Inflation and the accelerating universe

11

- Einstein introduced the cosmological constant in his field equation so as to obtain a static universe solution.
- The cosmological constant is the vacuum energy of the universe: This constant energy density corresponds to a negative pressure, giving rise to a repulsive force that increases with distance. A vacuum-energy dominated universe expands exponentially.
- The inflationary theory of cosmic origin—the universe had experienced a huge expansion at the earliest moment of the big bang—can provide the correct initial conditions for the standard FLRW model of cosmology: solving the flatness, horizon problems, and providing an origin of matter/energy, as well as giving just the right kind of density perturbation for subsequent structure formation.
- The primordial inflation leaves behind a flat universe, which can be compatible with the observed matter density being less than the critical density and a cosmic age greater than 9 Gyr if there remains a small but nonvanishing cosmological constant—a dark energy. This would imply a universe now undergoing an accelerating expansion.
- The measurement of supernovae at high redshift provided direct evidence for an accelerating universe. Such data, together with other observational results, especially the anisotropy of the cosmic microwave background and large structure surveys, gave rise to a concordant cosmological picture of a spatially flat universe $\Omega = \Omega_\Lambda + \Omega_M = 1$, dominated by dark energy $\Omega_\Lambda \simeq 0.75$. Most of the matter $\Omega_M \simeq 0.25$ is exotic dark matter $\Omega_{DM} \simeq 0.21$, compared to the ordinary (baryonic) matter $\Omega_B \simeq 0.04$ The cosmic age $t_0 \simeq 14\,\text{Gyr}$ comes out to be close to the Hubble time.
- The cosmological constant and the cosmic coincidence problems point to the need for new fundamental physics.

As we have discussed in Sections 9.1.3 and 10.2, Newton's and the original Einstein's equations would lead us to expect the expansion of the universe to slow down because of gravitational attraction. In this chapter, we shall see how a modification of the Einstein equation, with the introduction of the **cosmological constant** Λ, allows for the possibility of a gravitational repulsive force that increases with distance. This effect was first discovered by Einstein in his effort of seeking a static solution to the GR field equation. It also allows for the possibility that the universe had undergone an extraordinarily rapid expansion at an early moment (the inflationary epoch). The inflationary scenario of the big bang brings about just the correct initial conditions for the then standard cosmology (the FLRW model of Box.11.1) and predicts a flat geometry for the universe at large. Finally, a nonvanishing Λ term can account for the recently discovered evidence of an accelerating universe in the present epoch. An accelerating expansion means slower expansion in the past, hence a longer age for the naively expected decelerating universe—long enough to account for the oldest objects observed in the universe. The cosmological constant also provides us with a dark energy that, together with the observed matter content, fulfills the inflationary cosmology's prediction of a flat universe, which requires the mass/energy density of the universe to be equal to the critical density.

11.1 The cosmological constant

Before Hubble's discovery in 1929 of an expanding universe, just about everyone, Einstein included, believed that we lived in a static universe. Recall that the then-observed universe consisted essentially of stars within the Milky Way galaxy. But gravity, whether nonrelativistic or relativistic, is a universal attraction. Hence, theoretically speaking, a static universe is an impossibility. Specifically, as we have demonstrated, the Friedmann cosmological equations (10.1) and (10.2) have solutions corresponding always to a **dynamic** universe—a universe which is either contracting or expanding. Namely, these equations are not compatible with the static condition of an unchanging scale factor $\dot{a} = \ddot{a} = 0$, which would lead to a trivial empty universe,[1] $\rho = p = 0$.

[1]For the Einstein equation without a cosmological constant, a static solution necessarily corresponds to an empty universe. On the other hand, an empty universe is compatible with an expanding universe with negative spatial curvature. See Problem 10.4.

Λ as a modification of the geometry side Recall our brief discussion in Section 6.3.2 of the GR field equation $G_{\mu\nu} = \kappa T_{\mu\nu}$ with $\kappa = -8\pi c^{-4} G_{\mathrm{N}}$. The Einstein tensor $G_{\mu\nu}$ on the left-hand side (LHS) is the curvature of spacetime and $T_{\mu\nu}$ on the right-hand side (RHS), the energy–momentum source term for gravity (the curved spacetime). The goal of obtaining a static universe from general relativity (GR) led Einstein to alter his field equation to make it contain a repulsion component. This could, in principle, balance the usual gravitational attraction to yield a static cosmic solution. Einstein discovered that the geometry side of his field equation could naturally accommodate an additional term. As will be discussed in Section 14.4.3, the simplest term that is mathematically compatible with Einstein's field equation (6.37) is the metric tensor $g_{\mu\nu}$,

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}. \tag{11.1}$$

Such a modification will, however, alter its nonrelativistic limit to differ from Newton's equation. In order that this alteration is compatible with known phenomenology, it must have a coefficient Λ so small as to be unimportant in all situations except on truly large cosmic scales. Hence, this additional constant Λ has come to be called the **cosmological constant**.

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 239 of 276 September 24, 2009 17:51

Λ as a vacuum energy momentum contribution While we have introduced this term as an additional geometric term, we could just as well move it to the RHS of the equation and view it as an additional source term of gravity. In particular, when the regular energy–momentum is absent $T_{\mu\nu} = 0$ (the vacuum state),

$$G_{\mu\nu} = \Lambda g_{\mu\nu} \equiv \kappa T^{\Lambda}_{\mu\nu}$$

where $T^{\Lambda}_{\mu\nu} = \kappa^{-1}\Lambda g_{\mu\nu} = (-c^4\Lambda/8\pi G_{\rm N})g_{\mu\nu}$ can be interpreted as the energy–momentum tensor of the vacuum.[2] Just as the $T_{\mu\nu}$ for ordinary radiation and matter depends on two functions of energy density ρ and pressure p, this vacuum-energy–momentum tensor $T^{\Lambda}_{\mu\nu}$ can be similarly parametrized by "**vacuum-energy density**" ρ_Λ and "**vacuum pressure**" p_Λ. As we shall demonstrate in Section 14.4.3 (after we have properly studied the energy–momentum tensor in Section 12.3), these two quantities are related to a positive cosmological constant Λ as follows: the vacuum energy per unit volume,

$$\rho_\Lambda = \frac{\Lambda c^2}{8\pi G_{\rm N}} > 0, \tag{11.2}$$

is a constant[3] (in space and in time) and the corresponding vacuum pressure,

$$p_\Lambda = -\rho_\Lambda c^2 < 0, \tag{11.3}$$

is negative, corresponding to an equation-of-state parameter $w = -1$ as defined in Eq. (10.4). Such a density and pressure, as we shall presently show, are compatible with basic physics principles, and, most relevant for our cosmological discussion, they give rise to a gravitational repulsion.

[2]It is appropriate that the vacuum energy momentum tensor is proportional to the metric $g_{\mu\nu}$, which is Lorentz invariant in local inertial coordinates. This must be the case as such a $T^{\Lambda}_{\mu\nu}$ should not pick out any preferred direction.

[3]In nonrelativistic physics only the relative value of energy is meaningful—the motion of a particle with potential energy $V(x)$ is exactly the same as one with $V(x) + C$, where C is a constant. In GR, since the whole energy–momentum tensor is the source of gravity, the actual value of energy makes a difference.

Λ as constant energy density and negative pressure What is a negative pressure? Consider the simple case of a piston chamber filled with ordinary matter and energy, which exerts a positive pressure by pushing out against the piston. If it is filled with this Λ energy, Fig. 11.1, it will exert a negative pressure by **pulling in** the piston. Physically this is sensible because, as its energy per unit volume $\rho_\Lambda c^2$ is a constant, the change in system's energy is strictly proportional to its volume change $dE = \rho_\Lambda c^2 dV$. The system would like to lower its energy by volume-contraction (pulling in the piston). When we increase the volume of the chamber $dV > 0$ (hence its energy $dE > 0$) by pushing out the piston, we have to do positive work to overcome the pulling by the Λ energy. Energy conservation is maintained in such a situation because the negative pressure $p < 0$ is just what is required by the first law of thermodynamics: $dE = -p dV$ when both dE and dV have the same sign. In fact the first law also makes it clear that if the energy density

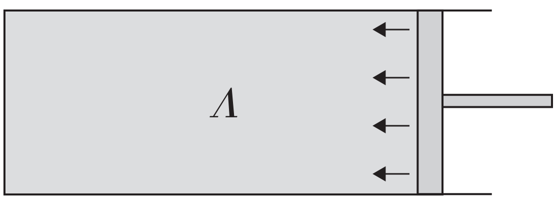


Fig. 11.1 The Λ energy in a chamber has negative pressure and pulls in the piston.

is a constant $dE = \rho c^2 dV$ so that the dV factors cancel from both sides, the pressure must equal the negative of the energy density $p = -\rho c^2$ as shown in (11.3).

11.1.1 Vacuum energy as source of gravitational repulsion

To see that the negative pressure can give rise to a repulsive force, let us first discuss the Newtonian limit of the Einstein equation with a general source, composed of mass density ρ as well as pressure p (as is the case for a cosmology with an ideal fluid as the source). It can be shown (see Box 14.1 for details) that the limiting equation, written in terms of the gravitational potential Φ, is

$$\nabla^2 \Phi = 4\pi G_N \left(\rho + 3\frac{p}{c^2}\right) = 4\pi G_N (1 + 3w)\, \rho. \tag{11.4}$$

This informs us that not only mass, but also pressure, can be a source of gravity. For the nonrelativistic matter having a negligible pressure term, we recover the familiar equation (6.36) of Newton.

Explicitly displaying the contributions from ordinary matter and vacuum energy (thus density and pressure each have two parts: $\rho = \rho_M + \rho_\Lambda$ and $p = p_M + p_\Lambda$), the Newton/Poisson equation (11.4) becomes

$$\begin{aligned}\nabla^2 \Phi &= 4\pi G_N \left(\rho_M + 3\frac{p_M}{c^2} + \rho_\Lambda + 3\frac{p_\Lambda}{c^2}\right) \\ &= 4\pi G_N \rho_M - 8\pi G_N \rho_\Lambda = 4\pi G_N \rho_M - \Lambda c^2,\end{aligned} \tag{11.5}$$

where we have used (11.3), $p_\Lambda = -\rho_\Lambda c^2$, and set $p_M = 0$ because $\rho_M c^2 \gg p_M$. For the vacuum-energy dominated case of $\Lambda c^2 \gg 4\pi G_N \rho_M$, the Poisson equation can be solved (after setting the potential to zero at the origin) by

$$\Phi_\Lambda (r) = -\frac{\Lambda c^2}{6} r^2. \tag{11.6}$$

Between any two mass points, this potential corresponds to a repulsive force (per unit mass) that increases with separation r,

$$\vec{g}_\Lambda = -\vec{\nabla}\Phi_\Lambda = +\frac{\Lambda c^2}{3}\vec{r}, \tag{11.7}$$

in contrast to the familiar $-\vec{r}/r^3$ gravitational attraction. With this pervasive repulsion that increases with distance, even a small Λ can have a significant effect on truly large dimensions. It would be possible to counteract the gravitational attraction and allow for the static solution sought by Einstein.

11.1.2 Einstein's static universe

We now consider the Friedmann equations (10.1) and (10.2) with a nonvanishing cosmological constant,

$$\frac{\dot{a}^2 + kc^2/R_0^2}{a^2} = \frac{8\pi G_N}{3}(\rho_M + \rho_\Lambda), \tag{11.8}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{c^2}\left[(p_M + p_\Lambda) + \frac{1}{3}(\rho_M + \rho_\Lambda)c^2\right]. \tag{11.9}$$

The RHS of (11.9) need not necessarily be negative because of the presence of the negative pressure term $p_\Lambda = -\rho_\Lambda c^2$. Consequently, a decelerating universe is no longer the inevitable outcome. For nonrelativistic matter ($p_M = 0$), we have

$$\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3}(\rho_M - 2\rho_\Lambda). \tag{11.10}$$

The static condition of $\ddot{a} = 0$ now leads to the constraint:

$$\rho_M = 2\rho_\Lambda = \frac{\Lambda c^2}{4\pi G_N}. \tag{11.11}$$

That is, the mass density ρ_M of the universe is fixed by the cosmological constant. The other static condition of $\dot{a} = 0$ implies, through (11.8), the static solution $a = a_0 = 1$

$$\frac{kc^2}{R_0^2} = 8\pi G_N \rho_\Lambda = \Lambda c^2. \tag{11.12}$$

Since the RHS is positive, we must have

$$k = +1. \tag{11.13}$$

Namely, the static universe has a positive curvature (a closed universe) and finite size. The "radius of the universe" is also determined, according to (11.12), by the cosmological constant:

$$R_0 = \frac{1}{\sqrt{\Lambda}}. \tag{11.14}$$

Thus, the basic features of such a static universe, the density and size, are determined by the arbitrary input parameter Λ. Not only is this a rather artificial arrangement, but also the solution is, in fact, unstable. That is, a small variation will cause the universe to deviate from this static point. A slight increase in the separation will cause the gravitational attraction to decrease and repulsion to increase, causing the system to deviate further from the initial point. A slight decrease in the separation will increase the gravitational attraction to cause the separation to decrease further, until the whole system collapses.

