18 Cosmology

Very quickly after Einstein published his general theory, a number of researchers attempted to apply Einstein's equations to the universe as a whole. This was a natural, if quite radical, move. In Einstein's theory the distribution of energy and momentum in the universe determines the structure of space–time, and this applies as much to the universe as a whole as to the region of space, say, around a star. To get started, these early researchers made an assumption which, while logical, may seem a bit bizarre. They took the principles enunciated by Copernicus to their logical extreme and assumed that space–time was homogeneous and isotropic, i.e. that there is no special place or direction in the universe. They had virtually no evidence for this hypothesis at the time – definitive observations of galaxies outside of the Milky Way were only made a few years later. It was only decades later that evidence in support of this *cosmological principle* emerged. As we will discuss, we now know that the universe is extremely homogeneous when viewed on sufficiently large scales.

18.1 The cosmological principle and the FRW universe

To implement the principle, just as for the Schwarzschild solution we begin by writing the most general metric consistent with an assumed set of symmetries. In this case the symmetries are homogeneity and isotropy in space. A metric of this form is called a *Friedmann–Robertson–Walker* (FRW) metric. We can derive this metric by imagining our three-dimensional space, at any instant, as a surface in a four-dimensional space. There should be no preferred direction on this surface; in this way we impose both homogeneity and isotropy. The surface will then be one of constant curvature. Consider, first, the mathematics required to describe a $(2+1)$-dimensional space–time of this sort. The three *spatial* coordinates would satisfy

$$x_1^2 + x_2^2 = k\left(R^2 - x_3^2\right), \tag{18.1}$$

where whether $k = \pm 1$ is positive or negative depends on whether the space has positive or negative curvature. Then the line element on the surface is (for positive k):

$$d\vec{x}^2 = dx_1^2 + dx_2^2 + dx_3^2 = dx_1^2 + dx_2^2 + \frac{(x_1 dx_1 + x_2 dx_2)^2}{x_3^2}. \tag{18.2}$$

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press

The equation of the hypersurface gives

$$x_3^2 = R^2 - x_1^2 - x_2^2. \tag{18.3}$$

Setting $x_1 = r' \cos\theta$, $x_2 = r' \sin\theta$, we have

$$d\vec{x}^2 = \frac{R^2 dr'^2}{R^2 - r'^2} + r'^2 d\theta^2. \tag{18.4}$$

It is natural to rescale according to $r' = r/R$. Then the metric takes the form, now for general k,

$$d\vec{x}^2 = \frac{dr'^2}{1 - kr^2} + r^2 d\theta^2. \tag{18.5}$$

Here $k = 1$ for a space of positive curvature; $k = -1$ for a space of negative curvature; $k = 0$ is a spacial case, corresponding to a flat universe.

We can immediately generalize this to three dimensions by allowing the radius R to be a function of time, $R \to a(t)$. In this way we obtain the Friedmann–Robertson–Walker (FRW) metric:

$$ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2\right). \tag{18.6}$$

By general coordinate transformations this can be written in a number of other convenient and commonly used forms, which we will encounter in the following.

First we will evaluate the connection and the curvature (see Section 17.1). Again, the reader should evaluate a few of these terms by hand and perform the complete calculation using one of the programs mentioned in the exercises in the previous chapter. The non-vanishing components of the Christoffel connection are

$$\Gamma^i_{0j} = \frac{\dot{a}}{a}\delta^i_j, \quad \Gamma^0_{ij} = g_{ij}\frac{\dot{a}}{a}, \quad \Gamma^i_{jk} = \frac{g^{il}}{2}(g_{lj,k} + g_{lk,j} - g_{jk,l}) \tag{18.7}$$

and those of the curvature are

$$\mathcal{R}_{00} = -3\frac{\ddot{a}}{a}, \quad \mathcal{R}_{ij} = g_{ij}\left(\frac{\ddot{a}}{a} + 2H^2 - 2\frac{k}{a^2}\right). \tag{18.8}$$

Here H is known as the Hubble parameter,

$$H = \frac{\dot{a}}{a}, \tag{18.9}$$

and represents the expansion rate of the universe. Today

$$H = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}, \quad h = 0.73 \pm 0.03. \tag{18.10}$$

