# 9 Successes and Problems with the Hot Big Bang Model

## 9.1 Primordial Nucleosynthesis

When the temperature of the universe is $T$, the typical energy of a particle is $E_{\rm part} \approx k_B T$. From the temperature–time relation (8.35), we can convert this to a timescale

$$\frac{t}{\rm sec} \approx \left(\frac{T}{1.5 \times 10^{10}\,{\rm K}}\right)^{-2} \approx \left(\frac{k_B T}{2\,{\rm MeV}}\right) \tag{9.1}$$

for times before the epoch of matter–radiation equality, i.e., when the universe is dominated by relativistic particles and radiation. Thus, when the universe was about 1 second old, the typical particle energies were about 2 MeV.

Now, 1 MeV corresponds to the nuclear binding energy andis roughly the energy it takes to rip apart an atomic nucleus. As a result, above this energy scale – that is, above a temperature of $10^{10}$ K or before one second had elapsed since the Big Bang – atomic nuclei could not have existed. The photons would have been so energetic that if a nucleus were to form, the photons would have very rapidly broken it apart when they collided with it. (This is analogous to the process that prevented the formation of stable atoms in the universe prior to the epoch of decoupling – see Section 8.3).

Thus, in the first second of the universe's history, the universe comprised a sea of photons, neutrinos electrons and baryons (neutrons and protons). Conversely, below 1 MeV, the energy of the photons had reduced sufficiently on average to allow stable nuclei to form. The process of nuclei formation in the early universe is known as *primordial nucleosynthesis*.

The nuclear physics is well understood and for our discussion we need three important pieces of information:

• Neutrons are more massive than protons (see Appendix A) and the difference in mass–energy is

$$(m_n - m_p)\,c^2 = (938.3 - 939.6)\,{\rm MeV} = 1.3\,{\rm MeV} \tag{9.2}$$

• As a result, a free neutron will decay into a proton. This process is radioactive, with a half–life of $t_{1/2} = 940$ s.

• Once bound in nuclei, neutrons are stable.

The decay of a free neutron, $n^o$, into a proton, $p^+$, also produces an electron, $e^-$, to conserve electric charge, together with a neutrino particle, $\nu$. Some pure energy, in the form of a gamma–ray photon, with an energy 0.8 MeV, is also released (in effect the excess mass of the neutron is converted into pure energy via Einstein's relation $E = mc^2$). The whole process can be expressed as

$$n^o \longrightarrow p^+ + e^- + \bar{\nu} + 0.8{\rm MeV} \tag{9.3}$$

where $\bar{\nu}$ represents an anti-neutrino.

At very high temperatures ($T \gg 1\,\text{MeV}$), the reverse reaction may also proceed. The proton can transform into a neutron by absorbing the energy of the gamma–ray photon. This process is made possible at high enough temperatures because the photon energy is on average sufficiently large to allow its energy to be converted into mass when the photons and protons collide. Consequently, the neutrons and protons attain thermal equilibrium. If too many protons are present, say, then more of them become converted into neutrons, and vice–versa.

Let us consider a time when the temperature was higher than 1 MeV, so that nuclei have yet to form, but low enough so that the baryons have become non–relativistic, i.e., $k_B T \ll m_p c^2 \approx 939\,\text{MeV}$. When a non–relativistic particle species is in equilibrium, its number density, $N$, is determined by the particle's mass and the temperature of the universe. The relationship is known as the Maxwell–Boltzmann distribution and we quote the result:

$$N \propto m^{3/2} \exp\left(-\frac{mc^2}{k_B T}\right) \tag{9.4}$$

Thus, the number of neutrons relative to the number of protons is given by

$$\frac{N_n}{N_p} = \left(\frac{m_n}{m_p}\right)^{3/2} \exp\left[-\frac{(m_n - m_p)\,c^2}{k_B T}\right] \tag{9.5}$$

It follows that the number densities are approximately equal, $N_n \approx N_p$, at temperature in excess of the mass difference, $k_B T \gg (m_n - m_p)c^2 \approx 1.3\text{MeV}$, because the exponent in Eq. (9.5) becomes very small and the exponential factor itself is of the order unity.