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 242 of 276 September 24, 2009 17:51

Box 11.1 Some historical tidbits of modern cosmology

FLRW cosmology The Friedmann equations with both ordinary and vacuum energies (11.8) and (11.9) are sometimes call **the Friedmann–Lemaître equations**. That Einstein's equation had expanding, or contracting, solutions was first pointed out in the early 1920s by A.A. Friedmann. His fundamental contribution to cosmology was hardly noticed by his contemporaries.[4] It was to be rediscovered later by the Belgian civil engineer and priest Georges Lemaître, who published in 1927 his model of cosmology with a contribution coming from both ρ_M and ρ_Λ. More importantly, Lemaître was the first one, having been aware of Hubble's work through his contact with Harvard astronomers (he spent three years studying at Cambridge University and MIT), to show that the linear relation between distance and redshift (Hubble's law) follows from such cosmological considerations. The original derivations by Friedmann and Lemaître were somewhat awkward. Modern presentations have mainly followed the approach initiated by Howard Percy Robertson and Arthur G. Walker. Thus the framework using Einstein's equation for a homogeneous and isotropic universe has come to be known as the **FLRW** (Friedmann–Lemaître–Robertson–Walker) **cosmological model**.

Einstein's greatest blunder? Having missed the chance of predicting an expanding universe before its discovery, Einstein came up with a solution which did not really solve the perceived difficulty. (His static solution is unstable.) It had often been said that later in life Einstein considered the introduction of the cosmological constant to be "the biggest blunder of his life!" This originated from a characterization by George Gamow in his autobiography (Gamow, 1970):

> *Thus, Einstein's original gravity equation was correct, and changing it was a mistake. Much later, when I was discussing cosmological problems with Einstein, he remarked that the introduction of the cosmological term was the biggest blunder he ever made in his life.*

Then Gamow went on to say,

> *But this blunder, rejected by Einstein, is still sometimes used by cosmologists even today, and the cosmological constant* Λ *rears its ugly head again and again and again.*

What we can conclude for sure is that Gamow himself considered the cosmological constant 'ugly' (because this extra term made the field equation less simple). Generations of cosmologist kept on including it because the quantum vacuum energy gives rise to such a term (cf. Section 11.7.1) and there was no physical principle one could invoke to exclude this term. (If it is not forbidden, it must exist!) In fact the discovery of the cosmological constant as the source of a new cosmic repulsive force must be regarded as one of Einstein's great achievements.[5]Now, as we shall see, the idea of a nonzero cosmological constant was the key in solving a number of

[4] On the other hand, the comoslogical model with only ρ_Λ, studied by the Dutch astronomer W. de Sitter soon after Einstein's 1917 paper, was widely discussed.

[5] One can speculate that, if there were regret on Einstein's part, it would be the missed opportunity of predicting the expanding universe before its observational discovery.

fundamental problems in cosmology. That is, Einstein taught us the way to bring about gravitational repulsion. Although the original goal of a static universe solution was misguided, this "tool" of the cosmological constant (a repulsive force) was needed to account for the explosion that was the big bang (inflationary epoch), and was needed to explain how the expansion of the universe could accelerate.

11.2 The inflationary epoch

The standard model of cosmology (the FLRW model) has been very successful in presenting a self-contained picture of the evolution and composition of the universe: how the universe expanded and cooled after the big bang; how the light nuclear elements were formed; after the inclusion of the proper density inhomogeneity, how in an expanding universe matter congealed to form stars, galaxies, and clusters of galaxies. It describes very well the aftermath of the big bang. However, the model says very little about the nature of the big bang itself: how did this "explosion of the space" come about? It assumes that all matter existed from the very beginning. Furthermore, it must assume certain very precise initial conditions (see the flatness and horizon problems discussed later) that just clamor for an explanation.

The inflationary cosmology is an attempt to give an account of this big bang back to an extremely short instant (something like 10^{-36} s) after the $t = 0$ cosmic singularity.[6] During this primordial inflation, the universe had a burst of expansion during which the scale factor increased by more than 30 orders of magnitude, see Fig. 11.2. In this inflationary process, all the matter and energy could have been created virtually from nothing. Afterwards, the universe followed the course of adiabatic expansion and cooling as described

[6]This is to be compared to the even earlier period, comparable to the Planck time $t_{\rm Pl} = O(10^{-43}$ s), when quantum gravity is required for a proper description. See Section 8.5.1.

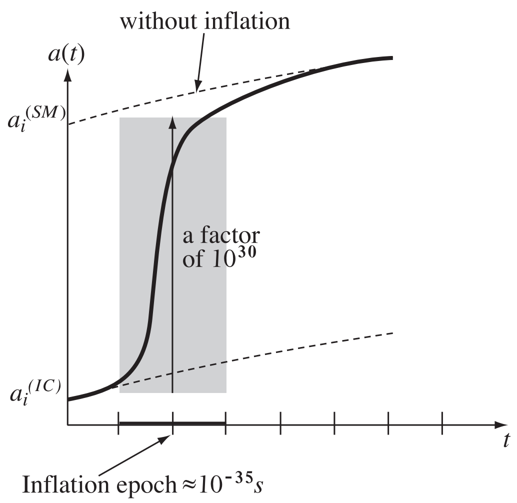


Fig. 11.2 Comparison of the scale factor's time evolution. The standard FLRW model curves are represented by dashed lines; the solid curve is that of the inflation model which coincides with the standard model curve after 10^{-35} s. The intercepts on the a-axis correspond respectively to the initial scales: $a_i^{(\rm SM)}$ in the standard model (without inflation) and $a_i^{(\rm IC)}$ in the inflation cosmology.

by the FLRW cosmology (cf. Chapter 10). Figure 11.2 also makes it clear that in the inflationary scenario, the observable universe originates from an entity some 10^{-30} times smaller than that which would have been the size in the case without inflation.

11.2.1 Initial conditions for the FLRW cosmology

The standard FLRW model requires a number of seemingly unnatural fine-tuned initial conditions. As we shall see, they are just the conditions that would follow from an inflationary epoch. We start the discussion of initial conditions by listing two such theoretical difficulties, two "problems."

The flatness problem

Because of the gravitational attraction among matter and energy, we would expect the expansion of the universe to slow down. This deceleration $\ddot{a}(t) < 0$ means that $\dot{a}(t)$ must be a decreasing function. This is exemplified by the specific case of a radiation-dominated universe $a \sim t^{1/2}$, thus $\dot{a} \sim t^{-1/2}$, or a matter-dominated universe $a \sim t^{2/3}$, and $\dot{a} \sim t^{-1/3}$, as derived in (10.30). Recall that the Friedmann equation can be written in terms of the mass density parameter Ω as in (10.7):

$$1 - \Omega(t) = \frac{-kc^2}{\dot{a}(t)^2 R_0^2}. \tag{11.15}$$

This displays the connection between geometry and matter/energy: if $k = 0$ (a flat geometry), we must have the density ratio $\Omega = 1$ exactly; when $k \neq 0$ for an universe having curvature, then $|1 - \Omega(t)|$ must be **ever-increasing** because the denominator on the RHS is ever decreasing. Thus, the condition for a flat universe $\Omega = 1$ is an **unstable equilibrium point**—if Ω ever deviates from 1, this deviation will increase with time. Or, we may say: gravitational attraction always enhances any initial curvature. In light of this property, it is puzzling that the present mass density Ω_0 has been found observationally (see Section 9.2) to be not too different from the critical density value $(1 - \Omega_0) = O(1)$. This means that Ω must have been extremely close to unity (an extremely flat universe) in the cosmic past. Such a fine-tuned initial condition would require an explanation.

We can make this statement quantitatively. Ever since the radiation–matter equality time $t > t_{\mathrm{RM}}$, with $z_{\mathrm{RM}} = O(10^4)$, cf. (10.68) the evolution of the universe has been dominated by nonrelativistic matter: $a(t) \sim t^{2/3}$ or $\dot{a} \sim t^{-1/3} \sim a^{-1/2}$. We can then estimate the ratio in (11.15):

$$\begin{aligned} \frac{1 - \Omega(t_{\mathrm{RM}})}{1 - \Omega(t_0)} &= \left[\frac{\dot{a}(t_{\mathrm{RM}})}{\dot{a}(t_0)}\right]^{-2} = \left[\frac{a_{\mathrm{RM}}}{a_0}\right] \\ &= (1 + z_{\mathrm{RM}})^{-1} = O(10^{-4}). \end{aligned} \tag{11.16}$$

Successful prediction of light element abundance by primordial nucleosynthesis gave us direct evidence for the validity of the FLRW model of cosmology back to the big bang nucleosynthesis time $t_{\mathrm{bbn}} = O(10^2\,\mathrm{s})$. The time evolution for $t < t_{\mathrm{RM}}$ was radiation dominated: $a(t) \sim t^{1/2}$ or $\dot{a} \sim t^{-1/2} \sim a^{-1}$.

This would then imply

$$\frac{1-\Omega(t_{\rm bbn})}{1-\Omega(t_{\rm RM})} = \left[\frac{\dot a(t_{\rm bbn})}{\dot a(t_{\rm RM})}\right]^{-2} = \left[\frac{a(t_{\rm bbn})}{a(t_{\rm RM})}\right]^{2}$$
$$= \left[\frac{k_{\rm B}T_{\rm bbn}}{k_{\rm B}T_{\rm RM}}\right]^{-2} \simeq O(10^{-11}), \tag{11.17}$$

where we have used the scaling behavior of the temperature, and (10.53) $k_{\rm B}T_{\rm bbn} = O(\text{MeV})$ and (10.69) $k_{\rm B}T_{\rm RM} = O(2\,\text{eV})$ to reach the last numerical estimate. Thus, in order to produce a $(\Omega_0 - 1) = O(1)$ now, the combined result of (11.16) and (11.17) tells us that one has to have at the epoch of primordial nucleosynthesis a density ratio equal to unity to an accuracy of one part in 10^{15}. Namely, we must have $\Omega(t_{\rm bbn}) - 1 = O(10^{-15})$. That the FLRW cosmology requires such an unnatural initial condition constitutes the flatness problem.

The horizon problem

Our universe is observed to be very homogeneous and isotropic. In fact, we can say that it is "too homogeneous and isotropic." Consider two different parts of the universe that are outside of each other's horizons. They are so far apart that no light signal sent from one at the beginning of the universe could have reached the other. Yet they are observed to have similar properties. This suggests their being in thermal contact sometime in the past. How can this be possible?

This horizon problem can be stated most precisely in terms of the observed isotropy of the CMB radiation (up to one part in 100 000, after subtracting out the dipole anisotropy due to the peculiar motion of our Galaxy). When pointing our instrument to measure the CMB, we obtain the same blackbody temperature in all directions. However, every two points in the sky with an angular separation on the order of a degree actually correspond to a horizon separation back at the photon-decoupling time t_γ, see (11.31). The age of the universe at the photon decoupling time was about 360 000 years, yet the observed isotropy indicates that regions far more than the horizon distance 360 000 light-year apart were strongly correlated. This is the horizon problem of the standard FLRW cosmology.