The assumptions of homogeneity and isotropy greatly restrict the form of the stress tensor: $T_{\mu\nu}$ must take the *perfect fluid* form

$$T_{00} = \rho, \quad T_{ij} = p g_{ij}, \tag{18.11}$$

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press

where ρ and p are the energy density and the pressure and are assumed to be functions only of time. Then the (0, 0) component of the Einstein equation (17.61) gives the *Friedmann equation*,

$$\frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \frac{8\pi G_{\rm N}}{3}\rho, \tag{18.12}$$

where $G_{\rm N}$ is Newton's gravitational constant (see Eq. (17.67)). The i, j components give:

$$\frac{2\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = -8\pi G_{\rm N}\rho. \tag{18.13}$$

There is also an equation which follows from the conservation of the energy momentum tensor, i.e. $T^{\mu\nu}_{\ \ ;\nu} = 0$. This is

$$d(\rho a^3) = -pd(a^3). \tag{18.14}$$

This equation is familiar in thermodynamics as the equation of energy conservation if we interpret a^3 as the volume of a system. Suppose that we have the equation of state $p = w\rho$, where w is a constant. Then Eq. (18.14) says that

$$\rho \propto a^{-3(1+w)}. \tag{18.15}$$

Three special cases are particularly interesting. For non-relativistic matter, the pressure is negligible compared with the energy density, so $w = 0$. For radiation (relativistic matter), $w = 1/3$. For a Lorentz-invariant stress tensor $T_{\mu\nu} = \Lambda g_{\mu\nu}$, we have $p = -\rho$ so $w = -1$. For these cases, it is worth remembering that

$$\text{radiation, } \rho \propto a^{-4}; \quad \text{matter, } \rho \propto a^{-3}; \quad \text{vacuum, } \quad \rho = \text{const.} \tag{18.16}$$

After taking account of the conservation of stress–energy and the Bianchi identities, only one of the two Einstein equations we have written down is independent; and it is conventional to take this as the Friedmann equation. This equation can be rewritten in terms of the Hubble parameter:

$$\frac{k}{H^2a^2} = \frac{8\pi G_{\rm N}\rho}{3H^2} - 1. \tag{18.17}$$

Examining the right-hand side of this equation, it is natural to define a *critical density*

$$\rho_{\rm c} = \frac{3H^2}{8\pi G_{\rm N}}, \tag{18.18}$$

and to define Ω as the ratio of the density and the critical density,

$$\Omega = \frac{\rho}{\rho_{\rm c}}. \tag{18.19}$$

So $k = 1$ corresponds to $\Omega > 1$, $k = -1$ to $\Omega < 1$ and $k = 0$, a flat universe, to $\Omega = 1$. It is also natural to break up Ω into various components, such as those due to radiation, matter or cosmological constant. As we will discuss shortly, Ω today is equal to unity within experimental errors; its main components are some unknown form of matter, baryons and dark energy (perhaps a cosmological constant):

$$\Omega_{\rm dm} = 0.267, \quad \Omega_{\rm b} = 0.049, \quad \Omega_{\rm de} = 0.683. \tag{18.20}$$

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press

The present error bars are of order 3% or less on these quantities (the most recent data is from the Planck satellite). Note that the total is close to unity. The present expansion rate is also known to be at the 2% level.

The history of the universe divides into various eras, in which different forms of energy were dominant. The earliest era for which we have direct observational evidence is a period lasting from a few seconds after the big bang to about 100 000 years, during which the universe was radiation dominated. From the Friedmann equation, setting $k = 0$, we have that

$$a(t) = a(t_0)t^{1/2}, \quad H = \frac{1}{2t}. \tag{18.21}$$

For the period of matter domination, which began about 10^5 years after the big bang and lasted almost to the present:

$$a(t) \propto t^{2/3}, \quad H = \frac{2}{3t}. \tag{18.22}$$

The universe appears today to be passing from an era of matter domination to a phase in which a (positive) cosmological constant dominates. Such a space is called a de Sitter space, with Hubble parameter M_d:

$$a(t) \propto e^{H_\mathrm{d}t}, \quad H_\mathrm{d} = \frac{8\pi G_\mathrm{N}}{3}\Lambda. \tag{18.23}$$

In the radiation-dominated and matter-dominated periods, H is, as we can see from the formulas above, roughly a measure of the age of the universe. One can define the age of the universe more formally as:

$$t = \int^{a(t)} \frac{da}{\dot{a}} = \int \frac{da}{aH}. \tag{18.24}$$

The present value of the Hubble constant corresponds to $t \approx 13.8$ billion years. To obtain this correspondence between the age and the measured H_0, it is important to include both the matter and the cosmological constant parts of the energy density. Note, in particular, that the integral is dominated by the most recent times, where H is smallest.

