}$$

$$t_0 - t_e = \frac{1}{H_0}\left[z - \left(1 + \frac{1}{2}q_0\right)z^2\right] \tag{1.7}$$

$$\chi = \frac{c}{R(t_0)H_0}\left[z - \frac{1}{2}(1 + q_0)z^2\right] \tag{1.8}$$

These relationships between z, H_0 and q_0 hold in the case of a redshift due to the expansion of the universe. Superimposed on the expansion of the universe are local peculiar velocities (e.g. blue- and redshifts) due to the motions caused by local anisotropies in the matter distribution; an example is local motions in clusters and groups of galaxies due to the gravitational potential of the cluster itself. These effects are minimized by observing suitably distant objects, where the velocities corresponding to the expansion of the universe become so large that they make local peculiar motions negligible.

From an observational point of view, the Hubble law needs, in addition to the measurements of the redshift z, an estimate of galaxy distances. This is usually done by comparing the observed flux l received from certain standard candles (i.e. objects of known intrinsic luminosity L) with their intrinsic luminosities. Traditionally one uses the inverse square law to determine the distance:

$$d = \left(\frac{L}{4\pi l} \right)^{1/2} \tag{1.9}$$

This result is based on the conservation of energy and assumes a flat static space. In cosmology, the distance obtained through Equation (1.9) is called the luminosity distance, and is denoted by d_{L}.

Consider a light source located at a radial co-moving coordinate χ; at a given cosmic time t_{e} the source emits photons that reach the observer located at $\chi = 0$ at time t_0. By the time the light reaches the observer it is distributed uniformly across a sphere of coordinate radius $R(t_0)\chi$. The area of the spherical surface at the observer location centred at the source is therefore given by $4\pi R(t_0)^2\chi^2$. The photons emitted by the source are redshifted by the expansion of the universe, and their energy is therefore reduced by a factor $(1 + z)$ when measured by the observer; this is because the wavelength is increased by a factor $(1 + z)$ and the photon energy is proportional to the inverse of its wavelength. There is also an additional reduction by a factor $(1 + z)$ due to the so-called time dilation effect, i.e. the observer receives less photons per unit time than emitted at the source. This can easily be understood by means of the same arguments as were applied in the case of the wave maxima, that led to the notion of redshift. We found before that the time between two consecutive maxima at emission is different from that at reception; the same holds for the time interval between photons emitted by the source, and implies that the rate of reception of photons is different from the rate of emission. Taking into account these two effects, conservation of energy dictates that:

$$l = \frac{1}{(1+z)^2} \frac{L}{4\pi R(t_0)^2 \chi^2}$$

We now define the luminosity distance d_{L} of the observed source, according to Equation (1.9); one obtains $d_{\mathrm{L}} = R(t_0)\chi(1+z)$, which can be rewritten using Equation (1.8) as (retaining the terms up to the second order in z):

$$d_{\mathrm{L}} = \frac{cz}{H_0}\left[1 + \frac{1}{2}(1 - q_0)z\right] \tag{1.10}$$

The first term is the empirical Hubble law, with the recession velocity given by the product cz. The higher-order correction term is proportional to the deceleration parameter q_0 and starts to play a role when $z > 0.1$.

Another way to determine cosmological distances is to consider objects (e.g. galaxies) with known diameter D_{p}, and compare the measured angular diameters Θ with the intrinsic ones. One can define a diameter distance $d_{D_{\mathrm{p}}}$ as

$$d_{D_{\mathrm{p}}} = \frac{D_{\mathrm{p}}}{\Theta} \tag{1.11}$$

which is equal to d_{L} for a flat static space. Consider an object located at the radial co-moving coordinate χ, that emits light at time t_{e}; if the observer is located at $\chi = 0$ and receives the light from the object at t_0, the relationship between D_{p} and Θ can easily be obtained by determining $\sqrt{\Delta s^2}$ where Δs is obtained integrating the FRW metric with $dt = d\chi = d\phi = 0$. This provides $d_{D_{\mathrm{p}}} = R(t_{\mathrm{e}})\chi$. By comparing the latter equation with $d_{\mathrm{L}} = R(t_0)\chi(1+z)$ obtained before and using the definition of z we obtain

$$d_{D_{\mathrm{p}}} = \frac{d_{\mathrm{L}}}{(1+z)^2}$$

In principle $d_{D_{\mathrm{p}}}$ is different from d_{L}, but the two distances converge to the same value when $z \to 0$.

It should be clear from this brief discussion that the empirical study of the trends of d_{L} and $d_{D_{\mathrm{p}}}$ with redshift z provides an estimate of the kinematical parameters H_0 and q_0. A third possible method to determine the kinematical status of the universe involves number counts of galaxies with a flux greater than some specified value l $(N(l))$. Assuming there are n galaxies per unit volume, in a static flat universe (with uniform distribution of galaxies) one expects

$$N(l) = \frac{4}{3}\pi n \left(\frac{L}{4\pi l}\right)^{3/2}$$

where L is the intrinsic galaxy luminosity, supposed constant. For an expanding universe it can be shown that (as a second-order approximation in z)

$$N(l) = \frac{4\pi n(t_0)}{3}\left(\frac{L}{4\pi l}\right)^{3/2}\left[1 - \frac{3H_0}{c}\left(\frac{L}{4\pi l}\right)^{1/2}\right]$$

where $n(t_0)$ is the number density at the present time (e.g. in the low redshift universe); notice that by a fortuitous cancellation of terms this relationship does not depend on q_0. The correction term to the static flat case is always negative, so that in principle one should always observe fewer sources than predicted by the simple $l^{-3/2}$ formula.

There are many practical difficulties in implementing these three tests; the reason is that we are assuming the existence of perfect standard candles and the absence of evolutionary effects on the size, and brightness of galaxies. Evolutionary effects are particularly important since a high redshift means a time far in the past, when galaxies had a very different age from the present one. A detailed discussion of these classical cosmological tests and the related observational problems can be found in [187]. In recent years the class of stellar objects called *Type Ia supernovae* (see Section 7.6) has been used as an effective standard candle and applied with great success to study the d_{L}–z relationship (see [146]).