Once the temperature drops below the critical value of 1 MeV, the reverse reaction where a proton is converted into a neutron becomes progressively more unlikely, because there are fewer photons in the universe with a sufficiently high energy. Thus, the number of neutrons in the universe decreases relative to the number of protons. Indeed, if the photon energy is below 0.8 MeV, then the reverse reaction is no longer possible and the neutrons and protons are said to fall out of equilibrium. By the time this occurs, the number density of neutrons is about one fifth that of the protons:

$$\frac{N_n}{N_p} \approx \exp\left(-\frac{1.3}{0.8}\right) \approx \frac{1}{5} \tag{9.6}$$

It is important to emphasize that the number density of neutrons is fewer than that of the protons due to the slightly larger mass of the former.

The production of the light elements can now proceed. In general, this occurs through a complex chain of nuclear reactions [20] Most of the neutrons end up forming

[20] Here we denote a nucleus, $X$, containing $p$ protons and $n$ neutrons as $X^{p+q}(p, n)$.

helium–4 nuclei, which consists of two neutrons and two protons, $\mathrm{He}^4(2,2)$. The remaining protons ultimately form hydrogen nuclei, $\mathrm{H}^1(1,0)$. A small amount of deuterium, $D^2(1,1)$, helium–3, $\mathrm{He}^3(2,1)$, and lithium–7, $\mathrm{Li}^7(3,4)$, are also produced.

In what follows we can neglect to a good approximation these additional nuclei and assume that all of the neutrons end up in helium–4 nuclei. Now, the formation of helium–4 takes time – it does not happen instantaneously. A key intermediate reaction in the formation of helium–4 is the synthesis of deuterium from a neutron and proton collision:

$$p^+ + n^0 \quad \longrightarrow \quad \mathrm{D}^2 + \gamma \tag{9.7}$$

where $\gamma$ denotes a photon that carries away excess energy. Helium-4 is then produced when two deuterium nuclei fuse:

$$\mathrm{D}^2 + \mathrm{D}^2 \quad \longrightarrow \quad \mathrm{He}^4 + \gamma \tag{9.8}$$

The production of helium–4 is delayed because deuterium is a very fragile nucleus. Its binding energy is only 0.1 MeV and so stable deuterium nuclei can not form until the temperature has fallen to this critical value. Consequently, the production of helium–4 only starts when the universe has cooled to $k_BT \approx 0.1\,\mathrm{MeV}$. From the temperature–time relation (8.35), the age of the universe at this temperature was about $t_{\mathrm{nuc}} \approx 400\,\mathrm{s}$. Note that this is comparable to the neutron's half–life, $t_{1/2} = 940\,\mathrm{s}$. Consequently, the number of neutrons will have dropped further from the relative value quoted in Eq. (9.6).

We can estimate this by recalling that if initially there are $N_{\mathrm{init}}$ particles then after a time $\Delta t$, the number of particles remaining after radioactive decay with a half–life $t_{1/2}$ is given by

$$N = N_{\mathrm{init}} \exp\left[-\left(\frac{\Delta t}{t_{1/2}}\right) \times \ln 2\right] \tag{9.9}$$

Here, we are interested in the time that elapses between the neutrons and protons falling out of equilibrium and the formation of stable helium–4. Hence, we consider $\Delta t \approx 400$, i.e., we take the initial number density of neutrons at the time when $k_BT = 0.8\,\mathrm{MeV}$. Thus, dividing through by the number of protons[21]

$$\begin{aligned}\left(\frac{N_n}{N_p}\right) &= \left(\frac{N_n}{N_p}\right)_{0.8\,\mathrm{MeV}} \exp\left(-\frac{t_{\mathrm{nuc}} \times \ln 2}{t_{1/2}}\right) \\ &\approx \frac{1}{5} \times \exp\left(-\frac{400 \times \ln 2}{940}\right) \approx \frac{1}{7}\end{aligned} \tag{9.10}$$

That is, by the time helium–4 nuclei can form there is 1 free neutron for every 7 protons.