18.2 A history of the universe

As little as 50 years ago, most scientists would have been surprised at just how much we would eventually know about the universe: its present composition, its age and its history, back to times a couple of minutes after the big bang. We have direct evidence of phenomena at much earlier times, though the full implications of this evidence are difficult to interpret. We understand how galaxies formed and the abundance of the light elements. And we have a treasure trove of plausible speculations, some of which we should be able to test over time.

In this section we outline some basic features of this picture. Examining the FRW solution of Einstein's equations, we see that the scale factor $a(t)$ gets monotonically smaller

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press

in the past. The Hubble parameter H becomes larger. So, at some time in the past, the universe was much smaller than it is today. More precisely, the objects we see, or their predecessors, were far closer together. Far enough back in time, the material we currently see was highly compressed and hot. So, at some stage, it is likely that the universe was dominated by radiation. Recall that, during a radiation-dominated era,

$$a \sim t^{1/2}, \quad H = \frac{1}{2t}. \tag{18.25}$$

If we suppose that the universe remains radiation dominated as we look further back in time, we face a problem. At $t = 0$ the metric is singular – the curvature diverges. This is a finite time in the past, since

$$\int_0^{\text{today}} dt \sqrt{-g_{00}} \tag{18.26}$$

converges as $t \to 0$. This is not simply a feature of our particular assumptions about the equation of state or the precise form of the metric but a feature of solving Einstein's equations; it is a consequence of the singularity theorems due to Penrose and Hawking. The meaning of this singularity is a subject of much speculation. It might be smoothed out by quantum effects, or it might indicate something else. For now we simply have to accept that extremely early times are inaccessible to us. To start, we will suppose that at time t_0 the universe was extremely hot, with temperature T_0, and reasonably homogeneous and isotropic. We will then allow the universe to evolve, using Einstein's equations, the known particles and their interactions and the basic principles of statistical mechanics. As we will see, we can safely take T_0 to be at least as large as several MeV (corresponding to temperatures larger than 10^{10} K).

To make further progress we need to think about the content of the universe and how it evolves as the universe expands. The universe cannot be precisely in thermal equilibrium but, for much of its history, it has been very nearly so, with matter and radiation evolving adiabatically. To understand why the expansion is adiabatic, note first that H^{-1} is the time scale for the expansion. If the universe is radiation dominated,

$$H \sim \frac{T^2}{M_{\rm p}}, \tag{18.27}$$

where $M_{\rm p}$ is the Planck mass. The rate for interactions in a gas will scale with T, multiplied perhaps by a few powers of coupling constants. For temperatures well below the Planck scale the reaction rates will be much more rapid than the expansion rate. So, at any given instant, the system will nearly be in equilibrium.

It is worth reviewing a few formulas from statistical mechanics. These formulas can be derived by elementary considerations or by using the methods of finite-temperature field theory, as discussed in Appendix C. For a relativistic weakly coupled Bose gas,

$$\rho = \frac{\pi^2}{30} g T^4, \quad p = \frac{\rho}{3}, \tag{18.28}$$

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press

while, for a similar Fermi gas,

$$\rho = \frac{7}{8}\frac{\pi^2}{30}gT^4, \quad p = \frac{\rho}{3}. \tag{18.29}$$

Here g is a degeneracy factor that counts the number of physical helicity states of each particle type. In the non-relativistic limit, for both bosons and fermions we have

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} \exp\left[-\frac{(m-\mu)}{T}\right] \tag{18.30}$$

$$\rho = mn, \quad p = nT \ll \rho. \tag{18.31}$$

For temperatures well below m, the density rapidly goes to zero unless $\mu \neq 0$. Note that μ may be non-zero when there is a (possibly approximately) conserved quantum number. Perhaps the most notable example is the baryon number.