We conclude this section by discussing briefly the concept of particle horizon in an FRW expanding universe. In general, as the universe expands and ages, a generic observer is able to see increasingly distant objects as the light they emitted has time to arrive at the observer's location. This implies that as time increases, increasingly larger regions of space come into causal contact with the observer, who will therefore be able to 'see' increasingly larger portions of the universe. We can ask ourselves what is the co-moving coordinate χ_H of the most distant galaxy we can see at a given cosmic time t. Increasing values of χ_H with time mean that we are actually seeing more and more distant galaxies (supposed to be at rest with respect to the cosmic expansion) as the time increases. Consider a radially travelling photon, for which $ds = 0$. From the FRW metric we obtain

$$\int_0^t \frac{dt'}{R(t')} = \frac{1}{c}\int_0^{\chi_H} \frac{d\chi}{\sqrt{1-k\chi^2}}$$

and therefore

$$\begin{aligned}
\chi_H &= \sin\left(c\int_0^t \frac{dt'}{R(t')}\right) \quad k=+1 \\
\chi_H &= c\int_0^t \frac{dt'}{R(t')} \quad k=0 \\
\chi_H &= \sinh\left(c\int_0^t \frac{dt'}{R(t')}\right) \quad k=-1
\end{aligned} \tag{1.12}$$

If the space has $k=0$ or $k=-1$ it is in principle possible, for specific forms of $R(t)$, to have an infinite χ_H; this means that all galaxies in the universe might eventually be visible at a certain time t for particular forms of the function $R(t)$. If $k=+1$ the behaviour of χ_H is periodic, and if the argument of the sine function is equal to or larger than π, one can sweep the entire universe.

1.3 Cosmic dynamics

The previous discussion about the kinematics of the cosmic fluid was based exclusively on the properties of the FRW metrics which, in turn, depend only on the hypothesis of homogeneous and isotropic cosmic fluid. To determine the behaviour of $R(t)$ with cosmic time t and the value of k we need to apply a theory for the physical force(s) governing the evolution of the cosmic fluid. The only fundamental interaction able to bridge the relevant cosmological scale is the gravitational force, therefore we need to use a theory of gravity – the general relativity theory – to describe the evolution of FRW universes.

The case of a space with the FRW metrics provides the equation

$$\left(\frac{dR(t)}{dt}\right)^2 = -kc^2 + \frac{8\pi G\rho(t)R(t)^2}{3} \tag{1.13}$$

where G is the gravitational constant ($6.6742 \times 10^{-8}\,\mathrm{dyn\,cm^2\,g^{-2}}$) and ρ is the matter density. Equation (1.13) was obtained in 1922 by Friedmann, who solved Einstein's field equations for an isotropic and homogeneous universe. As we will see in a moment, these equations predict an expanding universe. A more general form of the field equations contains the constant Λ – called the cosmological constant – introduced by Einstein in 1917 in order to obtain static universes (the expansion of the universe had not been discovered yet). It is important to notice that the value of Λ must be small in absolute terms, since the planetary motions in the solar system are well described by the Einstein field equations with $\Lambda = 0$. Including Λ in the gravitational field equations provides

$$\left(\frac{dR(t)}{dt}\right)^2 = -kc^2 + \frac{(8\pi G\rho(t) + \Lambda)R(t)^2}{3} \tag{1.14}$$

It is clear that the evolution of $R(t)$ is controlled by the density (ρ), the geometry (k) and the cosmological constant (Λ). By using the definition of $H(t)$ one can rewrite Equation (1.14) as

$$H(t)^2 = -\frac{kc^2}{R(t)^2} + \frac{8\pi G\rho(t)}{3} + \frac{\Lambda}{3} \tag{1.15}$$

It is customary to introduce the critical density $\rho_c \equiv 3H(t)^2/(8\pi G)$ and define the density parameter $\Omega_\rho = \rho/\rho_c$, an equivalent for the cosmological constant $\Omega_\Lambda = \Lambda/(3H(t)^2)$, and the sum $\Omega = \Omega_\rho + \Omega_\lambda$. With these definitions Equation (1.15) becomes

$$(1 - \Omega)H(t)^2R(t)^2 = -kc^2 \tag{1.16}$$

We can immediately see from this form of the Friedmann equation that there is an intimate connection between the density of matter plus the cosmological constant, and the geometry of space. $\Omega = 1$ gives a flat space, $\Omega > 1$ a positive curvature, and $\Omega < 1$ a negative curvature. It is also important to notice the obvious fact that Ω changes with time, since $H(t)$ and $R(t)$ both change with t, but the product kc^2 is a constant.

1.3.1 Histories of $R(t)$

Equation (1.14) enables us to perform a simple analysis of the behaviour of $R(t)$ for various model universes, once an additional equation for the density is obtained; this equation can be determined by applying the first principle of thermodynamics to the cosmic fluid. In an isolated system the first law of thermodynamics states that $dU = -PdV$ where U is the internal energy of the system, V its volume and P the pressure. The internal energy is ρc^2 times the volume V (i.e. the energy associated with the rest mass of the matter) so that the time evolution of the system according to the first law is

$$\frac{d(\rho(t)c^2V(t))}{dt} = -P\frac{dV(t)}{dt}$$

which can also be rewritten as

$$\frac{d(\rho(t)c^2R(t)^3)}{dt} = -P\frac{dR(t)^3}{dt} \tag{1.17}$$

using the fact that the volume V scales as $R(t)^3$. Let us now assume that the density is dominated by matter and not by radiation; this is a very good assumption since observationally – as discussed before – one finds that at the present time the matter density is about three orders of magnitude larger than the density associated with radiation ($\rho_{\mathrm{r}} = \epsilon_{\mathrm{r}}/c^2$, where ϵ_{r} is the photon energy density). If the matter is non-relativistic (a correct assumption for almost the whole evolution of the universe) its pressure is negligible with respect to ρc^2 and Equation (1.17) provides

$$\frac{d\rho(t)}{dt}\frac{1}{\rho(t)} = -3\frac{dR(t)}{dt}\frac{1}{R(t)} \tag{1.18}$$

which implies

$$\rho(t)R(t)^3 = \rho(t_0)R(t_0)^3$$

where t_0 is the present cosmic time and t a generic value. This reflects the simple fact that the density of non-relativistic matter is decreasing because of dilution as space is expanding. If photons were to be the dominant contributor to the total density, the previous relationship would be different. In fact, for photons (and more generally for relativistic particles) $P = (\rho_r c^2)/3$ and P is no longer negligible with respect to U. Therefore Equation (1.17) would provide

$$\frac{d\rho_r(t)}{dt}\frac{1}{\rho_r(t)} = -4\frac{dR(t)}{dt}\frac{1}{R(t)} \tag{1.19}$$

and

$$\rho_r(t)R(t)^4 = \rho_r(t_0)R(t_0)^4$$

The scaling of $\rho(t)$ with $R(t)^4$ is firstly due to the decrease of the number density of photons as $R(t)^{-3}$ when the universe expands (since the volume increases as $R(t)^3$). In addition, the energy of individual photons decreases as $R(t)^{-1}$ because of the cosmological redshift and therefore both ϵ_r and ρ_r decrease with time as $R(t)^{-4}$, faster than the matter density.