Initial conditions required for the standard cosmic evolution

We have discussed the horizon problem and flatness problem, etc. as the shortcomings of the standard big bang model. Nevertheless, it must be emphasized that they are not contradictions since we could always assume that the universe had just these conditions initially to account for the observed universe today. For example, the horizon problem can be interpreted simply as reflecting the fact that the universe must have been very uniform to begin with. These "problems" should be viewed as informing us of the correct initial conditions for the cosmic evolution after the big bang: "The initial conditions must be **just so**." What we need is a theory of the initial conditions. Putting it in another way, the FLRW model is really a theory for the evolution of the universe **after** the big bang. We now need a theory of the big bang **itself**. A correct theory

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 246 of 276 September 24, 2009 17:51

should have the feature that it would automatically leave behind a universe with just these desired conditions.

11.2.2 The inflation scenario

The initial condition problems can be solved if, in the early moments, the universe had gone through an epoch of extraordinarily rapid expansion. This can solve the flatness problem, as any initial curvature could be stretched flat by the burst of expansion, and can solve the horizon problem if the associated expansion rate could reach superluminal speed. If the expansion rate could be greater than the light speed, then one horizon volume could have been stretched out to such a large volume that corresponded to many horizon volumes after this burst of expansion. This rapid expansion could happen if there existed then a large cosmological constant Λ, which could supply a huge repulsion to the system. The question is, then, what kind of physics can give rise to such a large Λ? In this section, we explain how modern particle physics can suggest a possible mechanism to generate, for a short instant of time, such a large vacuum energy.

False vacuum, slow rollover phase transition and an effective Λ

Here we discuss the possibility of a field system that can give rise to an effective cosmological constant Λ_{eff} that then brings about an explosion of the space.

Hierarchy of particle physics unification and the invention of inflationary cosmology The inflationary cosmology was invented in 1980 by Alan Guth in his study of the cosmological implications of the grand unified theories (GUTs) of particle interactions. The basic idea of a GUT is that particle interactions possess certain symmetry.[7] As a result, all the fundamental forces (the strong, weak, and electromagnetic interactions, except for gravity) behave similarly at high energy. In fact they are just different aspects of the same (unified) interaction like the different faces of the same die. However, the structure of the theory is such that there is a phase transition at a temperature corresponding to the grand unification energy scale, around 10^{15}–10^{16} GeV. In the energy regime higher than this scale, the system is in a symmetric phase and the unification of particle interactions is manifest (i.e. all interactions behave similarly); when the universe cooled below this scale, the particle symmetry became hidden, showing up as distinctive forces.[8]

Higgs phenomenon in field theory In quantum field theory, particles are quantum excitations of their associated fields: electrons of the electron field, photons of the electromagnetic field, etc. New fields are postulated to exist, related to yet to be discovered particles. What brings about the above-mentioned spontaneous symmetry breaking and its associated phase transition is the existence of a certain spin-zero field $\phi(x)$, called the Higgs field (its quanta being Higgs particles). Such a field, just like the familiar electromagnetic field, carries energy. What is special about a Higgs field is that it possesses a potential energy density function $V(\phi)$ much like the potential energy function in the ferromagnet example of Section 11.6.

[7] "Particle interaction symmetry" has the same meaning as "symmetry in particle physics" as explained in Chapter 1: physics equations are unchanged under some transformation. However, instead of transformations of space and time coordinates as in relativity, here one considers transformations in some "internal charge space." The mathematical description of symmetry is group theory. An example of a grand unification group is $SU(5)$ and particles form multiplets in this internal charge space. Members of the same multiplet can be transformed into each other: electrons into neutrinos, or into quarks, and the GUT physics equations are covariant under such transformations. After spontaneous symmetry breaking, the interactions possess less symmetry: for example, $SU(5)$ is reduced down to $SU(3) \times SU(2) \times U(1)$, which is the symmetry group of the low energy effective theory known as the Standard Model of quantum chromodynamics and electroweak interactions.

[8] For a discussion of spontaneous symmetry breakdown, that is, hidden symmetry, as illustrated by spontaneous magnetization of a ferromagnet, see Section 11.6 (Appendix C).

Normally one would expect field values to vanish in the vacuum state (the state with the lowest energy). A Higgs field, surprisingly, can have a nonzero vacuum state field permeating throughout space, cf. Fig. 11.3 (a) and (b). The effect of this hidden symmetry can then spread to other particles through their interaction of the Higgs field. For example, a massless particle can gain its mass when propagating in the background of such a Higgs field. Different Higgs fields are posited to exist. Here we are referring to the Higgs particles[9] in GUTs, which may have a mass $O(10^{16}\ \mathrm{GeV}/c^2)$.

Slow rollover from a false vacuum gives rise to Λ In the cosmological context, such a postulated field is simply referred to as the **inflation field**, or **inflation/Higgs field**. In order to have a large Λ_{eff} over an interval long enough to produce the desired initial conditions for the FLRW cosmology, it was suggested that parameters of the unified theory were such that the potential energy function of the inflation field had a very small slope around the $\phi = 0$ origin as in Fig. 11.3(c). As the universe cools, the temperature dependent parameters change so that the potential energy function changes from Fig. 11.15(a) to (b). The prior lowest energy point at zero field value became a local maximum and the system would rollover to the new asymmetric vacuum state where the Higgs field would have a nonvanishing vacuum value. But the parameters are such that this rollover was slow. During this transition, we could regard the system, compared to the true (asymmetric) vacuum state, as having an extra energy density. We say the system (i.e. the universe) was temporarily in a **false vacuum**. Having this vacuum-energy density, which is time and position independent, the universe effectively had a large cosmological constant.

Exponential expansion in a vacuum-energy dominated universe

Let us consider the behavior of the scale factor $a(t)$ in a model with $\Lambda > 0$ when the matter density can be ignored. In such a vacuum-energy dominated situation, the expansion rate $\dot{a}(t)$ is so large, cf. (11.22) that we can always ignore the curvature term in Eq. (11.8):

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G_{\mathrm{N}}}{3}\rho_\Lambda = \frac{\Lambda c^2}{3}. \tag{11.18}$$

Thus $\dot{a}$ is proportional to the scale factor a itself. Namely, we have the familiar rate equation. It can be solved to yield an exponentially expanding universe (called the **de Sitter universe**):

$$a(t) \equiv a(t_1)\, e^{(t-t_1)/\Delta\tau} \tag{11.19}$$

with the time constant

$$\Delta\tau = \sqrt{\frac{3}{\Lambda c^2}} = \sqrt{\frac{3}{8\pi G_{\mathrm{N}}\rho_\Lambda}}, \tag{11.20}$$

where we have expressed the cosmological constant in terms of the vacuum-energy density $\rho_\Lambda c^2$ as in (11.2). Physically we can understand this exponential expansion result because the repulsive expansion is self-reinforcing: as the energy density ρ_Λ is a constant, the more the space expands, the greater is the

[9] These Higgs particles should not be confused with the Standard Model's electroweak Higgs particle, thought to have a mass on the order of $10^2\ \mathrm{GeV}/c^2$, which is responsible for giving masses to electrons and quarks as well as the W and Z bosons that mediate weak interactions. Our discussion of the inflation scenario is couched in the language of the grand unified Higgs field. It should be understood the grand unified theories themselves have not been verified experimentally in any detail because its intrinsic energy scale of 10^{16} GeV is so much higher than the highest energy $\approx 10^3$ GeV reachable by our accelerators. On the other hand, we are confident that some version of grand unification is correct, as the simplest GUTs can already explain several puzzles of the Standard Model of particle physics, such as why the strong interaction is strong, the weak interaction weak, and why the quarks and leptons have the charges that they do. Nevertheless, the connection between grand unification and inflation cosmology has remained only as a suggestive possibility. It was our knowledge of the grand unification theory that allowed the construction of a physically viable scenario that could give rise to an inflationary epoch. But what precisely is the inflation field, and what parameters actually govern its behavior remain as topics of theoretical discussion. The remarkable fact is that some reasonable speculation of this type can already lead to the resolution of many cosmological puzzles, and have predictions that have been consistently checked with observation.

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 248 of 276 September 24, 2009 17:51

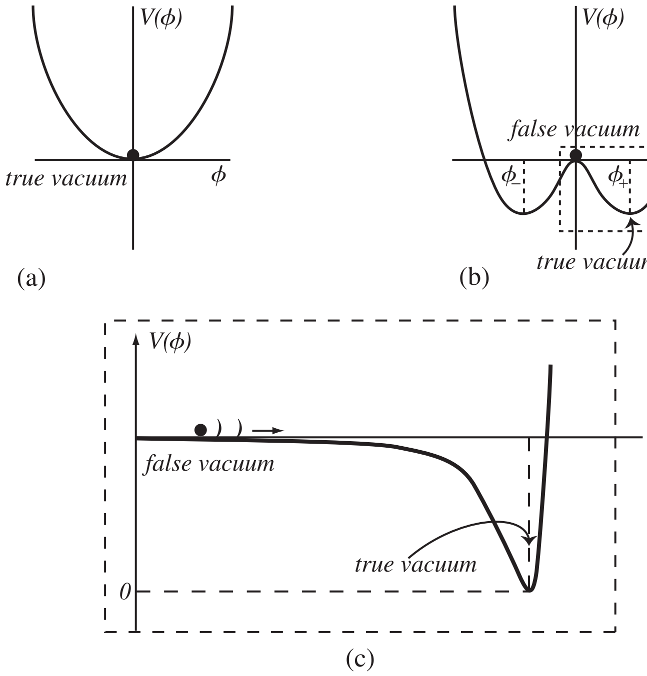


Fig. 11.3 Potential energy function of a Higgs field $\phi(x)$ is illustrated by the simple case of $V(\phi) = \alpha\phi^2 + \lambda\phi^4$, possessing a discrete symmetry $V(-\phi) = V(\phi)$. The parameter α has temperature-dependence, for example, $\alpha = \alpha_0(T - T_c)$, with positive constants α_0 and λ. (a) Above the critical temperature ($T > T_c$, hence $\alpha > 0$), we have the normal case of the lowest energy state (the vacuum) being at $\phi_0 = 0$, which is symmetric under $\phi \to -\phi$. (b) Below T_c (hence $\alpha < 0$), the symmetric $V(\phi)$ has the lowest energy at points $\phi_\pm = \pm\sqrt{-\alpha/2\lambda}$ while $V(\phi = 0)$ is a local maximum. The choice of the vacuum state being either of the asymmetric ϕ_+ or ϕ_- breaks the symmetry, cf. similar plot in Fig 11.15(b). The dashed box in (b) is displayed in (c) to show that the inflation/Higgs potential $V(\phi)$ has an almost flat portion at the $\phi = 0$ origin for a slow rollover transition. The dot represents the changing location of the system—rolling from a high plateau of the false vacuum towards the true vacuum at the bottom of the trough.

vacuum energy and negative pressure, causing the space to expand even faster. In fact, we can think of this Λ repulsive force as residing in the space itself, so as the universe expands, the push from this Λ energy increases as well.[10] We note that the total energy was conserved during the inflationary epoch's rapid expansion because of the concomitant creation of the gravitational field, which has a **negative** potential energy (cf. Section 10.3.1).

[10]Because Λ represents a constant energy density, it will be the dominant factor $\rho_\Lambda \gg \rho_M$ at later cosmic time, as $\rho_M \sim a^{-3}$. This dominance means that it is possible for the universe to be geometrically closed ($\Omega > 1$ and $k = +1$), yet does not stop expanding. Namely, with the presence of a cosmological constant, the mass/energy density Ω (hence the geometry) no longer determines the fate of the universe in a simple way. In general, a universe with a nonvanishing Λ, regardless of its geometry, would expand forever. The only exception is when the matter density is so large that the universe starts to contract before ρ_Λ becomes the dominant term.