Now, the mass of the helium–4 nuclei is $M_{\mathrm{He}} \approx 4m_n$ (neglecting the small mass difference between the neutrons and the protons). Furthermore, the number density

[21] We assume this number stays constant; it is only a small correction.

| | Nucleosynthesis | Decoupling |
|---|---|---|
| Time | Few minutes | 300,000 yrs |
| Temperature | $10^{10}$ K | 3000 K |
| Typical Energy of Particles | 1 MeV | 1 eV |
| Physical Process | $p^+$ & $n^o$ form nuclei. | Nuclei & $e^-$ |
| | $e^-$ remain free | form atoms |
| Radiation | Interacts with nuclei & $e^-$ | No interaction with matter |

Table (9.1): Comparing the similarities and differences between primordial nucleosynthesis and the decoupling of matter and radiation.

of helium–4 nuclei is half that of the number density of neutrons at that time, since two neutrons go into each helium–4 nucleus, i.e., $N_{\mathrm{He}} = N_n/2$. The fraction of mass in helium–4 nuclei is defined by

$$Y_4 \equiv \frac{4N_{\mathrm{He}}}{N_n + N_p} \tag{9.11}$$

and substituting and rearranging we deduce that

$$Y_4 = \frac{4\,(N_n/2)}{N_n + N_p} = \frac{2}{1 + \frac{N_p}{N_n}} \approx \frac{2}{1+7} = 0.25 \tag{9.12}$$

In other words, the theoretical prediction is that *the fraction in mass in the universe that forms into helium–4 nuclei is about 25 %*. It is important to realize that this is a key theoretical prediction of the big bang model.

The crucial point about helium–4 nuclei is that they are very stable. It is takes a great deal of energy to attach further neutrons or protons to them to form heavier nuclei, such as oxygen and carbon. By the time helium–4 nuclei had formed in the early universe the temperature had dropped below that required to produce heavier elements. Consequently, the carbon and oxygen so crucial for life in the universe could *not* have formed in the big bang. (They are actually formed during supernovae explosions). Moreover, because helium–4 nuclei are stable, it is very difficult to break them apart. This implies that they should have survived through to the present epoch. So, by looking in regions of the universe where there are stars are rare, we can compare theory with observation to test the viability of the big bang model[22]. Good agreement between theory and observation would indicate that our understanding of the physics of the universe when it was just one second old is reliable.

Remarkably, the helium abundance can be measured, as can the abundances of the other light elements such as deuterium and lithium–7. For example, the deuterium

[22]As helium is a by-product of the main sequence phase of a stellar lifetime, the primordial abundance of helium becomes polluted by stellar material released during supernovae explosions. It is necessary to measure helium abundance in regions of the universe where stellar populations are negligible.

abundance is measured by observing the absorption of quasar light as it passes through a cloud of primordial gas located between us and the quasar – more light is absorbed if the abundance of deuterium is higher.

The comparison between theory and observation is shown in Fig. (9.1). The broad bands represent the observed abundances and the curved lines the theoretical predictions. The density of baryons, $\Omega_B$, at the present epoch, in units of the critical density, is plotted on the horizontal axis. As might be expected, the theoretical predictions depend quite sensitively on the density of baryons. For example, the deuterium abundance falls rapidly with increasing baryon density, because a higher density implies that collisions between the baryons at the epoch of nucleosynthesis are more likely and so more deuterium is converted into helium.

Remarkably there is a greement between theory and observation if the baryon density is in the rather narrow range

$$\Omega_B h^2 = 0.013 \pm 0.002 \tag{9.13}$$

In particular, note how the helium–4 prediction is close to what is observed. This agreement is another major success of the big bang model. It implies that we have a good understanding of the universe when it was just a few minutes old. It also has a major implication in that *baryonic mattre can contribute at most about 5 % of the critical density of the universe. This implies that most of the matter in the universe must be non–baryonic.*

Table (9.1) compares the similarities and differences between the physics of primordial nucleosynthesis and that of the decoupling of matter and radiation.

## 9.2 Problems with the Big Bang Model

### 9.2.1 Flatness Problem

To appreciate the so–called 'flatness' problem, let us recall the Friedmann equation expressed in terms of the $\Omega$–parameter, Eq. (4.4):

$$\Omega - 1 = \frac{kc^2}{a^2 H^2} \tag{9.14}$$

Taking the modulus of both sides implies

$$|\Omega - 1| = \frac{1}{\dot{a}^2} \tag{9.15}$$

if $k \neq 0$.