By considering a matter dominated universe one now can rewrite Equation (1.14) as

$$\left(\frac{dR(t)}{dt}\right)^2 = -kc^2 + \frac{8\pi G\rho(t_0)R(t_0)^3}{3R(t)} + \frac{\Lambda R(t)^2}{3} \tag{1.20}$$

Differentiation of this equation with respect to t provides:

$$\frac{d^2R(t)}{dt^2} = -\frac{4\pi G\rho(t_0)R(t_0)^3}{3R(t)^2} + \frac{\Lambda R(t)}{3} \tag{1.21}$$

This equation shows clearly how the self gravitation of matter (represented by ρ) acts to slow down the expansion of the universe, because it appears as a negative contribution to the acceleration of $R(t)$. On the other hand, a positive Λ acts like a negative density and tends to accelerate the expansion of the universe; a particular choice of Λ makes the universe static (although in a situation of unstable equilibrium). The term $(\Lambda R(t))/3$ is often called the cosmic repulsion term.

It is now easy to determine some general properties of $R(t)$ in a matter dominated universe. If Λ is zero or negative the acceleration of $R(t)$ is always negative; at some time in the past $R(t)$ must have reached zero and therefore ρ was infinite (i.e. a singularity is attained). It is natural to set the zero point of the cosmic time at this instant, which can also be considered the origin of the universe. As for the future evolution, if Λ is negative, $R(t)$ will also intersect the t axis some time in the future (hence a final implosion) since the expansion will slow down, eventually stop and then reverse to a contraction. If Λ is zero the acceleration can become zero in the future if $R(t)$ becomes infinite, and therefore the expansion can slow down without ever being followed by a contraction. The precise behaviour depends in this case on the value of k. If $k = -1$ or 0 the future collapse is avoided, but not if $k = +1$.

If Λ is positive then $R(t)$ is not always decelerating and there is the possibility of avoiding a singularity in the past. In fact, if $k = +1$ one can obtain from Equations (1.20) and (1.21) that in the past there has been a minimum of $R(t)$ different from zero, given by $R^3_{\min} = (4\pi G\rho(t_0)R(t_0)^3)/\Lambda$ if the cosmological constant satisfies the following relation: $\Lambda < (c^6)/(4\pi G\rho(t_0)R(t_0)^3)^2$. As for the future evolution, if $k = 0$

or -1 the expansion continues forever, whereas if $k=+1$ the expansion may vanish and then be followed by a contraction, depending upon the value of Λ.

For historical interest we show briefly how it is possible to obtain a static universe by tuning the value of Λ. In a static universe both $R(t)$ and $\rho(t)$ are constant, and both velocity and acceleration of $R(t)$ are equal to zero. With these constraints Equations (1.20) and (1.21) provide $\Lambda = 4\pi\rho(t_0)G$, $k/R^2 = (4\pi\rho(t_0)G)/c^2$, where R denotes the constant value of $R(t)$. Since R has to be positive and k can be only equal to 0, $+1$, -1, we have that a static universe will have $k=+1$ and $R = c/\sqrt{4\pi\rho(t_0)G}$.

We conclude by providing analytical relationships between $R(t)$ and t for the case of flat geometry, i.e. $\Omega = 1$ and $k=0$, and arbitrary values of Λ, which are relevant to the presently favoured cosmological model. With this choice of parameters the universe began from a singular state ($R=0$ and $\rho=\infty$ at $t=0$) and Equation (1.20) gives (see also Figure 1.3):

$$
\begin{aligned}
R(t) &= R(t_0)\left(\frac{8\pi G\rho(t_0)}{\Lambda}\right)^{1/3}\sinh^{2/3}\left(\frac{1}{2}t\sqrt{3\Lambda}\right) \quad \Lambda > 0 \\
R(t) &= R(t_0)(6\pi G\rho(t_0))^{1/3}t^{2/3} \quad \Lambda = 0 \\
R(t) &= R(t_0)\left(\frac{8\pi G\rho(t_0)}{|\Lambda|}\right)^{1/3}\sin^{2/3}\left(\frac{1}{2}t\sqrt{3|\Lambda|}\right) \quad \Lambda < 0
\end{aligned} \tag{1.22}
$$

For $\Lambda = 0$ one obtains the very simple result $q_0 = 1/2$, $H(t) = 2/(3t)$ and therefore the age of the universe is $t_0 = 2/(3H_0)$. The quantity $1/H_0$ is often called Hubble time.

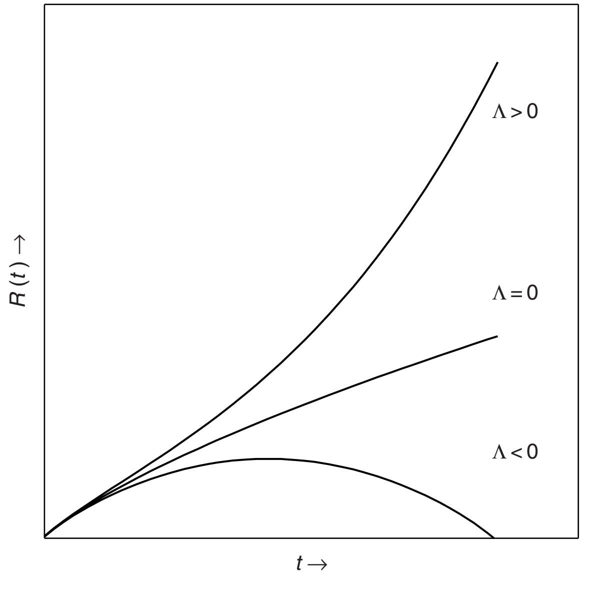

Figure 1.3 Qualitative behaviour of the scale factor $R(t)$ with respect to the cosmic time t for models with $\Omega = 1$ and $k=0$

1.4 Particle- and nucleosynthesis

We have already noticed that the density of matter in an expanding universe decreases with time slower than the density of photons. This means that as we go backwards in time the radiation density increases faster than the density of matter. Therefore, there must be a point in time when the two densities were equal and before that the universe was radiation dominated. If the actual densities of matter and radiation are $\rho(t_0)$ and $\rho_r(t_0)$, respectively, the equality is attained at

$$\frac{R(t_0)}{R(t_E)} = \frac{\rho(t_0)}{\rho_r(t_0)} = 1 + z_E$$

where t_E and z_E are, respectively, the cosmic time and redshift of matter–radiation equality; their values are of the order of 10^4–10^5 years and 10^3, respectively. It is worth noticing at this stage that the results about the trend of $R(t)$ with t we gave in the previous section were obtained assuming a matter dominated universe (negligible radiation density) at all time. The onset of a radiation dominated universe at the beginning of the evolution does not, however, alter the general results regarding the occurrence of an initial singularity, and also the quantitative relationship between $R(t)$ and t is not substantially changed, since – as we will soon see – the radiation dominated era lasts only a short time compared with the timescale of cosmological evolution.