11.2.3 Inflation and the conditions it left behind

In the previous section we have described how the grand unification Higgs field associated with spontaneous symmetry breaking can serve as the inflation field. A patch of the universe with this "inflation/Higgs matter" might have undergone a slow rollover phase transition and thus lodged temporarily in a false vacuum with a large constant energy density. The resultant effective cosmological constant Λ_{eff} provided the gravitational repulsion to inflate the scale factor exponentially. The a grand unification thermal energy scale is $E_{GU} = O(10^{16}\,\text{GeV})$, that is, a temperature $T_{GU} = O(10^{29}\,\text{K})$, which according to (10.44) corresponds[11] to the cosmic time $t_{GU} \simeq 10^{-36}\,\text{s}$. The energy density $\rho_{GU}c^2$ can be estimated as follows: in a relativistic quantum system (such as quantum fields) there is the natural energy length scale given by the product of Planck's constant (over 2π) times the velocity of light: $\hbar c = 1.97 \times 10^{-16}\,\text{GeV} \cdot \text{m}$. Using this conversion factor we have the energy density scale for grand unification

$$\rho_{GU}c^2 \simeq \frac{(E_{GU})^4}{(\hbar c)^3} \simeq 10^{100}\,\text{J/m}^3. \tag{11.21}$$

[11]One can check this estimate of $t_{GU} \simeq 10^{-36}\,\text{s}$ for a thermal energy $E_{GU} = 10^{16}\,\text{GeV}$ with a primordial nucleosynthesis time of $t_{NS} \simeq 10^2\,\text{s}$ when the thermal energy was $E_{NS} = 1\,\text{MeV}$ in this way.

$$\frac{E_{GU}}{E_{NS}} = \frac{kT_{GU}}{kT_{SN}} = \frac{a\,(t_{NS})}{a\,(t_{GU})} = \sqrt{\frac{t_{NS}}{t_{GU}}}$$

because the period between inflation and the nucleosynthesis time was radiation dominated $a \sim \sqrt{t}$. The ratio of $E_{GU}/E_{NS} = 10^{19}$ indeed matches that of $t_{NS}/t_{GU} = 10^{38}$ for $t_{GU} \simeq 10^{-36}\,\text{s}$.

For a vacuum energy density $\rho_\Lambda \approx \rho_{GU}$, the corresponding exponential expansion time constant $\Delta\tau$ of (11.20) had the value $\Delta\tau \simeq 10^{-37}$ s. Namely, the exponential inflationary expansion took place when the universe was $t_{GU} \simeq 10^{-36}$ s old, with an exponential expansion time constant of $\Delta\tau = O(10^{-37}\text{ s})$. By a "slow" rollover phase transition we mean that the parameters of the theory are such that inflation might have lasted much longer than 10^{-37} s, for example, 10^{-35} s ($\lesssim 100$ e-fold), expanding the scale factor by more than 30 orders of magnitude, until the system rolled down to the true vacuum, ending the inflation epoch (cf. Fig. 11.3). Afterwards the universe commenced the adiabatic expansion and cooling according to the standard FLRW model until the present epoch.[12] Such dynamics have the attractive property that they would leave behind precisely the features that had to be postulated as the initial conditions for the standard FLRW cosmology.

The horizon and flatness problems solved

With the exponential behavior of the scale factor in (11.19), we can naturally have superluminal ($\dot{a}R_0 > c$) expansion as the rate $\dot{a}(t)$ also grows exponentially.[13] This does not contradict special relativity, which says that an object cannot pass another one faster than c in one fixed frame. Putting it another way, while an object cannot travel faster than the speed of light through space, there is no restriction stipulating that space itself cannot expand faster than c. Having a superluminal expansion rate, this inflationary scenario can solve the horizon problem, because two points that are a large number of horizon lengths apart now (or at the photon decoupling time when the CMB was created) could still be in causal contact before the onset of the inflationary epoch. They started out being thermalized within one horizon volume before the inflation epoch, but became separated by many horizon lengths due to the superluminal expansion.

This inflationary scenario can solve the flatness problem because the space was stretched so much that it became, after the inflationary epoch, a geometrically flat universe to a high degree of accuracy. When this exponential expansion (11.19) is applied to the Friedmann equation (11.15), it yields the ratio

$$\frac{1-\Omega(t_2)}{1-\Omega(t_1)} = \left[\frac{\dot{a}(t_2)}{\dot{a}(t_1)}\right]^{-2} = e^{-2(t_2-t_1)/\Delta\tau}. \tag{11.22}$$

Just as the scale factor was inflated by a large ratio, say, $e^{(t_2-t_1)/\Delta\tau} = 10^{30}$, we can have the RHS as small as 10^{-60}. Starting with any reasonable value of $\Omega(t_1)$ we can still have, after the inflation, a $\Omega(t_2) = 1$ to a high accuracy. While the cosmic time evolution in the FLRW model, being determined by gravitational attraction, always enhances the curvature by driving the universe away from $\Omega = 1$ (hence the flatness problem), the accelerating expansion due to the vacuum repulsion always pushes the universe (very rapidly) toward the $\Omega = 1$ point. Thus a firm prediction by the inflationary scenario is that the universe left behind by inflation must have a flat geometry and, according to GR, a density equal to the critical value (11.15)—although it does not specify what components make up such a density.

[12] It had generally been assumed that the effective cosmological constant, associated with the false vacuum, vanished at the end of the inflationary epoch. The general expectation was that the standard FLRW cosmology that followed the inflation epoch was one with no cosmological constant. Part of the rationale was that a straightforward estimate of the cosmological constant, as due to the zero-point energy of a quantum vacuum, yielded such an enormously large Λ (see Section 11.7) that many had assumed that there must be some yet-to-be discovered symmetry argument that would strictly forbid a nonzero cosmological constant. However, as we shall see below, more recent discoveries point to a nonvanishing, but small, Λ. The challenge is now how to explain the presence of such a "dark energy" in the universe.

[13] The Hubble constant, being the ratio of scale change rate per unit scale $\dot{a}/a$, does not change under such an exponential expansion of the scale factor.

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 250 of 276 September 24, 2009 17:51

The origin of matter/energy and structure in the universe

Besides the flatness and horizon problems, the standard FLRW cosmology requires as initial conditions that all the energy and particles of the universe be present at the very beginning. Furthermore, this hot soup of particles should have just the right amount of **initial density inhomogeneity** (density perturbation) which, through subsequent gravitational clumping, formed the cosmic structure of galaxies, clusters of galaxies, voids, etc. we observe today. One natural possibility is that such a density perturbation resulted from quantum fluctuation of particle fields in a very early universe. However, it is difficult to understand how such microscopic fluctuations can bring forth the astrophysical-sized density nonuniformity required for the subsequent cosmic construction. Remarkably, the inflationary cosmology can provide us with an explanation of the origin of matter/energy, as well as the structure of the universe.

The inflation model suggests that at the beginning of the big bang a patch of the inflation/Higgs matter (smaller than the size of a proton) underwent a phase transition bringing about a huge gravitational repulsion. This is the driving force behind the space-explosion that was the big bang. While this inflation material (the Λ energy) expanded exponentially in size to encompass a space that eventually developed into our presently observed universe, its energy density remained essentially a constant. In this way more and more particle/field energy was "created" during the inflationary epoch. When it ended with the universe reaching the true vacuum, its oscillations at the trough in Fig. 11.3 showed up, according to quantum field theory, as a soup of ordinary particles. According to the inflation theory, the initial potential energy of the inflation/Higgs field (having little kinetic energy) was the origin of our universe's matter content when it was converted into relativistic particles. In short, it is the vacuum energy that drove the inflation that would in the end decay into radiation and matter.

The phenomenon of particle creation in an expanding universe can be qualitatively understood as follows: according to quantum field theory, the quantum fluctuations of the field system can take on the form of the appearance and disappearance of particle–antiparticle pairs in the vacuum. Such energy nonconserving processes are permitted as long as they take place on a sufficiently short time-scale Δt so that the uncertainty relation $\Delta E \Delta t \leq \hbar$ is not violated. In a static space, such "virtual processes" do not create real particles. However, when the space is rapidly expanding, that is, the expansion rate was larger than the annihilation rate, real particles were created.[14] Thus, inflation in conjunction with quantum field theory naturally gives rise to the phenomenon of particle creation. This hot, dense, uniform collection of particles is just the postulated initial state of the standard big bang model. Furthermore, the scale factor had increased by such a large factor that it could stretch the subatomic size fluctuation of a quantum field into astrophysical sized density perturbation to seed the subsequent cosmic structure formation. The resultant density fluctuation was Gaussian (i.e. maximally random) and scale-invariant (i.e. the same fluctuation, of the order of 10^{-5}, in the gravitational potential on all length-scales) as will be discussed in Box 11.2 below.

[14] This way of seeding the structure formation can be viewed as "Hawking radiation from inflation." Recall our discussion in Section 8.5 of Hawking radiation from black holes in which virtual particles are turned into real ones because of a black hole event horizon. Here the production of real particles from quantum fluctuation comes about because of the horizon created by the hyper-accelerating expansion of the universe.

11.3 CMB anisotropy and evidence for a flat universe

As discussed in Section 11.2.3, inflationary cosmology predicts that the spacetime geometry of our universe must be flat. This prediction received more direct observational support through detailed measurement of the temperature anisotropy of the CMB radiation.[15]

The CMB is the earliest and largest observable thing in cosmology. Its remarkable uniformity over many horizon lengths reflects its origination from a single pre-inflation horizon volume. Just before the photon decoupling time t_γ, the universe was composed of dark matter and a tightly bound photon–baryon fluid. The inflationary scenario, with its associated phenomenon of particle creation, also generated a small density perturbation on a wide range of distance scales onto this overall homogeneity. Because of gravitational instability, this nonuniform distribution of matter eventually evolved into the cosmic structure we see today. In the early universe up till t_γ, the gravitational clumping of baryons was resisted by photon radiation pressure. This set up acoustic waves of compression and rarefaction with gravity being the driving force and radiation pressure the restoring force. All this took place against a background of dark matter fluctuations, which started to grow right after the radiation–matter equality time because dark matter did not interact with radiation.[16] Such a photon–baryon fluid can be idealized by ignoring the dynamical effects of gravitation and baryons (because the photon number density is much higher than that of baryons). This leads to a sound wave speed

$$c_s \simeq \sqrt{\frac{p}{\rho}} \simeq \frac{c}{\sqrt{3}} \tag{11.23}$$

as pressure and density being approximated by those for radiation $p \approx \rho c^2/3$. This cosmic sound left an imprint that is still discernible today. The compression and rarefaction was translated through gravitational redshift into a temperature inhomogeneity. By a careful analysis of this wave pattern, we can garner much information about the universe at this early epoch.

[15] The purpose of Section 11.3 is to present the observational evidence for a **flat universe**. As the discussion is somewhat more difficult, some readers may wish to skip it and proceed directly to Section 11.4.

[16] We remind ourselves that the dominant form of matter in the universe is cold dark matter. We can have the somewhat simplified picture: gravitational clumping took place principally among such nonbaryonic dark matter particles after t_{RM}. The baryonic matter inhomogeneity was not amplified until t_γ (when its resistance by radiation pressure disappeared); thereafter it fell into the dominant gravitational potential (the cosmic scaffoldings of Section 10.5.4) formed principally by the dark matter.

11.3.1 Three regions of the angular power spectrum

We shall present only a qualitative discussion of the power spectrum of the temperature anisotropy to give the reader some general idea of how a detailed analysis will allow one to fix a number of important cosmological parameters. From (10.86) and (10.88) for the correlation function, we see that the mean-square temperature anisotropy may be written for large multipole number l as

$$\left\langle \left(\frac{\delta T}{T}\right)^2 \right\rangle = \frac{1}{4\pi}\sum_{l=0}^{\infty}(2l+1)C_l \approx \int \frac{l(l+1)C_l}{2\pi} d\,(\ln l)\,. \tag{11.24}$$

$(l(l+1)/2\pi)C_l$ is approximately the power per logarithmic interval, and is the quantity presented in the conventional plot of the power spectrum against a logarithmic multipole number (cf. Figs. 11.4 and 11.13).

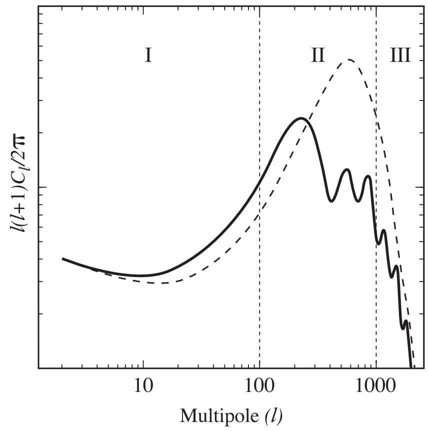


Fig. 11.4 CMB power spectrum as a function of the multipole moments. The solid curve with peaks and troughs (the acoustic peaks) is the prediction by the inflation model (with cold dark matter). The physics corresponding to the three marked regions is discussed in the text. The dashed curve is that by the topological defect model for the origin of the cosmic structure.