Now, when we discussed the acceleration equation (2.23) for a universe dominated by either pressureless matter or radiation (or both), we saw that $\ddot{a} < 0$ is always satisfied, because the density and pressure of the material are both non–negative quantities (see Section 2.4). This implies that the expansion of the universe is slowing

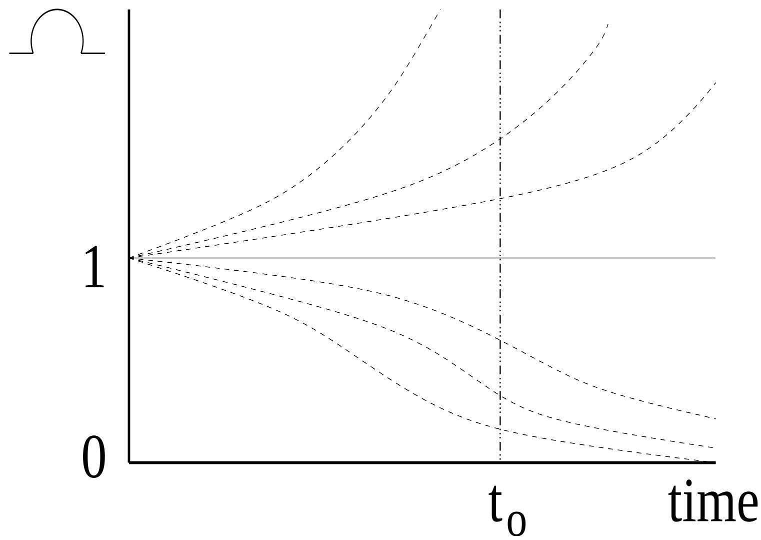


Fig. (9.2). Showing how the $\Omega$–parameter moves away from unity in an open or closed universe. Note that if $\Omega = 1$ precisely then it remains fixed at this value for all time, since $k = 0$ in this case. This represents a critical value. At any given earlier time $\Omega$ is closer to unity than it is at present (assuming of course that it is different to unity).

down with time and, consequently, that $\dot{a}$ is *decreasing* as time proceeds. It follows from Eq. (9.15), therefore, that the quantity $|\Omega - 1|$ *increases* with time. In other words, $\Omega$ moves progressively further away from unity as the universe expands.

Conversely, as we trace the evolution of the $\Omega$–parameter back in time, it gets progressively closer to unity, since the expansion rate is higher at earlier times. This behaviour is shown qualitatively in Fig. (9.2). Another way to see this is to recall that the density of radiation falls as $\rho_{\rm rad} \propto 1/a^4$, whereas that of pressureless matter falls as $\rho_{\rm mat} \propto 1/a^3$. Both of these decrease more rapidly than the curvature term $k/a^2$ that appears in the Friedmann equation (2.27).

Observations indicate that $\Omega_0 = \mathcal{O}(1)$ at the present epoch – that is, the universe is close to being spatially flat. However, the above discussion implies that if $|\Omega - 1|$ is small today, it must have been *much* smaller at earlier times. The question is: how close to unity was the $\Omega$–parameter when the universe was very young?

We have seen that we can be reasonably confident that we understand the state of the universe at the epoch of primordial nucleosynthesis, $t_{\rm nuc} \approx 1\,{\rm s}$. What was the value of $\Omega$ at that time, i.e., how close was the density to the critical density? Suppose the universe has always been radiation dominated, $a \propto t^{1/2}$ (the correction for accounting for the fact that universe became matter dominated after the epoch of matter–radiation equality turns out to be negligible). Then it follows from Eq. (9.15) that

$$\frac{|\Omega - 1|_0}{|\Omega - 1|_{\rm nuc}} = \frac{\dot{a}^2_{\rm nuc}}{\dot{a}^2_0} = \frac{t_0}{t_{\rm nuc}} \tag{9.16}$$

where a subscript ‘nuc’ denotes values are to be evaluated at the epoch of nucleosynthesis. By taking the values $t_0 \approx 10^{17}\,$s and $t_{\rm nuc} \approx 1\,$sec for the age of the universe and onset of nucleosynthesis, respectively, and further assuming that $|\Omega - 1|_0 \approx 1$, we deduce that when the universe was 1 second old, the $\Omega$–parameter satisfied the constraint

$$|\Omega - 1|_{\rm nuc} \approx 10^{-17} \tag{9.17}$$

or, equivalently,

$$0.99999999999999999 \leq \Omega_{\rm nuc} \leq 1.00000000000000001 \tag{9.18}$$

In other words, the density of the universe was equal to the critical density to within 1 part in $10^{17}$. To have a value for the $\Omega$–parameter that is of order unity today implies that $\Omega$ had to have been exceptionally close to unity at the time of nucleosynthesis. (On a positive note, this does indicate that the approximation that the early universe was effectively flat is a very good one.)