In case of radiation $\epsilon_r = aT_r^4$ (where T_r is the radiation temperature and $a = 7.566 \times 10^{-15}\,\mathrm{erg\,cm^{-3}\,K^{-4}}$) $T_r(t) \propto R(t)^{-1}$ and therefore the radiation temperature was steadily increasing in the past. Calculations of the interaction cross section between photons and matter and the expansion rate of the universe show that during the radiation dominated epoch the interaction rate was high enough to ensure that at each instant there was (to a good approximation) thermodynamical equilibrium (see Chapter 2) i.e. photons followed a black-body distribution of energies characterized by the same temperature $T = T_r$ for both radiation and matter. During the radiation era $\rho_r(t)$ becomes so large that the contribution of the terms containing k and Λ in Equation (1.14) are negligible and we can write

$$\left(\frac{dR(t)}{dt}\right)^2 = \frac{8\pi G \rho_r(t) R(t)^2}{3}$$

This equation, in conjunction with $\epsilon_r = aT^4$ and $T_r(t)R(t) = T_r(t_0)R(t_0)$ does provide

$$\begin{aligned}
R(t) &= R(t_0)T_r(t_0)\left(\frac{32\pi G a}{3c^2}\right)^{1/4} t^{1/2} \\
T(t) &= \left(\frac{3c^2}{32\pi G a}\right)^{1/4} t^{-1/2} \\
\rho(t) &= \frac{3}{32\pi G} t^{-2}
\end{aligned} \tag{1.23}$$

These formulae for the early evolution of ρ and T do not contain adjustable constants; however, they do not strictly apply when approaching t_E, since in this case the matter contribution to ρ is not negligible, and also the contributions of the curvature and the cosmological constant may play a role.

There is also a minimum time t_P (Planck time) below which we cannot describe the evolution of the universe with Equation (1.23) due to quantum uncertainty. This stems from the uncertainty principle applied to the pair of physical variables energy E and time t, i.e. $\Delta E \Delta t > h/(2\pi)$ where h is the Planck constant ($h = 6.626 \times 10^{-27}$ erg s). Consider a length (Planck length) $l_P = ct_P$ that defines a region in causal contact at time t_P. A mass $m_P \sim \rho_P l_P^3$ is associated with this length scale (ρ_P is the density of matter at $t = t_P$) hence an energy $m_P c^2 = \rho_P (ct_P)^3 c^2$. The uncertainty relationship can therefore be rewritten as $\rho_P (ct_P)^3 c^2 t_P = \rho_P c^5 t_P^4 > (h/2\pi)$. From Equation (1.23) we have $\rho_P \approx 1/(Gt_P^2)$, and consequently $\rho_P c^5 t_P^4 \approx (c^5 t_P^4)/(Gt_P^2) > (h/2\pi)$, that provides $t_P > ((hG)/(2\pi c^5))^{1/2} \sim 10^{-43}$ s. Due to this quantum uncertainty we cannot be completely sure that there has really been a singularity at $t = 0$.

When $t \sim 10^{-43}$ s the universe was extremely hot, the temperature being of the order of 10^{32} K, that corresponds to an energy of $\sim 10^{19}$ GeV (an energy of 1 eV corresponds to a temperature of 1.1605×10^4 K from the relationship $E = K_B T$ where K_B is the Boltzmann constant equal to 1.3807×10^{-16} erg K). According to the currently accepted cosmology and particle physics theories, it is during the first epochs after the singularity that today's stable particles – the proton–neutron pair, electrons in a number that compensates for the electric charge of the protons, neutrinos – were produced. A description of what happened during those first moments of the evolution of the universe – the so-called Big Bang – has to be based on the knowledge of the four fundamental interactions (gravitational, strong, weak, electromagnetic) which we briefly summarize below.

According to the standard model of particle physics, the fundamental interacting particles – quarks and leptons – are all fermions (particles with spin 1/2). Leptons are the negatively charged particles electron, μ and τ, and the associated neutrinos ν_e, ν_μ, ν_τ. The rest masses are $\sim$0.0005 GeV for the electron (1 eV $= 1.7827 \times 10^{-33}$ g using the relationship $E = mc^2$), 0.106 GeV for μ and 1.178 GeV for τ. Neutrinos are supposed to be massless although recent experiments suggest a mass different from zero, but not yet well determined. There are six quark species (positively and negatively charged) called 'down', 'up', 'strange', 'charm', 'bottom' and 'top'. Their mass increases from $\sim$0.31 GeV for the 'down' quark up to $\sim$177 GeV for the 'top' quark. In addition, there are antiparticles for each lepton and quark.

The gravitational interaction involves all particles, it is described by general relativity, and is supposed to be mediated by a boson (particle with integer spin) called graviton. At the moment there is no established quantum theory of gravity, which is the reason why we cannot try to described what happened at $t < t_P$. The strong interaction is a short-range interaction mediated by gluons, a family of eight massless bosons, and involves the so-called hadrons (respectively baryons, like protons and neutrons, and mesons, the most relevant of them being the pions π^0, π^+, π^-) which are made of combinations of quarks. Baryons are made of triplets of quarks, mesons of pairs quark–antiquark (e.g. the proton is made by two ‘up’ and one ‘down’ quark, the neutron by two ‘down’ and one ‘up’ quark). The weak interaction involves all particles, has short range and is mediated by the W^+, W^- and Z^0 bosons, with masses of the order of $\sim$90 GeV. The electromagnetic interaction is a long-range interaction acting among charged particles, and is mediated by the photon (a massless boson).

According to the so-called Grand Unified Theories (GUT) the strong, weak and electromagnetic interactions were all unified into a single force mediated by superheavy bosons with masses of the order of 10^{15} GeV. This idea stems from the successful unification of weak and electromagnetic interactions into the electroweak force that separates into the two components at sufficiently low energies. If it is possible to unify the strong with the electroweak force at even higher energies has yet to be seen; there are various theories that are, however, difficult to test experimentally. Interestingly a GUT prediction is that the proton should decay with a timescale of $\sim 10^{32}$–10^{33} yr (this hypothetical decay has not been observed yet). A further goal of physics is to unify gravity with the other three forces (a unification which should happen at energies higher than GUT). Whether this is possible – in spite of various attempts – remains to be seen.