On small sections of the sky where curvature can be neglected, the spherical harmonic analysis becomes ordinary Fourier analysis in two dimensions. In this approximation the multipole number l has the interpretation as the Fourier wavenumber. Just as the usual Fourier wavenumber is $k \approx \pi/x$, the multipole moment number is $l \approx \pi/\theta$: large l corresponds to small angular scales with $l \approx 10^2$ corresponding to degree scale separation.

The CMB anisotropy is observed to be adiabatic (all particle species varied together) and this is consistent with the idea that the density fluctuation at t_γ was due to the primordial wrinkles of spacetime left behind by the earlier inflationary epoch. The inflationary scenario left behind density fluctuations that were Gaussian and scale invariant (cf. Box 11.2). Such the initial density perturbation, together with the assumption of a dark matter content dominated by nonrelativistic particles (the "cold dark matter" model) leads to a power spectrum as shown in Fig. 11.4. We can broadly divide it into three regions:

Region I ($l < 10^2$) This flat portion at large angular scales (the "Sachs–Wolfe plateau") corresponds to oscillations with a period larger than the age of the universe at the photon decoupling time. These waves are essentially frozen in their initial configuration. The flatness of the curve reflects the scale-invariant nature of the initial density perturbation as given by the inflation cosmology (cf. Box 11.2).

Region II ($10^2 < l < 10^3$) At these smaller angular scales (smaller than the sound horizon), there had been enough time for the photon–baryon fluid to undergo oscillation. The peaks correspond to regions having higher, as well as lower, than average density.[17] The troughs are regions with neutral compression, thus have maximum velocity (recall our knowledge of oscillators). The CMB from such regions underwent a large Doppler shift. In short, here is a snapshot of the acoustic oscillations with modes (fundamental plus harmonics) having different wavelengths and different phases of oscillations. The relative heights of the acoustic peaks are related to cosmological parameters

[17] This is so because the power spectrum is the square of a_{lm} and hence indifferent to their signs.

such as the baryon and the cold dark matter densities. Higher Ω_B enhances the odd-numbered acoustic peaks relative to the even-numbered ones, while higher Ω_{DM} lowers all of the peaks. The positions of the peaks depend on the characteristic length on the surface of last scattering as well as on the spatial curvature of the intervening universe since (see the subsection below).

Region III ($l > 10^3$) Photon decoupling did not take place instantaneously, i.e. the last scattering surface had a finite thickness.[18] Photons can diffuse out from any over-dense region if it was smaller than the photon's mean free path, which was increasing as the universe expanded. The net effect was an exponential damping of the oscillation amplitude in these sub-arcminute scales.

[18] The transition took about 50 000 years.

Box 11.2 Density fluctuation from inflation is scale-invariant

Inflation produces such a huge expansion that subatomic size quantum fluctuations were stretched to astrophysical dimensions.[19] For fluctuations larger than the sound horizon $\approx c_s H^{-1}$ one can ignore pressure gradients, as the associated sound waves cannot have crossed the perturbation in a Hubble time. The density perturbation without a pressure gradient would evolve like the homogeneous universe (Problem 11.1):

$$\rho a^2(\Omega^{-1} - 1) = \text{const.} \tag{11.25}$$

where a is the scale factor. With $\Omega = 1 + \Delta\Omega$ and $\rho = \rho_c + \Delta\rho$, the above relation implies, for small $\Delta\Omega$ and $\Delta\rho$,

$$\rho_c a^2 \Delta\Omega = a^2 \Delta\rho = \text{const.} \tag{11.26}$$

We now consider the implication of this scaling behavior for the perturbation in a gravitational potential on a physical distance scale of aL,

$$\Delta\Phi = \frac{G_N \Delta M}{aL} = \frac{4\pi}{3}\frac{G_N \Delta\rho\,(aL)^3}{(aL)} = \left(\frac{4\pi L^2 G_N}{3}\right) a^2 \Delta\rho.$$

Because of (11.26), the gravitational potential perturbation $\Delta\Phi$ over a comoving length L is scale invariant. During the inflationary epoch the.scale factor a would change by something like 30 decades; yet we would have the same $\Delta\Phi$ for a huge range[20] of physical distances of aL. Thus, inflationary cosmology makes the strong prediction of a scale-invariant density perturbation—the same fluctuation (of 10^{-5}) on all distance scales. It can be shown that such a density fluctuation, called the Harrison–Zel'dovich spectrum, would produce an angular power spectrum for the CMB anisotropy of the form

$$C_l = \frac{\text{const.}}{l\,(l+1)}.$$

Thus in the plot of $l\,(l+1)\,C_l$ vs. l in Fig 9.4 the power spectrum for the large angle region ($l < 100$) is a fairly flat curve.

[19] This turns the inevitable quantum effect into the seeds of structure in our universe.

[20] The same level of distortion (the warping of spacetime due to quantum fluctuation) was imprinted on all scales.

In Box 11.2 we have presented the power spectrum as predicted by the inflationary cosmology: a Gaussian density perturbation leading to a random distribution of hot and cold spots on the temperature anisotropy map, and a power spectrum displaying peaks and troughs. It is illuminating to contrast this with an alternative theory of cosmic structure origin, the topological defect model. In this scenario, one posits that as the universe cooled to a thermal energy of 10^{16} GeV, the phase transition that breaks the associated grand unification symmetry also produces defects in the fabric of spacetime—in the form of strings, knots, and domain walls, etc. This introduced the initial density perturbation that seeded the subsequent structure formation. Such a density fluctuation would produce line-like discontinuities in the temperature map and a smooth power spectrum (instead of the wiggly features as predicted by the inflation model), see Fig. 11.4. As we shall discuss in the next subsection, the observed CMB anisotropy favors inflation over this topological defect model for the origin of the cosmic structure.

11.3.2 The primary peak and spatial geometry of the universe

Consider the oscillatory power spectrum in region II of Fig. 11.4. The temperature anisotropy of the CMB is the result of a pattern of density fluctuations on a spherical surface centered on us. It reflects the sound wave spectrum of the photon–baryon fluid at the photon decoupling time, i.e. on the surface of last scattering. There would be standing waves having wavelength $\lambda_n = \lambda_1/n$, with the fundamental wavelength given by the sound horizon,[21] cf. (9.45):

$$\lambda_1 = \int_0^{t_\gamma} \frac{c_s dt}{a(t)} \approx c_s \int_0^{t_\gamma} \frac{dt}{a(t)}. \tag{11.27}$$

[21]This is the distance that a light signal ($ds^2 = 0$) would have traveled since the beginning of the universe ($t = 0$).

Now such a wavelength on the surface of last scattering would appear as angular anisotropy of scale

$$\alpha_1 \simeq \lambda_1/d(t_\gamma), \tag{11.28}$$

where $d(t_\gamma)$ is the (proper) radial distance between us now (t_0) and the photon decoupling time (t_γ). Namely, it is the comoving distance a photon would have traveled to reach us (t_0) from the surface of last scattering (also called the angular diameter distance)

$$d\left(t_\gamma\right) = c \int_{t_\gamma}^{t_0} \frac{dt}{a(t)}. \tag{11.29}$$

When evaluating the integrals in (11.27) and (11.29), we shall assume a matter-dominated flat universe with time dependence of the scale factor $a(t) \propto t^{2/3}$ as given by (10.30),

$$\int \frac{dt}{a(t)} \propto \int a^{-1/2} da \propto a^{1/2} = (1+z)^{-1/2}. \tag{11.30}$$

Matter-domination is plausible because the radiation–matter equality time is almost an order of magnitude smaller than the photon decoupling time, that is, according to (10.68) the redshift is $z_{\mathrm{RM}} \gg z_\gamma$. Thus the fundamental

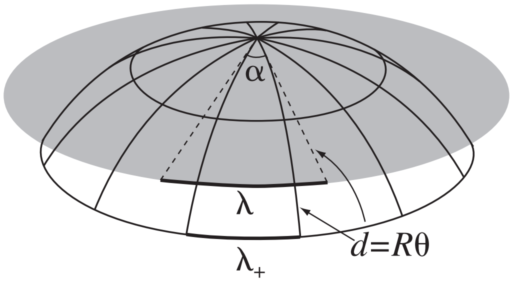


Fig. 11.5 A comparison of subtended lengths in a flat (shaded) vs. positively curved surface. For the same angular diameter distance d, the same angle α subtends a smaller wavelength λ_+ in a closed universe when compared to the corresponding $\lambda = \lambda_+ \left[(R_0/d)\sin(R_0/d)\right]^{-1} > \lambda_+$ in a flat universe.

wavelength corresponds to an angular separation of

$$\alpha_1 \approx \frac{\lambda_1}{d(t_\gamma)} = \frac{c_s(1+z_\gamma)^{-1/2}}{c(1+z_0)^{-1/2} - (1+z_\gamma)^{-1/2}}$$
$$\simeq \frac{(1+z_\gamma)^{-1/2}}{\sqrt{3}} \simeq 0.017\,\text{rad} \simeq 1^\circ, \tag{11.31}$$

where we have used $z_0 = 0$, $z_\gamma \simeq 1100$ and, as discussed in (11.23), a sound speed $c_s \simeq c/\sqrt{3}$. This fundamental wave angular separation in turn translates into the multipole number

$$l_1 \simeq \frac{\pi}{\alpha_1} \simeq \pi\sqrt{3(1+z_\gamma)} \approx 200. \tag{11.32}$$

Thus, in a flat universe we expect the first peak of the power spectrum to be located at this multipole number.

The above calculation was performed for a flat universe. What would be the result for a spatially curved universe? We will simplify our discussion by the suppression of one dimension and consider a 2D curved surface. In a positive curved closed universe ($k = +1$), light travels along longitudes (Fig. 11.5). A physical separation λ_1 at a fixed latitude, with polar angle θ and a coordinate distance $d = R_0\theta$, subtends an angle

$$\alpha_{1+} = \frac{\lambda_1}{R_0 \sin\theta} = \frac{\lambda_1}{R_0 \sin(d/R_0)} = \frac{\lambda_1}{d}\left(1 + \frac{d^2}{3R_0^2} + \cdots\right) > \frac{\lambda_1}{d}.$$

Namely, at a given scale (λ_1) at a fixed distance (d) the separation angle (α_{1+}) would appear to be larger (than the case of a flat universe). For a negatively curved open universe ($k = -1$), one simply replaces the sine by the hyperbolic sine:

$$\alpha_{1-} = \frac{\lambda_1}{R_0 \sinh(d/R_0)} = \frac{\lambda_1}{d}\left(1 - \frac{d^2}{3R_0^2} + \cdots\right) < \frac{\lambda_1}{d}.$$

At a given scale at a fixed distance, the separation angle would appear to be smaller. With the multipole number being inversely proportional to the separation angular scale, in a universe with spatial curvature the first peak would be shifted away from $l_1 \approx 200$, to a smaller (larger) multipole number for a closed (open) universe.

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 256 of 276 September 24, 2009 17:51

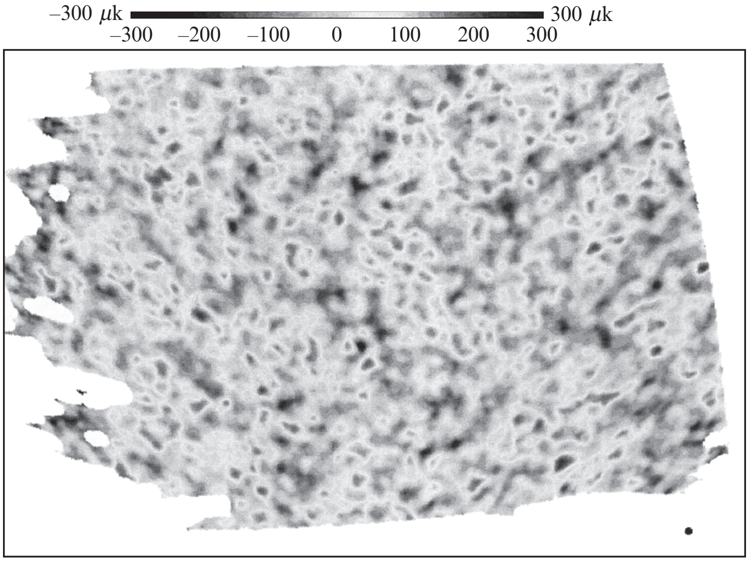


Fig. 11.6 Image of the complex temperature structure of CMB over 2.5% of the sky as captured by the Boomerang balloon-borne detector. It is interpreted as a freeze-frame picture of the sound wave patterns in the universe at the photon decoupling time. The black dot at the lower right-hand corner represents the size of a full moon subtending an angle of about half-a-degree.