We should pause here and discuss an aspect of general relativity which we have not considered up to now. A gravitational field alters the behavior of clocks. This is known as the gravitational red shift. We can understand this in a variety of ways. First, if we have a particle at rest in a gravitational field then the proper time is related to the coordinate time by a factor $\sqrt{g_{00}}$. Consider, alternatively, the equation for a massless scalar field with momentum k in an expanding FRW universe. This is just $D^\mu \partial_\mu \phi = 0$. Using the non-vanishing Christoffel symbols, with $\phi(\vec{x}, t) = e^{i\vec{k}\cdot\vec{x}}\phi(t)$,

$$\ddot{\phi}(k) + 3H\dot{\phi}(k) + \frac{k^2}{a^2(t)}\phi = 0. \tag{18.32}$$

As a result of the last term, the wavelength effectively increases as $a(t)$. A photon red-shifts in precisely the same way.

The implications of this for the statistical mechanical distribution functions are interesting. Consider, first, a massless particle such as the photon. For such a particle, the distribution is

$$\int \frac{d^3k}{(2\pi)^3} \frac{1}{e^{k/T} - 1}. \tag{18.33}$$

The effect of the red shift is to maintain this form of distribution but to change the temperature, $T(t) \propto 1/a(t)$. So even if the particles are not in equilibrium, they maintain an equilibrium distribution appropriate to the red-shifted temperature. This is not the case for massive particles.

Let us imagine, then, starting the clock when the universe is at temperatures well above the scale of QCD but well below the scale of weak interactions, say at 10 GeV. In this regime the density of Ws and Zs is negligible, but the quarks and gluons behave as nearly free particles. So we can take an inventory of the bosons and fermions that are light compared with T. The bosons include the photon and the gluons; the fermions include all the quarks and leptons except the top quark. So the effective g, which we might call g_{10}, is approximately 98. This means, for example, that

$$\rho \approx \frac{g_{10}\pi^2}{30}T^4 \tag{18.34}$$

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press

and the Hubble constant is related to the temperature through

$$H = \left(\frac{8\pi}{3} G_N \frac{\pi^2}{30} g_{10} T^4 \right)^{1/2}, \tag{18.35}$$

where G_N is Newton's gravitational constant (see Eq. (17.67)). This allows us to write a precise formula for the temperature as a function of time:

$$T = \left(\frac{16\pi}{3} G_N \frac{\pi^2}{30} g_{10} \right)^{-1/4} \left(\frac{1}{t} \right)^{1/2}. \tag{18.36}$$

As the universe cools, QCD changes from a phase of nearly free quarks and gluons to a hadronic phase. At temperatures below m_π, the only light species are the electron and the neutrinos. By this time, the antineutrons have annihilated with neutrons and the antiprotons with protons, leaving a small net baryon number, the total number of neutrons and protons. There is, at this time, of order one baryon per billion photons. We will have much more to say about this slight excess later.

At this stage, interactions involving neutrinos maintain an equilibrium distribution of protons and neutrons. We can give a crude, but reasonably accurate, estimate of the temperature at which neutrino interactions drop out of equilibrium by asking when the interaction rate becomes comparable to the expansion rate. The cross section for neutrino–proton interactions is

$$\sigma(\nu + p \to n + e) \approx G_F^2 E^2, \tag{18.37}$$

where G_F is the Fermi constant (see Eq. (3.3)), and the number density of neutrinos is

$$n_\nu \approx \frac{\pi^2}{30} g_T T^3. \tag{18.38}$$

Combining this with our formula Eq. (18.35) for the Hubble constant as a function of T gives, for the decoupling temperature T_ν,

$$T_\nu^3 \approx G_F^{-2} M_p^{-1} \tag{18.39}$$

or

$$T_\nu \approx 2 \text{ MeV}. \tag{18.40}$$

This corresponds to a time of order 100 s after the big bang. At this point neutron decays are not compensated by the inverse reaction, but many neutrons will pair with protons to form stable light elements such as D and He. At about this time the abundances of the various light elements are fixed.