If GUT are a viable proposition (at the moment there is no experimental confirmation of their predictions) the physical conditions during the first moments right after the singularity were adequate to attain the unification of the four fundamental interactions. Following this line of thought one expects that the steady temperature decrease caused by the expansion of the universe has caused a number of spontaneous symmetry breaking, that have generated the separate interactions we see today. At $t = 10^{-43}$ s the gravitational force has separated from the other three interactions which are still unified into a single force. At energies between 10^{16} and 10^{14} GeV (between 10^{-38} and 10^{-35} seconds after the singularity) the strong force separated from the electroweak one. The superheavy bosons disappear rapidly due to annihilation or decay processes. At this stage the universe is made of leptons, antileptons, quarks, antiquarks, gluons and four bosons that mediate the electroweak interaction (and probably gravitons). At energies of the order of 10^2 GeV (about 10^{-11} seconds after the singularity) the electroweak interaction separates into the electromagnetic and weak one. The leptons (massless until this moment) acquire mass through the Higgs mechanism (probably also the neutrinos) and the bosons that mediated the electroweak interaction give rise to the massive W^+, W^-, Z^0 bosons and photons. Below $\sim$90 GeV the massive bosons disappear through annihilation or decay. At this stage the universe was made of photons (and probably gravitons) quark–antiquark and lepton–antilepton pairs. By about 10^{-6} seconds after the singularity, quarks combine into hadrons.

Photons and matter are in equilibrium through absorption and creation–annihilation processes. Particles and antiparticles continually annihilate each other but more pairs are produced from the high-energy photon field as long as $K_B T > 2mc^2$ where m is the rest mass of the particle and antiparticle pair, and T is the temperature. This means that the number of a given particle species with mass m and the photon number are about the same as long as the previous inequality is satisfied. When the temperature goes below $2mc^2/K_B$, the particle–antiparticle pairs of mass m annihilate, without being replaced by newly produced pairs. If there is an asymmetry between the number of particles and antiparticles, after the annihilation only the residual number of surviving particles or antiparticles will be left. We do not have, to date, any empirical evidence for the existence of antimatter in the universe. The antiprotons observed in the cosmic rays are consistent with the hypothesis of production by interaction of cosmic rays with the interstellar medium; therefore, at least for the Galaxy, there is no evidence of antimatter. Absence of γ rays from a cluster of galaxies due to nucleon–antinucleon annihilations is further evidence against the existence of antimatter in the universe. In order to have a universe populated only by matter, it is necessary to postulate an asymmetry between matter and antimatter so that the annihilation processes destroyed all antimatter leaving the excess of matter that we see today. To explain the observed ratio between photons and matter an initial matter–antimatter asymmetry of only $\sim(1/10^8)$ particles is needed. Proposed mechanisms to explain this asymmetry involve processes acting when the temperature drops below the threshold for the separation of the strong force from the electroweak one, but no definitive solution to this problem (unless one invokes an ad hoc initial condition) has been found yet.

At energies of the order of 1 GeV (about 10^{-5} s after the singularity) nucleons and antinucleons annihilate, leaving the small excess of nucleons arising from the asymmetry discussed before. When the energy is down to about 130 MeV pairs π^+–π^- annihilate and π^0 particles decay into photons

The μ leptons annihilate with the corresponding antiparticles at about 100 MeV (the more massive τ leptons annihilated at higher energies). As for the nucleons, protons (p) and neutrons (n) were constantly being transformed into each other via the following reactions:

$$\mathrm{n} \leftrightarrow \mathrm{p} + \mathrm{e}^- + \bar{\nu}_e$$

$$\mathrm{n} + \mathrm{e}^+ \leftrightarrow \mathrm{p} + \bar{\nu}_e$$

$$\mathrm{n} + \bar{\nu}_e \leftrightarrow \mathrm{p} + \mathrm{e}^-$$

involving electrons e^-, positrons e^+, electron neutrinos ν_e and electron antineutrinos $\bar{\nu}_\mathrm{e}$. The conversions from one particle to the other were easily accomplished as long as the energy was above 1.293 MeV, e.g. the energy corresponding to the mass difference between proton and neutron. The direct and inverse reactions were so frequent that an equilibrium was established between the number densities of protons and neutrons, given by

$$\frac{n_\mathrm{n}}{n_\mathrm{p}} = \mathrm{e}^{(-1.293\mathrm{MeV})/K_\mathrm{B}T}$$

At 100 MeV $n_\mathrm{n}/n_\mathrm{p} \sim 0.99$, decreasing down to ~ 0.22 when the energy is ~ 1 MeV. At about 3 MeV the neutrinos decouple, i.e. they do not interact any longer with the rest of the matter[1]; their decoupling happens after the annihilation of muons and

[1] Neutrinos have a very small interaction cross section. However, right after the singularity the universe was so dense that neutrinos were also tightly coupled to the other components of the cosmic fluid.

before the annihilation of electrons and positrons. From this moment on neutrinos travel essentially undisturbed by the other particles, their temperature still decreasing (they are relativistic particles) as $1/R(t)$. When the energy goes below ~ 1 MeV electron–positron pairs annihilate, leaving a small remainder excess of electrons. Neutrons cannot any longer be replenished fast enough, due to the electron–positron annihilation and the fast expansion of the universe (it is now about 20 seconds after the singularity) and the number ratio $(n_n/n_p) \sim 0.224$ attained at electron–positron annihilation decreases due to the neutron decay (half-life of the order of 10 minutes). Between 100 and 200 seconds after the singularity the energy has fallen to about 0.1 MeV ($T \sim 10^9$ K) and the nuclear fusion reaction

$$p + n \rightarrow {}^2D + \gamma$$

becomes an efficient producer of deuterium (at higher energies the deuterium produced was photodissociated by the energetic photons). Helium production reactions become efficient when ${}^2D$ is sufficiently abundant:

$$ {}^2D + {}^2D \rightarrow {}^3H + p $$

$$ {}^3H + {}^2D \rightarrow {}^4He + n $$

$$ {}^2D + p \rightarrow {}^3He + \gamma $$

$$ {}^3He + n \rightarrow {}^4He + \gamma $$

Since there are no stable nuclei with atomic mass 5 to 8, and because of the fast expansion of the universe that lowers the energies of the particles involved on very short timescales – the energy drops below ~ 30 keV (1 keV$=10^3$ eV) e.g. $\sim 3 \times 10^8$ K about 20 minutes after the singularity – the nucleosynthesis leaves about a fraction of 0.75 (by mass) of protons (hydrogen), ~ 0.25 of helium, $\sim 10^{-4}$ of deuterium, $\sim 10^{-5}$ of ^{3}He and $\sim 10^{-10}$ of lithium. The precise values of these abundances (see Figure 1.4) are determined by the competition between the expansion rate of the universe and the nucleon density. The formation of He is limited only by the availability of neutrons; to a good approximation the helium abundance is therefore set by the neutron abundance at the beginning of the nucleosynthesis. It therefore depends, albeit only mildly, on the value of the matter density, increasing for increasing density.