Although the COBE satellite mapped the entire sky with high sensitivity discovering the CMB anisotropy at $\delta T/T = O(10^{-5})$, its relatively coarse angular resolution of $O(7^\circ)$ was not able to deduce the geometry of our universe. In the late 1990s a number of high altitude observations, e.g. MAT/TOCO (Miller *et al.,* 1999), and balloon-borne telescopes, Boomerang (de Bernardis *et al.,* 2000), and Maxima-1 (Hanany *et al.,* 2000), had detected CMB fluctuations on smaller sizes. These observations (see Fig. 11.6) produced evidence for a flat universe by finding the characteristic size of the structure to be about a degree wide and a power spectrum peaked at $l \approx 200$, see Fig. 11.4. The $k = 0$ statement is of course equivalent, via the Friedmann equation, to a total density $\Omega_0 = 1$. A careful matching of the power spectrum led to

$$\Omega_0 = 1.03 \pm 0.03. \tag{11.33}$$

In the meantime, another dedicated satellite endeavor, WMAP (Wilkinson Microwave Anisotropy Probe), had reported their results in a series of publications (Bennett, 2003; Hinshaw, 2009). Another influential cosmological project has been the survey of galaxy distributions by SDSS (Sloan Digital Sky Survey). Their high resolution result allowed them to extract many important cosmological parameters: H_0, Ω_0, $\Omega_{M,0}$, Ω_B, and the deceleration parameter q_0, etc. (to be discussed in Section 11.5).

11.4 The accelerating universe in the present epoch

Thus by mid/late-1990s there was definitive evidence that the geometry of the universe is flat as predicted by inflation. Nevertheless, there were several pieces of phenomenology that appeared in direct contradiction to such a picture.

A missing energy problem The Friedmann equation (10.7) requires a flat universe to have a mass/energy density exactly equal to the critical density, $\Omega_0 = 1$. Yet observationally, including both the baryonic and dark matter, we can only find less than a third of this value (radiation energy is negligibly small in the present epoch):

$$\Omega_{\mathrm{M}} = \Omega_{\mathrm{B}} + \Omega_{\mathrm{DM}} \simeq 0.25. \tag{11.34}$$

Thus, it appears that to have a flat universe we would have to solve a "missing energy problem."

A cosmic age problem From our discussion of the time evolution of the universe, we learned that the age of a flat universe should be two-third of the Hubble time, see (10.70),

$$(t_0)_{\mathrm{flat}} = \frac{2}{3} t_{\mathrm{H}} \lesssim 9\,\mathrm{Gyr}, \tag{11.35}$$

which is shorter than the estimated age of old stars. Notably the globular clusters have been deduced to be older than 12 Gyr (cf. Section 9.1.3). Thus, it appears that to have a flat universe we would also have to solve a "cosmic age problem."

Possible resolution with a dark energy A possible resolution of these phenomenological difficulties of a flat universe (hence inflationary cosmology) would be to assume the presence of a **dark energy**. A dark energy is defined as the "negative equation-of-state energy," $w < -1/3$ in Eq. (10.4). It gives rise to a gravitational repulsion, cf. Eq. (11.4). The simplest example of dark energy[22] is Einstein's cosmological constant, with $w = -1$. Such a cosmological constant assumed to be present even after inflation cannot have the immense size as the one it had during the inflation epoch. Rather, the constant dark energy density ρ_Λ should now be about three-quarters of the critical density to provide the required missing energy.

[22] One should not confuse dark energy with the energies of neutrinos, WIMPs, etc., which are also "dark," but are counted as parts of the "dark matter" (cf. Section 9.2), as the associated pressure is not negative.

$$\Omega = \Omega_{\mathrm{M}} + \Omega_\Lambda \stackrel{?}{=} 1, \tag{11.36}$$

where $\Omega_\Lambda \equiv \rho_\Lambda/\rho_{\mathrm{c}}$. A nonvanishing Λ would also provide the repulsion to accelerate the expansion of the universe. In such an accelerating universe the expansion rate in the past must be smaller than the current rate H_0. This means that it would take a longer period (as compared to a decelerating or empty universe) to reach the present era, leading to a longer age $t_0 > 2t_{\mathrm{H}}/3$ even though the geometry is flat. This just might possibly solve the cosmic age problem as well.

11.4.1 Distant supernovae and the 1998 discovery

In order to obtain observational evidence for any changing expansion rate of the universe (i.e. to measure the curvature of the Hubble curve), one would have to measure great cosmic distances. One needed a distance method that works to over 5 billion light years. Clearly some very bright light sources are required. Since this also means that we must measure objects back in a time interval that is a significant fraction of the age of the universe, the method must be applicable to objects present at the early cosmic era. As it turns out, supernovae are ideally suited for this purpose.

SNe as standard candles and their systematic search

After the suggestion made in the 1970s that type Ia supernovae (SNe Ia) could possibly serve as standard candles, the first SN Ia was discovered in 1988 by a Danish group at redshift $z = 0.3$. At their peaks SNe Ia produce a million times more light than Cepheid variables, the standard candle most commonly used in cosmology (cf. Section 9.4.2). SNe Ia begin as white dwarfs (collapsed old stars sustained by degenerate pressure of their electrons) with mass comparable to the sun. If the white dwarf has a large companion star, which is not uncommon, the dwarf's powerful gravitational attraction will draw matter from its companion. Its mass increases until the "Chandrasekhar limit" $\simeq 1.4\, M_{\odot}$. As it can no longer be countered by the electron pressure, the gravitational contraction develops and the resultant heating of the interior core triggers the thermonuclear blast that rips it apart, resulting in an SN explosion. The supernova eventually collapses into a neutron star. Because they start with masses in a narrow range, such supernovae have comparable intrinsic brightness. Furthermore, their brightness has a characteristic decline from the maximum which can be used to improve on the calibration of their luminosity (the light-curve shape-analysis), making SNe Ia standardizable candles (Phillips 1993). Supernovae are rare events in a galaxy. The last time a supernova explosion occurred in our Milky Way was about 400 years ago. However, using new technology (large mosaic CCD cameras), astronomers overcame this problem by simultaneously monitoring thousands of galaxies[23] so that on the average some 10–20 supernovae can be observed in a year.

[23]Two images of the sky containing thousands of galaxies were taken weeks apart and digitally subtracted; the supernova locations leaped out.

The discovery of an accelerating universe

Because light from distant galaxies was emitted long ago, to measure a star (or a supernova) farther out in distance is to probe the cosmos further back in time. An accelerating expansion means that the expansion rate was smaller in the past. Thus to reach a given redshift (i.e. recession speed) it must be located farther away[24] than expected, see Fig. 11.7. Observationally, the light source

[24]A Hubble curve (as in Fig. 11.7) is a plot of the luminosity distance versus the redshift (measuring recession velocity). A straight Hubble curve means a cosmic expansion that is coasting. This can only happen in an empty universe (cf. Section 9.1.3 and Fig. 10.2). If the expansion is accelerating, the expansion rate H must be smaller in the past ($H < H_0$). From Eq. (9.5): $H\Delta r = z$, we see that, for a given redshift z, the distance Δr to the light-emitting supernova must be larger than that for an empty or decelerating universe.

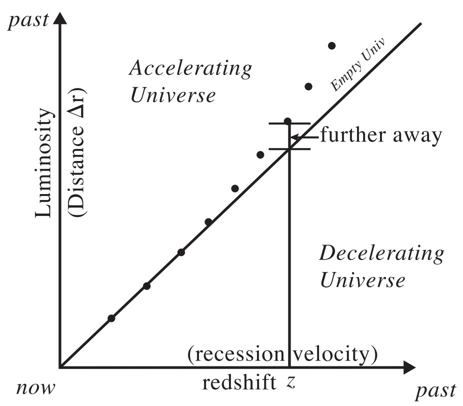


Fig. 11.7 Hubble diagram: the Hubble curve for an accelerating universe bends upwards. A supernova on this curve at a given redshift would be further out in distance than anticipated.

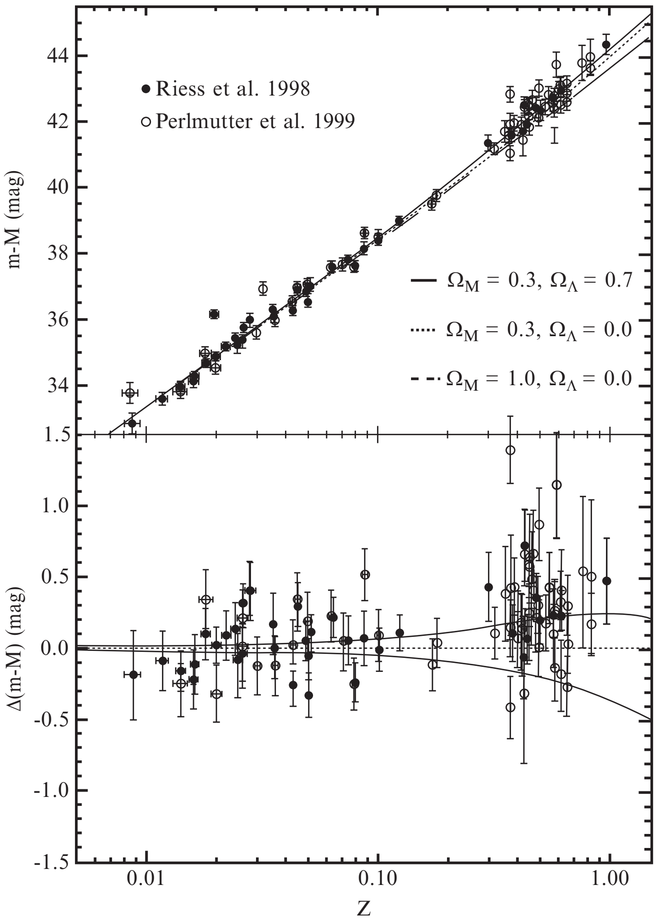


Fig. 11.8 Discovery of an accelerating universe. The Hubble plot showing the data points from Riess *et al.* (1998) and Perlmutter (1999). The horizontal axis is the redshift z; the vertical axes are the luminosity distance expressed in terms of distance modulus (i.e. logarithmic luminosity distance, cf. Box 9.1). In the lower panel $\triangle(m - M)$ is the difference after subtracting out the then expected value for a decelerating universe with $\Omega_M = 0.3$ and $\Omega_\Lambda = 0$. Three curves correspond to models with different matter/energy contents $(\Omega_M, \Omega_\Lambda)$ of the universe. The solid curve for nonvanishing cosmological constant has the best fit of the observational data.

in an accelerating universe would be measured to be dimmer than expected.

- By 1998 two collaborations: the Supernova Cosmological Project, led by Saul Perlmutter of the Lawrence Berkeley National Laboratory (Perlmutter *et al.*, 1999) and the High-z Supernova Search Team, led by Adam Riess of the Astronomy Department at UC Berkeley and Brian Schmidt of the Mount Stromlo and Siding Spring Observatories (Riess *et al.*, 1998), each had accumulated some 50 SNe Ia at high redshifts—z: 0.4–0.7 corresponding to SNe occurring five to eight billion years ago. They made the astonishing discovery that the expansion of the universe was actually accelerating, as indicated by the fact that the measured luminosities were on the average 25% less than anticipated, and the Hubble curve bent upward, Fig. 11.8.