It is important to emphasize that this conclusion is largely independent of the particular values we take for the present age of the universe and current value of the density. The key point is that if $\Omega_0$ is reasonably close to unity today, it must have been extremely close to unity in the early universe. The flatness problem, then, is the problem of understanding how such a precarious balance was achieved: *Why is the density of the universe so still close to the critical density ten billion years after the big bang?*.

### 9.2.2 Horizon Problem

It is not possible to communicate with a distant observer faster than the speed of light, since nothing can travel faster than light. Since the age of the universe is finite, there is a maximum distance that light could have travelled since the big bang. Consequently, for any given observer in the universe, such as ourselves, there is a maximal distance beyond which we can not observe. This is called the *horizon distance*. We saw in Section 6.1 that a light pulse must follow a radial trajectory ($d\theta = d\phi = 0$) and that the Robertson–Walker metric $ds^2$ must vanish. Hence, the comoving distance, $\chi$, is given by

$$a(t)d\chi = cdt \tag{9.19}$$

where we have taken the positive root of Eq. (6.1) since we are now interested in how far light could have travelled away from us since the big bang.

In general, the comoving distance travelled by a light pulse between a time $t_1$ and time $t_2$ is deduced by integrating Eq. (9.19):

$$\chi = c\int_{t_1}^{t_2} dt/a(t) \tag{9.20}$$

and the physical (proper) distance is then

$$D_P(t) = a(t)\chi \tag{9.21}$$

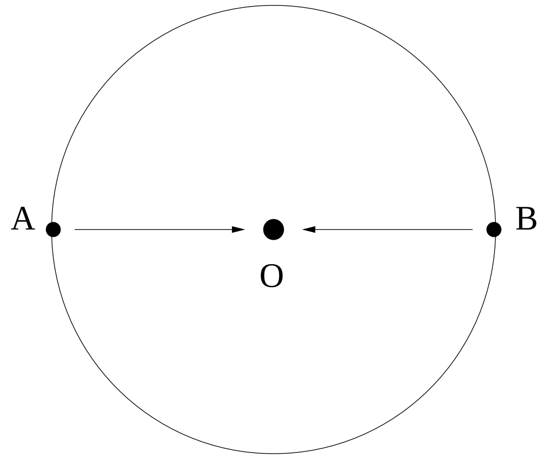


Fig. (9.3). Consider two regions $A$ and $B$ that are in opposite directions when viewed by us located at point $O$. If the radiation that we observe today as the CMB left these two regions at the epoch of decoupling, then the distance separating $A$ and $B$ from $O$ is $3ct_0$ and the distance between $A$ and $B$ is therefore $6ct_0$.

Thus, at a given time $t$, the distance a light pulse could travel since the origin of the universe at $t = 0$ is

$$d_H(t) = a(t) \int_0^t \frac{c}{a(t)} dt \tag{9.22}$$

This is the horizon distance.

Our horizon distance at the present epoch is evaluated by noting that the universe has been dominated by pressureless matter for most of its history, so $a \propto t^{2/3}$. Eq. (9.22) then implies that

$$d_H(t_0) = 3ct_0 \tag{9.23}$$

The factor of 3 arises due to the expansion of the universe. In a static universe, the distance is just $ct_0$, as expected.

At any given earlier time, the horizon distance may be viewed as a real physical distance between two points in the universe. This distance then scales with the scale factor from that time to the present. Specifically it grows by a factor $a_0/a(t)$. We may therefore define a rescaled horizon distance, $\tilde{d}_H(t)$, that represents the horizon distance at some earlier time rescaled up to the present day scale factor:

$$\tilde{d}_H(t) = \frac{a_0}{a(t)} d_H(t) = a_0 \int_0^t \frac{c}{a(t)} dt \tag{9.24}$$

The point of introducing Eq. (9.24) is that if two regions in the universe today are separated by more than this distance $\tilde{d}_H$, then they would have been separated by more than the horizon distance at that earlier time, $t$. Consequently, they would not have been able to communicate with each other at that time and, since they would have been isolated from one other, we would expect the physical conditions in those two regions to have been different. In particular, we would not expect the two regions to have the same temperature.