More precisely, the primordial element abundances depend on the density of baryonic matter that we denote with ρ_b and $\Omega_b = \rho_b/\rho_c$. It is important to notice that ρ_b is approximately the same as the density of baryonic plus leptonic matter, since the number of electrons equals the proton numbers to achieve charge neutrality, but electrons are about 10^3 times lighter than baryons. From now on we will denote with baryonic density the density of baryonic plus leptonic matter.

It is generally assumed that the density of dark matter does not play any role in this cosmological nucleosynthesis, since it is supposed to affect the cosmic fluid only via its contribution to the gravitational interaction after the end of the radiation dominated epoch.

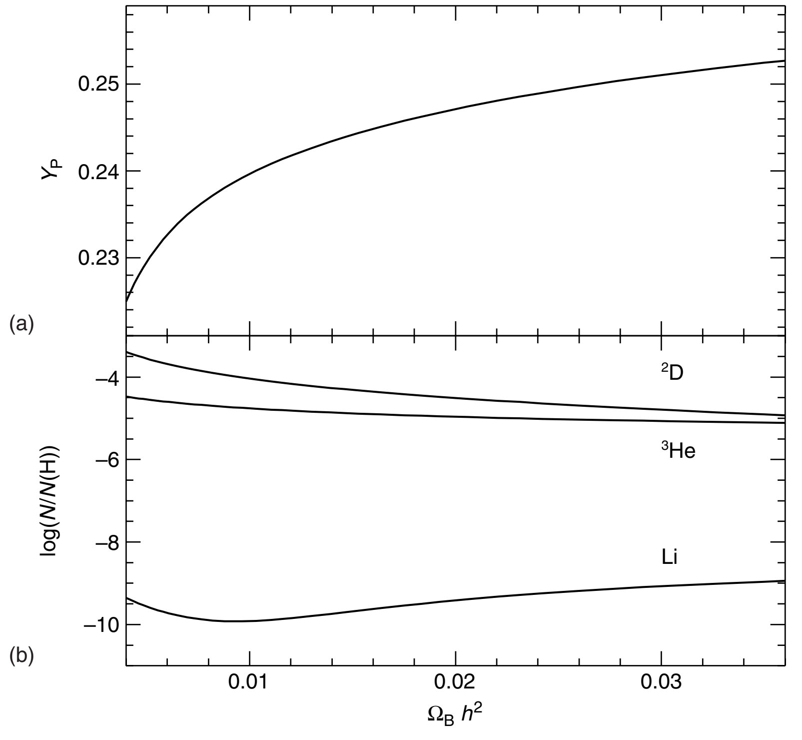

Figure 1.4 Abundances of ^{4}He ((a) in mass fraction), deuterium, lithium and ^{3}He ((b) number ratios with respect to hydrogen) produced during the primordial nucleosynthesis, as a function of the product $\Omega_B h^2$. Ω_B denotes the present (at $t = t_0$) value of the baryon density in units of the critical density, $h = H_0/(100\,\mathrm{km\,Mpc\,s^{-1}})$

The primordial abundances of deuterium and ^{3}He are decreasing functions of Ω_b; this is explained by the fact that the higher the matter density, the higher the temperature at which these elements reach an abundance high enough to start the production of helium, and consequently the higher their destruction rate. The behaviour of lithium is more complicated; for present values of Ω_b below ≈ 0.002[2] lithium is produced by direct fusion of helium with ^{3}H, whereas at higher densities (for present values of Ω_b larger than ≈ 0.02) it is produced by fusion of helium and ^{3}H producing ^{7}Be that transforms into lithium by electron captures. In both cases the final abundance of lithium increases with Ω_b. However, there is an intermediate region showing a dip in the abundance, due to the efficiency of a destruction reaction that involves a proton capture and a consequent decay into two helium nuclei.

[2] Once the present value of the density parameter Ω_b or Ω_r is given, its value at any earlier (or future) epochs can be obtained from Equations (1.18) and (1.19).

As the universe expands and cools, one reaches the time when the energy densities of matter and radiation are equal (at redshift ~ 3000) and immediately after that the matter density starts to dominate. Even so, matter and radiation are still tightly coupled through electron scattering processes. The universe is ionized, and matter is made mostly of protons (hydrogen nuclei) and free electrons. When the temperature drops below the ionization energy of hydrogen (13.6 eV) the ionization fraction stays close to one, due to the large excess number of photons over baryons (photons dominate by number, although matter dominates energetically and therefore gravitationally) so that the number of photons in the high-energy tails of the black-body spectrum is high enough to keep the matter fully ionized. Eventually the temperature, and therefore the number density, of sufficiently energetic photons drops so low that recombination prevails. It is at this time, $\sim 10^{5.5}$ years after the singularity (i.e. at redshift ~ 1000 and $T \sim 4000\,\mathrm{K}$) that the first atoms form. The resulting dearth of free electrons has the immediate consequence of reducing the efficiency of electron scattering, so that matter and radiation decouple. From this moment on the temperatures of radiation and matter become different and start to evolve separately; radiation no longer interacts with matter and can travel undisturbed through space, since the number of particles of matter is too low to produce significant interactions. The radiation temperature T_r is reduced according to $T_\mathrm{r} \propto R(t)^{-1}$, and the black-body spectrum it had at decoupling is preserved. This last point can be demonstrated as follows. For a black-body spectrum the number of photons with frequencies between ν and $\nu + d\nu$ contained in a volume of space $V(t)$ at cosmic time t is given by

$$dN(t) = \frac{8\pi\nu^2 V(t) d\nu}{c^3 (\mathrm{e}^{(h\nu/K_\mathrm{B} T_\mathrm{r}(t))} - 1)} \tag{1.24}$$

At a later time t' the frequency will be redshifted to $\nu' = \nu R(t)/R(t')$ and therefore $d\nu' = d\nu R(t)/R(t')$. The volume will have expanded to $V(t') = V(t)R(t')^3/R(t)^3$, but the number of photons within $V(t')$ will be the same as the number within $V(t)$ because of conservation (no appreciable interactions with the matter happen); the temperature $T_\mathrm{r}(t)$ will have also changed according to $T(t') = T(t)R(t)/R(t')$. By imposing $dN(t') = dN(t)$ and rewriting Equation (1.24) expressing ν, $V(t)$ and $T_\mathrm{r}(t)$ in terms of ν', $V(t')$ and $T_\mathrm{r}(t')$ according to the relationships given before, one obtains that

$$dN(t') = \frac{8\pi\nu'^2 V(t') d\nu'}{c^3 (\mathrm{e}^{(h\nu'/K_\mathrm{B} T_\mathrm{r}(t'))} - 1)}$$

i.e. of the same form as Equation (1.24).

This black-body radiation, homogeneous and isotropic (because of the cosmological principle) with a temperature T_r nowadays of the order of $\sim 3\,\mathrm{K}$ (as obtained from $T_\mathrm{r} \propto R(t)^{-1}$) is the theoretical counterpart of the observed CMB.