Extracting Ω_M and Ω_Λ from the measured Hubble curve From the Hubble curve plotted in the space of redshift and luminosity distance, one can then extract the mass and dark energy content of the universe. The proper distance d_p from a supernova with a redshift z in the present epoch $a(t_0) = 1$ has been worked out in (9.51). Combined with the result in (9.57),

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 260 of 276 September 24, 2009 17:51

this yields an expression for the luminosity distance:

$$d_{\mathrm{L}}(z) = c(1+z)\int_0^z \frac{dz'}{H(z')}, \tag{11.37}$$

where, using the Friedmann equation (10.1), we can express the epoch-dependent Hubble constant in terms of the scale factor and the density ratios in the present epoch (Problem 11.2), including in particular the cosmological constant density term:

$$H(t) = H_0\left(\frac{\Omega_{\mathrm{R},0}}{a^4} + \frac{\Omega_{\mathrm{M},0}}{a^3} + \Omega_\Lambda + \frac{1-\Omega_0}{a^2}\right)^{1/2}, \tag{11.38}$$

where $a(t)$ can in turn be replaced by the redshift according to (9.50),

$$\begin{aligned} H(z) &= H_0 \Omega_{\mathrm{R},0}(1+z)^4 + \Omega_{\mathrm{M},0}(1+z)^3 + \Omega_\Lambda + (1-\Omega_0)(1+z)^{21/2} \\ &\simeq H_0 \Omega_{\mathrm{M},0}(1+z)^3 + \Omega_\Lambda + (1-\Omega_{\mathrm{M},0}-\Omega_\Lambda)(1+z)^{21/2}. \end{aligned} \tag{11.39}$$

The resultant Hubble curves $d_L(z)$ in (11.37) with $H(z)$ in the form of (11.39) that best fitted the observation data yields values of $\Omega_{\mathrm{M},0}$ and Ω_Λ as shown in Fig. 11.9. If we further impose the requirement of a flat geometry, $\Omega_{\mathrm{M},0} + \Omega_\Lambda = 1$ as suggested by the CMB data, the favored values from Fig. 11.9 as well as from other supporting evidence obtained later on are

$$\Omega_{\mathrm{M},0} = 0.246 \quad \text{and} \quad \Omega_\Lambda = 0.757 \tag{11.40}$$

suggesting that most of the energy in our universe resides in this mysterious dark energy.[25]

The age of universe calculated These observed values for $\Omega_{\mathrm{M},0}$ and Ω_Λ can also be translated into an age for the flat universe. The Hubble constant being the rate of expansion $H = \dot{a}/a$, we can relate dt to the differential of the scale factor,

$$t_0 = \int_0^{t_0} dt = \int_0^1 \frac{da}{aH}. \tag{11.41}$$

From (11.38) for the scale-dependent Hubble constant, this yields an expression of the age[26] in terms of the density parameters

$$t_0 = t_{\mathrm{H}} \int_0^1 \frac{da}{\Omega_{\mathrm{R},0}a^{-2} + \Omega_{\mathrm{M},0}a^{-1} + \Omega_\Lambda a^2 + (1-\Omega_0)^{1/2}}. \tag{11.42}$$

The spatially flat universe with negligible amount of radiation energy, $\Omega_0 = \Omega_{\mathrm{M},0} + \Omega_\Lambda = 1$, leads to a simple expression of the age of the universe in terms of the densities

$$\begin{aligned} \frac{t_0}{t_{\mathrm{H}}} &= \int_0^1 \left(\Omega_{\mathrm{M},0}a^{-1} + \Omega_\Lambda a^2\right)^{-1/2} \\ da &= \frac{2}{3\sqrt{\Omega_\Lambda}} \ln \frac{\sqrt{\Omega_\Lambda} + \sqrt{\Omega_{\mathrm{M},0} + \Omega_\Lambda}}{\sqrt{\Omega_{\mathrm{M},0}}} = 1.02. \end{aligned} \tag{11.43}$$

[25] In the present discussion we shall for definiteness assume the dark energy as being the cosmological constant with an equation of state parameter $w = -1$.

[26] We can check the limit of (11.42) for a matter-dominated flat universe ($\Omega_{\Lambda,0} = \Omega_{R,0} = 0$ with $\Omega_0 = \Omega_{M,0} = 1$) which yields an age $t_0 = t_H \int_0^1 a^{1/2} da = \frac{2}{3} t_H$, in agreement with the result obtained in (10.30).

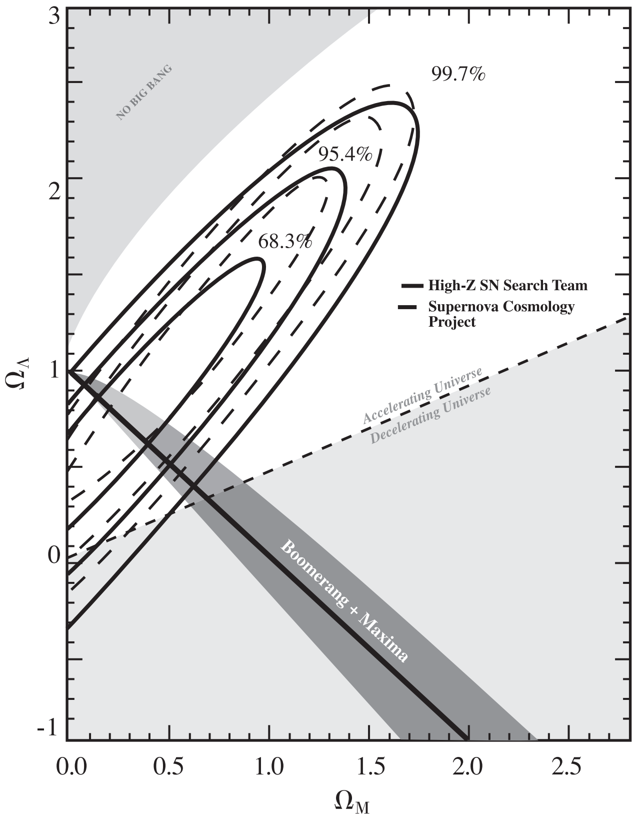


Fig. 11.9 Fitting Ω_Λ and Ω_M to the discovery data as obtained by the High-z SN Search Team and Supernova Cosmology Project. The favored values of Ω_Λ and Ω_M follow from the central values of CMB anisotropy $\Omega_\Lambda + \Omega_M \simeq 1$ (the straight line) and those of the SNe data represented by confidence contours (ellipses) around $\Omega_\Lambda - \Omega_M \simeq 0.5$.

For the density values given in (11.40), the RHS comes very close to unity. Thus the deceleration effect of $\Omega_{\mathrm{M},0}$ and the accelerating effect of Ω_Λ coincidentally cancel each other. The age is very close to that of an empty universe.

$$t_0 = 1.02\, t_{\mathrm{H}} = 13.9\,\mathrm{Gyr}. \tag{11.44}$$

11.4.2 Transition from deceleration to acceleration

Since the immediate observational evidence from these far away supernovae is a smaller-than-anticipated luminosity, one wonders whether there is a more mundane astrophysical explanation. There may be one or a combination of several mundane causes that can mimic the observational effects of an accelerating universe. Maybe this luminosity diminution is brought about not because the supernovae were further away than expected, but due to the absorption by yet-unknown[27] interstellar dust, and/or due to some yet-unknown evolution of supernovae themselves (i.e. supernovae's intrinsic luminosity were smaller in the cosmic past). However, all such scenarios would lead us to expect that the supernovae, at even greater distances (and even further back in time), should have their brightness continue to diminish.

[27]The absorption and scattering by ordinary dust shows a characteristic frequency dependence that can in principle be subtracted out. By the unknown dust we refer to any possible "gray dust" that could absorb light in a frequency-independent manner.

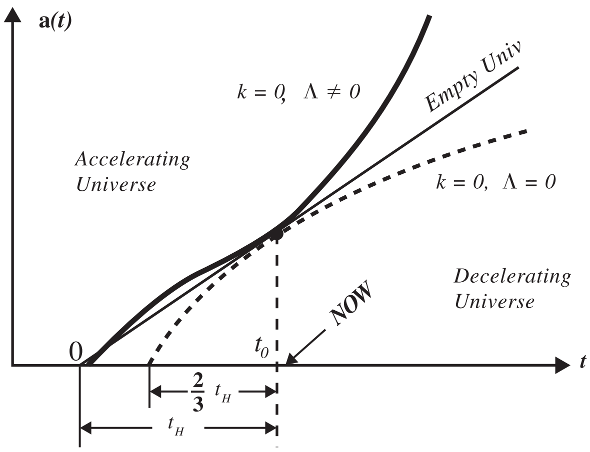


Fig. 11.10 Time evolution of an accelerating universe. It started out in a decelerating phase before taking on the form of an exponential expansion. The transition to an accelerating phase shows up as a "bulge;" this way it has an age longer than the $\Lambda = 0$ flat universe age of $2t_H/3$.

For the accelerating universe, on the other hand, this diminution of luminosity **would stop**, and the brightness would **increase** at even larger distances. This is so because we expect the accelerating epoch be proceeded by a decelerating phase. The dark energy should be relatively insensitive to scale change $\rho_\Lambda \sim a^0(t)$ (the true cosmological constant is a constant density, independent of scale change), while the matter or radiation energy densities, $\rho \sim a^{-3}(t)$ or $a^{-4}(t)$, should be more and more important in earlier times. Thus, the early universe could not be dark energy dominated, and it must be decelerating. This transition from a decelerating to an accelerating phase would show up as a bulge in the Hubble curve, see Fig. 11.10.

The cosmic age at transition Let us estimate the redshift when the universe made this transition. We define an epoch-dependent **deceleration parameter** which generalizes the q_0 parameter of Problem 9.10,

$$q(t) \equiv \frac{-\ddot{a}(t)}{a(t)H^2(t)}, \tag{11.45}$$

which, through the Friedmann equation, can be related to the density ratios (Problem 10.10)

$$\begin{aligned} q(t) &= \Omega_R(t) + \frac{1}{2}\Omega_M(t) - \Omega_\Lambda \\ &= \frac{\Omega_{R,0}}{a(t)^4} + \frac{\Omega_{M,0}}{2a(t)^3} - \Omega_\Lambda. \end{aligned} \tag{11.46}$$

After dropping the unimportant $\Omega_{R,0}$ and replacing the scale factor by z, we have

$$q(z) \simeq \frac{1}{2}\Omega_{M,0}(1+z)^3 - \Omega_\Lambda. \tag{11.47}$$

The transition from decelerating ($q > 0$) to the accelerating ($q < 0$) phase occurred at redshift z_{tr} when the deceleration parameter vanished $q(z_{tr}) \equiv 0$,

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 263 of 276 September 24, 2009 17:51

or

$$1 + z_{\mathrm{tr}} = \left(\frac{2\Omega_\Lambda}{\Omega_{\mathrm{M},0}}\right)^{1/3}. \tag{11.48}$$

The supernovae data translate into a transition redshift of $z_{\mathrm{tr}} \simeq 0.8$, corresponding to a scale factor of $a_{\mathrm{tr}} \simeq 0.56$ and a cosmic time, calculated similarly[28] as in (11.42) and (11.43), of $t_{\mathrm{tr}} = t(a = 0.56) \simeq 7\,\mathrm{Gyr}$—in cosmic terms, the transition took place only recently ("just yesterday")! This reflects the fact that the matter density in the present epoch $\Omega_{\mathrm{M},0}$ happens to be comparable to the dark energy density Ω_Λ.

[28]The relation between cosmic time and scale factor of a given epoch is

$$t(a) = t_{\mathrm{H}} \int_0^a \Omega_{\mathrm{M},0}/a' + \Omega_\Lambda a'^{2-1/2} da'$$

$$= \frac{2t_{\mathrm{H}}}{3\sqrt{\Omega_\Lambda}} \ln\left(\sqrt{\frac{\Omega_\Lambda}{\Omega_{\mathrm{M},0}} a^3} + \sqrt{1 + \frac{\Omega_\Lambda}{\Omega_{\mathrm{M},0}} a^3}\right).$$

Discovery of SNe prior to the accelerating phase Thus, the conclusive evidence for the accelerating universe interpretation of the supernovae data is to observe this bulge structure, which cannot be mimicked by any known astrophysical causes. The 1998 discovery data (z: 0.4–0.7) showed the rise of this bulge, but we need to see the falling part of the Hubble curve. SNe further out ($z > 0.8$) should be still in the decelerating phase; they should be brighter than what is expected of the continuing dimming scenario that a mundane interpretation would have us anticipate. Reassuringly, just such an early decelerating phase had been detected.