On the other hand, we have seen that the temperature of the CMB appears highly isotropic, to within 1 part in $10^5$. This indicates that it was in thermal equilibrium when it was formed at the epoch of decoupling, $t_{\rm dec}$, and moreover, implies that all the radiation that we observe today should have been within the horizon at the decoupling epoch.

To explore this further, consider the radiation that is travelling towards us from opposite regions of the universe (see Fig. (9.3)). Radiation from each of the two regions $A$ and $B$ has travelled a distance $d_H(t_0) = 3ct_0$ to reach us today, so the distance separating $A$ and $B$ is $6ct_0$. However, the horizon distance at the epoch of decoupling – rescaled to today's size – is given by

$$\tilde{d}_H(t_{\rm dec}) = ca_0 \int_0^{t_{\rm dec}} dt \frac{t_{\rm dec}^{2/3}}{a_{\rm dec} t^{2/3}} = 3ct_0 \left(\frac{t_{\rm dec}}{t_0}\right)^{1/3} \approx 0.1ct_0 < 6ct_0 \qquad (9.25)$$

since $a(t) = a_{\rm dec}(t/t_{\rm dec})^{2/3}$ and $t_{\rm dec} \approx 3 \times 10^{-5} t_0$ from Eq. (8.39).

This is less than the present distance between $A$ and $B$ and we deduce, therefore, that these two regions of the universe were isolated from each other when the matter and radiation decoupled. *The horizon problem is then the problem of understanding why it is that the temperature of the cosmic microwave radiation in these separate regions was effectively the same to within* $10^{-3}\%$.

### 9.2.3 Relic Particles

Grand Unified Theories (GUTs) of particle physics generically predict the existence of very massive, stable particles. These exotic particles come in many forms, depending on the theory in question. Generally, they arise as by products in the symmetry breaking process that causes the strong nuclear force to split away from the grand unified force. This occurred when the universe was just $10^{-35}$ seconds old. The specific properties of these particles are not important to us because we are only interested here in their cosmological consequences. What is important is their very high mass and the fact that they are stable. Here we focus on one such particle known as the magnetic monopole, although the discussion to follow is generic.

The mass of the magnetic monopole is $m_{\rm mono}c^2 \approx 10^{16}\,$GeV. (Compare this with the mass of the proton, $m_pc^2 \approx 1\,$GeV. Since these particles are so massive, their mass very rapidly comes to dominate their energy even though the temperature of the universe is so high. Recall that a massive particle becomes non–relativistic when the temperature of the universe falls to $k_BT \approx mc^2$. For the monopole this occurs when the temperature $T \approx 10^{28}\,$K. From the temperature–time relation (8.35), it follows that this is comparable to the temperature of the universe when the monopoles form. In other words, monopoles effectively behave as non–relativistic particles almost immediately after they have formed.

Now, this is potentially a disaster for the big bang model, because as we discussed in Section 7.2, the density of non–relativistic particles falls as $1/a^3$, whereas that of the

| Success | Problem |
|---|---|
| Explains Hubble Expansion | Flatness Problem |
| Origin of Microwave background | Horizon Problem |
| Primordial Nucleosynthesis | Relic Particle Problem |

Table (9.2): The successes and problems with the hot big bang model.

radiation and relativistic particles falls away more rapidly as $1/a^4$. Consequently, the monopoles will come to dominate the relativistic particles if sufficient time elapses. If they are formed in sufficient numbers, it is possible that they could come to dominate the universe before the epoch of primordial nucleosynthesis. In this case, the synthesis of helium–4 into hydrogen could not proceed because collisions between neutrons and protons would be too rare. On the other hand, we have seen in Section 9.1 that the agreement between the predicted and observed mass fraction of helium–4 is very good and this implies that primordial nucleosynthesis did indeed proceed according to our understanding of the universe when it was a few minutes old.

Effectively, this imposes an important constraint on the number density of monopoles when they form. The number that form must be sufficiently low that they do not dominate the density of the universe before one second (the epoch when neutrons and protons fall out of equilibrium). Question 7 of Exercise Sheet 7 asks you to calculate when monopoles are expected to dominate according to many popular particle physics theories – typically it is much earlier than the epoch of nucleosynthesis.

The conclusion is that either the particle physics theory is wrong or some mechanism for diluting the monopole density, i.e., increasing the volume of the universe, must be found.

Table (9.2) summarizes the three main success and problems with the hot big bang cosmological model.