1.5 CMB fluctuations and structure formation

According to the scenario presented above, the CMB is the relic of the hot phase before decoupling, and provides us with information about the state of the universe when its age was only about a few 10^5 yr. The wealth of structures populating the universe nowadays suggests the existence of some density inhomogeneities in the cosmic fluid that have grown with time; if the universe was perfectly isotropic and homogeneous no structures would have formed with time, whereas in case of inhomogeneities, regions denser than the background tend to contract and get denser still, inducing a growth of the initial perturbation.

In 1970 Peebles, Yu, Sunyaev and Zel'dovich predicted that these inhomogeneities had to be imprinted in the CMB as the tiny temperature fluctuations that have recently been detected. In very simple terms, fluctuations of the local density of matter would have behaved as sound waves (with their fundamental mode plus overtones) in the cosmic fluid before recombination, with the photons providing the restoring force. The matter we are considering here is the baryonic matter, to which the photons are tightly coupled, whereas the dark matter did not have any interaction with photons. At recombination the photons started to travel unimpeded through space for the first time; photons released from denser, hotter regions were more energetic than photons released from more rarefied regions. These temperature differences were thus frozen into the CMB at recombination and are detected today. The shape of the observed CMB power spectrum is explained when one assumes that the phases of all the sound waves were synchronized at birth – i.e. that they were all triggered at the same time – and that the initial disturbances were approximately equal on all scales, e.g. the fluctuations on small scales had approximately the same magnitude as those affecting larger regions.

The first and highest peak in the CMB power spectrum (see Figure 1.2) corresponds to the fundamental wave of this acoustic oscillation; subsequent peaks represent the overtones. The power spectrum shows a strong drop off after the third peak (an effect known as Silk damping) due to the process of recombination. Since the recombination is not instantaneous, during its development the photon mean free path starts to progressively increase, producing a flow of photons from regions of high densities to lower density zones, that smooths out the small-scale temperature fluctuations. Most importantly, amplitude and location of the peaks are closely related to a number of cosmological parameters (for more details see, for example, the discussion in [112]); in particular, the location of the first peak is mainly related to the geometry of the three-dimensional space, whereas the ratio of the heights of the first to second peak is strongly dependent on Ω_b. Also the values of the Hubble constant and of the cosmological constant affect both the location and the amplitudes of the peaks albeit with different sensitivities.

Up to decoupling the density perturbations of the baryonic matter did not grow substantially (in fact the fluctuations of the CMB temperature are extremely small) due to the damping effect of the tightly coupled photons, and their evolution can be followed analytically with linear approximations. However, the perturbations involving dark matter are supposed to be able to grow more substantially (no interaction with the photons and no imprint left on the CMB power spectrum) and after baryonic matter decoupled from the radiation field, it could fall into and enhance the potential well of the dark matter condensations, thus starting to build up the structures we see today. The end point of the evolution of the primordial density fluctuations is the present statistical distribution of matter. This is generally very complicated, varying from point to point with objects of different sizes and masses (alternatively, fluctuations of various wavelengths and amplitudes) and its study is fundamental in order to determine the mechanisms of structure formation. A particular sticking point is the so-called biasing, i.e. the fact that the light distribution may not faithfully trace the mass distribution (we can only detect the luminous matter directly, not the dark matter). Studies not only of the light but also of the gravitational field in the observable structures can overcome this problem, i.e. by determining the peculiar velocities induced by the mass distribution through the gravitational interaction.

Numerical simulations of the evolution of perturbation and structure formation (see, for example [110]) have shown that the best assumption about the dark matter dynamical status is that it is 'cold', i.e. it has a negligible velocity dispersion; a 'hot' dark matter, i.e. matter with a large velocity dispersion, has been excluded, since it does not allow the formation of galaxies. As for the nature of this mysterious dark matter, various hypothetical particles predicted by GUT have been proposed as viable candidates; at the moment the question is still wide open, since – as mentioned before – there is no experimental confirmation for any of the proposed GUT and therefore for their predictions.

1.6 Cosmological parameters

Combining the location and amplitude of the peaks in Figure 1.2 to a number of constraints obtained from the spatial distribution of galaxies, the d_{L}–z empirical relationship using Type Ia supernovae as standard candles shown in Figure 1.5, and the empirical determination of the Hubble law at low redshifts, one can obtain a consistent picture for the fundamental cosmological parameters, within the framework of the Big Bang cosmology.

According to these estimates (see Table 1.1) the three-dimensional geometry of the universe is flat to a high degree of accuracy and the total density is dominated by the cosmological constant, whereas matter makes only $\sim$25 per cent of the total. In addition, the matter density appears dominated by the elusive dark matter, while the familiar baryonic matter makes only a negligible fraction of the total Ω. The estimated (small) value of the present baryonic matter density also agrees with some determinations of the ^{2}D abundance in high redshift gas clouds (supposed to be of primordial origin) and recent estimates of the initial He content in Galactic globular clusters (made of stars formed close to the Big Bang epoch). Since we do not know what dark matter is and what physical energy is represented by the cosmological constant, we are in the situation of ignoring the origin of more than 90 per cent of

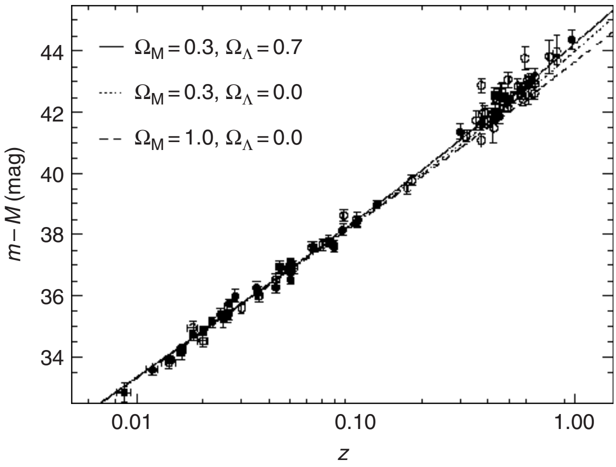

Figure 1.5 Empirical d_L–z relationship using Type Ia supernovae as distance indicators (filled and open circles; data from [146] and [164]) compared with the theoretical results for three different choices of the matter (baryonic plus dark – Ω_M) and cosmological constant (Ω_λ) density parameters (but the same value of H_0). The luminosity distance d_L is related to the displayed $(m - M)$ – called distance modulus – by $(m - M) = 5\log(d_L)\ -5$, where d_L is in parsec and $(m - M)$ is in magnitudes (see Chapter 8)

Table 1.1 Basic cosmological parameters (from [10])

Temperature CMB	T_{CMB} (K)	2.725 ± 0.002
Hubble constant	H_0 (km Mpc s^{-1})	71^{+4}_{-3}
Total density	Ω	1.02 ± 0.02
Cosmological constant density	Ω_Λ	0.73 ± 0.04
Baryon density	Ω_b	0.044 ± 0.004
Dark matter density	Ω_{DM}	0.22 ± 0.04
Photon density	$\Omega_\gamma (10^{-5})$	(4.800 ± 0.014)
Age of the universe	t_0 (Gyr)	13.7 ± 0.2
Redshift of matter–energy equality	z_{eq}	3233^{+194}_{-210}
Redshift of decoupling	z_{dec}	1089 ± 1
Age at decoupling	t_{dec} (10^3 yr)	379^{+8}_{-7}

the matter/energy content of the universe, although we can 'feel' its presence from its gravitational influence.