There is a long history of careful, detailed, calculations of the abundances of the light elements (H, He, D, Li, . . .) which result from this period of decoupling. The abundances turn out to be a sensitive function of the ratio of baryon and photons, n_B/n_γ. Astronomers have also made extensive efforts to *measure* this ratio. A comparison of observations and measurements gives, for the baryon to photon ratio,

$$\frac{n_B}{n_\gamma} = 6.1^{+0.3}_{-0.2} \times 10^{-10}. \tag{18.41}$$

We will see that this result receives strong corroboration from other sources.

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press

The universe continues to cool in this radiation-dominated phase for a long time. At $t \approx 10^5$ years the temperature drops to about 1 eV. At this time electrons and nuclei can combine to form neutral atoms. As the density of ionized material drops, the universe becomes essentially transparent to photons. This is referred to as *recombination.* The photons now stream freely. As the universe continues to cool the photons red-shift, maintaining a Planck spectrum. Today, these photons behave as if they had a temperature $T \approx 3\,\mathrm{K}$. They constitute the *cosmic microwave background radiation* (CMBR). This radiation was first observed in 1963 by Penzias and Wilson and has since been extensively studied. It is very precisely a black body, with characteristic temperature 2.7 K. We will discuss other features of this radiation shortly.

It is interesting that, given the measured value of the matter density, matter and radiation have comparable energy densities at the recombination stage. At later times matter dominates the energy density, and this continues to be the case to the present time.

In our brief history, another important event occurs at $t \simeq 10^9$ years. If we suppose that initially there were small inhomogeneities, these remain essentially frozen, as we will explain later, until the time of matter–radiation equality. They then grow with time. From observations of the CMBR we know that these inhomogeneities were at the level of one part in 10^5. At about 1 billion years after the big bang, these then grow enough to be non-linear, and their subsequent evolution is believed to give rise to the structure – galaxies, clusters of galaxies, and so on – that we see around us.

One surprising feature of the universe is that most of the energy density is in two forms which we cannot see directly. One is referred to as the *dark matter*. The possibility of dark matter was first noted by astronomers in the 1930s, from observations of the rotation curves of galaxies. Simply using Newton's laws one can calculate the expected rotational velocities and one finds that these do not agree with the observed distribution of stars and dust in the galaxies. This is true for structures on many scales, not only galaxies but clusters and larger structures. Other features of the evolution of the universe are not in agreement with observation unless most of the energy density is in some other form. From a variety of measurements, Ω_{m}, the fraction of the critical energy density (see Eq. (18.18)) in matter, is known to be about 0.3. Nucleosynthesis and the CMBR give a much smaller fraction in baryons, $\Omega_{\mathrm{b}} \approx 0.05$. In support of this picture, direct searches for hidden baryons give results that are compatible with the smaller number; they have failed to find anything like the required density to give Ω_{m}.

Finally, it appears that we are now entering a new era in the history of the universe. For the last 14 billion years, the energy density has been dominated by non-relativistic matter. But, at the present time, there is almost twice as much energy in some new form, with $p < 0$, referred to as dark energy. The dark energy is quite possibly a cosmological constant, Λ. Current measurements are compatible with $w = -1$ ($p = -\rho$).

The picture we have described has extensive observational support. We have indicated some of this: the light element abundances and the observation of the CMBR. The agreement of these two quite different sets of observations for the baryon to photon ratio is extremely impressive. Observations of supernovae, the age of the universe and features of structure at different scales all support the existence of a cosmological constant (dark energy) constituting about 70% of the total energy.

This is not a book on cosmology, and the overview we have presented is admittedly sketchy; there are many aspects of this picture we have not discussed. Fortunately there are many excellent books on the subject, some of which are mentioned in the suggested reading.

Suggested reading

There are a number of good books and lectures on the aspects of cosmology discussed here. Apart from the text of Weinberg (1972), mentioned earlier, these include the texts of Kolb and Turner (1990), Dodelson (2004) and Weinberg (2008).

Exercises

(1) Compute the Christoffel symbols and the curvature for the FRW metric. Verify the Friedmann equations.
(2) Verify Eq. (18.32).
(3) Consider the case of de Sitter space, $T_{\mu\nu} = -\Lambda g_{\mu\nu}$ with positive Λ. Show that the space expands exponentially rapidly. Compute the horizon, i.e. the largest distance from which light can travel to an observer.

https://doi.org/10.1017/9781009290883.024 Published online by Cambridge University Press