1.7 The inflationary paradigm

According to the results reported in Table 1.1 the universe is flat, with $\Omega = 1$ to a very good approximation. The flatness is potentially a problem because Equation (1.16) tells us that is an unstable condition. Consider a matter dominated universe. In this case one can neglect Λ in Equation (1.15), multiply both terms by $3/(8\pi G\rho(t))$ and obtain

$$\frac{3H(t)^2}{8\pi G\rho(t)} - 1 = -\frac{3kc^2}{8\pi G\rho(t)R(t)^2}$$

which can be rewritten as

$$(\Omega^{-1} - 1)\rho(t)R(t)^2 = \text{constant}$$

The right-hand side of this equation is a constant, and therefore we can write, for two different values of the cosmic time t,

$$(\Omega^{-1} - 1)\rho(t)R(t)^2 = (\Omega_0^{-1} - 1)\rho(t_0)R(t_0)^2$$

where the right-hand side contains the present values and the left-hand side the corresponding values at a given earlier time t. This latter equation can also be rewritten as

$$(\Omega^{-1} - 1) = (\Omega_0^{-1} - 1)\frac{\rho(t_0)}{\rho(t)}\left(\frac{R(t_0)}{R(t)}\right)^2$$

and since $R(t) = R(t_0)/(1+z)$, $\rho(t) = \rho(t_0)(1+z)^3$, we obtain

$$(\Omega^{-1} - 1) = \frac{(\Omega_0^{-1} - 1)}{(1+z)}$$

From this equation we can easily see that if Ω was at the beginning only slightly different from unity, then it could not possibly be equal to unity nowadays. For example, about 1 second after the Big Bang ($z \approx 10^{11}$) Ω had to be different from unity by less than $\approx 2 \times 10^{-13}$, in order to have Ω within 0.02 of unity today. This is the so called flatness problem: why was Ω so finely tuned?

The second problem faced by our understanding of the universe is the so-called horizon problem. The CMB across the sky is to a very good approximation isotropic, thus confirming one of the assumptions behind the cosmological principle. However, the size of the region in causal contact with a given observer increases with time for a flat universe, and its size at decoupling was much smaller than at the present time, corresponding to about only one degree in the sky today. Why are two points at the opposite sides of the sky at the same temperature (apart from the small primordial fluctuations) even though no information was able to travel from one to the other at decoupling?

A third question is: what is the origin of the primordial fluctuations and why were they triggered all at the same time?

One can, in principle, consider these three occurrencies as the initial conditions of our universe; however, to avoid such a finely tuned choice of initial conditions, the so-called inflationary paradigm was proposed in the 1980s by Guth, Linde, Sato, Albrecht and Steinhardt. The central idea is that there is a period in the early universe where a term $\Lambda_{\rm inf}$ – originated by some hypothetical quantum field – analogous to the cosmological constant dominates Equation (1.15), that can therefore be rewritten as

$$H(t)^2 = \frac{\Lambda_{\rm inf}}{3}$$

The solution of this equation, after recalling the definition of $H(t)$, and assuming a constant $\Lambda_{\rm inf}$, is

$$R(t) = R(t_{\rm i})e^{\sqrt{\Lambda_{\rm inf}/3}t} = R_{\rm i}e^{H(t)t}$$

if t is much larger than the cosmic time $t = t_{\rm i}$ of the beginning of the $\Lambda_{\rm inf}$ dominated epoch.

If this exponential expansion (inflation) is long enough, it will drive Ω towards 1, irrespective of its initial value; this happens because $R(t)$ increases exponentially, $H(t)$ is constant (its value set by the value of $\Lambda_{\rm inf}$) and therefore, following Equation (1.16), $\Omega \rightarrow 1$ if $R(t)$ has increased enough during this phase. Moreover, during inflation, a very small patch of the universe can grow to enormous dimensions, so that the isotropy of the CMB temperature, we see today, arose from a very small causally connected region that underwent an inflationary growth. An expansion by a factor of $\approx 10^{30}$ solves both the flatness and horizon problem without invoking ad hoc initial conditions. The quantum field that originated $\Lambda_{\rm inf}$ is expected to experience quantum fluctuations that were stretched by the inflation to the scales we see imprinted in the CMB. Therefore the simultaneous triggering of the primordial fluctuations is due to the onset of inflation. The general belief is that the inflation occurred when the strong force separated from the electroweak one, at about $t = 10^{-35}$ s, and lasted until about $t = 10^{-32}$ s.

1.8 The role of stellar evolution

The theory of stellar evolution that we will present in the following chapters is devoted to unveiling the physical and chemical properties of the stars populating the universe, and their development with time. The role played by stars in our understanding of the mechanisms driving the evolution of the matter created during the Big Bang is paramount. Take the human body as an example. Its chemical composition comprises $\sim$50 per cent of carbon, $\sim$20 per cent of oxygen, $\sim$8.5 per cent of nitrogen plus $\sim$10 per cent of heavier elements, the remaining fraction being hydrogen. Apart from hydrogen, no other elements were produced during the primordial nucleosynthesis.

After the groundbreaking study by Burbidge, Burbidge, Fowler and Hoyle ([26]) we know that all elements heavier than helium are produced in stars and injected into the interstellar medium by mass loss processes and explosions. Out of this chemically enriched matter new generations of stars are formed and the cycle is perpetuated until interstellar gas is available to form new stars. Stellar evolution theory is therefore a fundamental tool to understand the chemical composition of the present universe and how it evolved from the Big Bang epoch. It also provides powerful tools to study the timescale and mechanisms for the formation of the stellar populations and galaxies we see in the universe, as well as cosmic yardsticks to estimate distances and investigate the kinematical status of the universe.

Starting from the basics of stellar physics, in the following chapters we will build up a comprehensive picture of how stars form and evolve, how their physical and chemical properties change with time, and how one can take advantage of this knowledge to address broader questions related to galaxy formation and evolution and the estimate of the main cosmological parameters.