After the original discovery of an accelerating universe, researchers had searched for other supernovae at high z. The supernova labeled SN1997ff had been serendipitously recorded by the Hubble Space Telescope, and by other observational means (some intentionally, and some unpremeditated). Through a major effort at data analysis, its properties were deduced in 2001, showing that it is a type Ia SN having a redshift of $z \simeq 1.7$ and, thus an explosion occurring 10 billion years ago, making it by far most distant supernova ever detected. Remarkably, it is brighter by almost a factor of two (see Fig. 11.11) compared to the expectation of continual dimming as a mundane astrophysical explanation would require. This is the bulge feature unique to a Hubble curve for an accelerating universe—the light was emitted so long ago when the

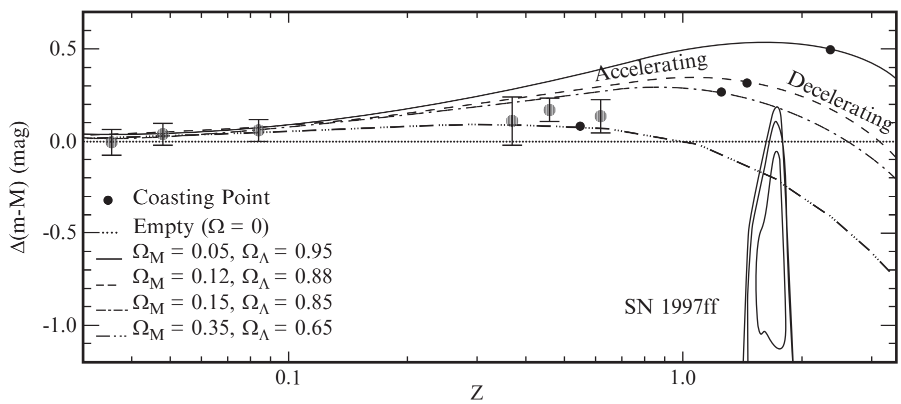


Fig. 11.11 Location of SN1997ff (because of measurement uncertainties, shown as a patch on the right side of the diagram) and other high z SNe are plotted with respect to those for an empty universe (the horizontal line) in a Hubble diagram. The black spots follow an up-turning curve which represents the luminosity and redshift relation showing continuing dimming.

expansion of the universe was still decelerating. Since this first confirmation finding, many more high-z SNe had been observed both from ground-based surveys and with Hubble Space Telescope. These data had provided conclusive evidence for cosmic deceleration that preceded the present epoch of cosmic acceleration (Riess *et al.*, 2004).

11.4.3 Dark energy: Further evidence and the mystery of its origin

After the supernovae discovery, the presence of dark energy $\Omega_\Lambda = 0.75$ was further confirmed by the analysis of the CMB anisotropy power spectrum, as well as of the distribution of galaxies. In the following paragraph we shall discuss the recent result showing that such a dark energy is just the agent needed to explain the observed slow down of galaxies' growth. On the other hand, even though the observational evidence for dark energy is strong, its physical origin remains mysterious. The quantum vacuum energy does have precisely the property of a density being constant with respect to volume changes and hence being a negative pressure, the estimated magnitude of such a vacuum quantum energy is something like 120 orders of magnitude too large. In this section we give a brief summary of these developments; more details can be found in a recent review (Frieman, Turner and Huterer, 2008).

Dark energy stunts the growth of galaxies We have calculated the cosmic time $t_{\rm tr} \simeq 7$ Gyr when the deceleration phase begin to be replaced by the acceleration phase, and the cosmic time $t_{\rm M\Lambda} \simeq 9.5$ Gyr when the energy content of the universe was just balanced between matter and dark energy (see Problem 11.5). Namely, only in the last 5–7 Gyr or so has the dark energy become the dominant force in the universe that turned the (decelerating) expansion of the universe into an accelerated expansion. Such an effect of having the repulsive gravity of dark energy overcoming the more familiar gravitational attraction also shows up in the slowing down of the growth of the largest conglomeration of matter in the universe, the galaxy clusters. They are relatively easy to find as clusters are filled with hot gas that emits X-rays. A group researcher (Vikhlinin *et al.,* 2009) used the Chandra X-ray satellite telescope to study the intensity and spectra of 86 clusters that had previously been found by the ROSAT (X-ray) All-Sky Survey. A set of 37 clusters at more than 5 gigalight-years away was compared with another set of 49 that are closer than half a gigalight-year. Theoretical models were used to calculate how the numbers of clusters with different masses would change during this span of $\Delta t \simeq 5$ Gyr under different conditions: with different amounts, and values of the equation of state parameter w, and with or without a dark energy. A good fit to the observation data[29] clearly required the presence of dark energy with $w \simeq -1$.

[29] For example, Vikhlinin *et al.* (2009) found only a fifth of the number of the most massive clusters that a universe without dark energy would have.

The problem of interpreting Λ as quantum vacuum energy The introduction of the cosmological constant in the GR field equation does not explain its physical origin. In the inflation model one postulates that it is the false vacuum energy of an inflation/Higgs field that acts like an effective

Table 11.1 Cosmological parameters deduced from an analysis (Tegmark 2006) based on data collected by WMAP and SDSS. The first column displays the parameter result from an analysis that assumes the cosmological constant as being the dark energy; the second column for a universe assumed to be flat. The equation numbers in the third column refer to part of the text, where such parameters were discussed. The parameter h_0 is the Hubble constant H_0 measured in units of 100 (km/s) /Mpc.

	$\begin{pmatrix} DE=\Lambda \\ w=-1 \end{pmatrix}$	$\begin{pmatrix} FlatU \\ \Omega_0=1 \end{pmatrix}$	Parameter description (equation number)
Ω_0	1.003 ± 0.010	1 (fixed)	density parameter (11.33)
Ω_Λ	0.757 ± 0.021	0.757 ± 0.020	dark energy density (11.40)
Ω_M	0.246 ± 0.028	0.243 ± 0.020	matter density (9.23, 11.40)
Ω_B	0.042 ± 0.002	0.042 ± 0.002	baryon density (10.58)
h_0	0.72 ± 0.05	0.72 ± 0.03	present expansion rate (9.7)
t_0	13.9 ± 0.6 Gyr	13.8 ± 0.2 Gyr	age of the universe (11.44)
T_0	2.725 ± 0.001 K	2.725 ± 0.001 K	CMB temperature (10.64)
q_0	-0.63 ± 0.03	-0.57 ± 0.1	deceleration parameter (11.45)
w	-1 (fixed)	-0.94 ± 0.1	dark energy equation of state (10.4)

cosmological constant driving the inflationary expansion. What is the physical origin of the dark energy that brings about the accelerating expansion of the present epoch? A natural candidate is the quantum vacuum energy. The zero point energy of a quantized field automatically has the property of having an energy density that is constant, giving rise to a negative pressure. However as explained in Appendix D (Section 11.7) such a vacuum energy while having the correct property is expected to be way too large to account for the observed $\Omega_\Lambda = O(1)$. What are the other possibilities? One chance is that the dark energy is associated with some yet-unknown scalar field (sometimes referred to as the "quintessence"), somewhat akin to the association of the inflationary expansion to the inflation/Higgs field. Such theories often have an equation-of-state parameter $w \neq w_\Lambda = -1$. However, observational data do not support a dark energy w significantly different from the value of -1 (see Table 11.1).

11.5 The concordant picture

An overall coherent and self-consistent picture of the cosmos has emerged that can account for the geometry and structure of the universe, as well as its evolution onward from a fraction of a second after the big bang. In this section, we first summarize the cosmological parameters and discuss the concordant cosmological model that has emerged. Even though we have a consistent picture, there are still many unsolved problems; we shall mention some of them at the end of this chapter.

Cosmological parameters from CMB and the galaxy distribution

Our previous discussion has concentrated on conceptually and technically simpler approaches in obtaining cosmological parameters—counting and weighing methods, plotting the Hubble curve (including data from high-redshift supernovae), and light nuclear element abundance, etc. These measurements have now been confirmed and hugely improved by the analysis of very different physical phenomena: the CMB temperature anisotropy (in particular as measured by WMAP) in combination with analysis of large-scale structure survey

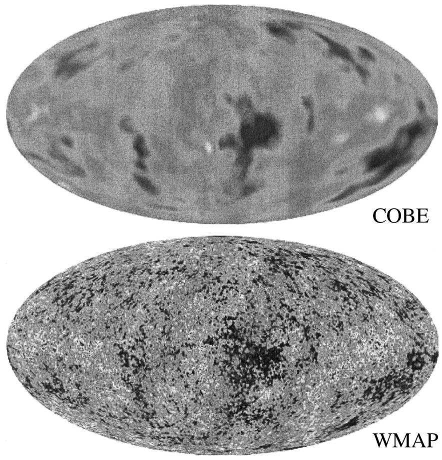


Fig. 11.12 The temperature fluctuation of the CMB is a snap-shot of the baby universe at the photon decoupling time. A comparison of the results by COBE vs. WMAP1 shows the marked improvement in resolution by WMAP. This allowed us to extract many more cosmological parameters from the latest observations.

data (obtained in particular by 2dF and SDSS). We have briefly discussed the CMB anisotropy (cf. Sections 10.5.4 and 11.3): a detailed study of the power spectrum through a spherical harmonics decomposition can be displayed as a curve (relative amplitude vs. angular momentum number) with a series of peaks. The primary peak (i.e. the dominant structure) is at the one degree scale showing that the spatial geometry is flat; the secondary peaks are sensitive to other cosmological parameters such as the baryon contents of the universe, $\Omega_B \simeq 0.04$, etc. WMAP has a much improved angular resolution compared to COBE, Fig. 11.12, The study of the large-scale cosmic distribution of galaxies is beyond the scope of this book, the underlying physics also reflects the relic imprints of primordial acoustic waves as in the case of CMB. The combined results of this array of observations allowed us to extract a large number of cosmological parameters at high accuracy (Table 11.1).

The standard model of cosmology

Cosmology has seen a set of major achievements over the past decade, to the extent that something like a standard model for the origin and development of the universe is now in place: the FLRW cosmology proceeded by an inflationary epoch. Many of the basic cosmological parameters have been deduced in several independent ways, arriving at a consistent set of results. These data are compatible with our universe being infinite and spatially flat, having matter/energy density equal to the critical density, $\Omega_0 = 1$. The largest energy component is consistent with it having Einstein's cosmological constant $\Omega_\Lambda \simeq 0.75$. In the present epoch this dark energy content is comparable in size to the matter density $\Omega_M \simeq 0.25$, which is made up mostly of cold dark matter. Thus this standard model is often call the ΛCDM cosmology model. The expansion of the universe will never stop—in fact having entered the accelerating phase, the expansion will be getting faster and faster.

Still many unsolved problems

Although we have a self-consistent cosmological description, many mysteries remain. We do not really know what makes up the bulk of the dark

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 267 of 276 September 24, 2009 17:51

matter, even though there are plausible candidates as predicted by some yet-to-be-proven particle physics theories. The most important energy component is the mysterious "dark energy," although a natural candidate is the quantum vacuum energy. Such an identification leads to an estimate of its size that is completely off the mark (cf. Section 11.7). If one can show that the quantum vacuum energy must somehow vanish due to some yet-to-be-found symmetry principle, a particular pressing problem is to find out whether this dark energy is time-independent, as is the case of the cosmological constant Λ, or is it more like a Λ_{eff} coming from some quintessence scalar field as in the case of inflation? Despite our lack of understanding of this dark energy, in recent discoveries constitute a remarkable affirmation of the inflationary theory of the big bang. Still, even here question remains as to the true identity of the inflation/Higgs field. We need to find ways to test the existence of such a field in some noncosmological settings.

Besides the basic mystery of dark energy ("the cosmological constant problem") there are other associated puzzles, one of them being the "cosmic coincidence problem:" we have the observational result that in the present epoch the dark energy density is comparable to the matter density, $\Omega_\Lambda \simeq \Omega_M$. Since they scale so differently ($\Omega_M \sim a^{-3}$ vs. $\Omega_\Lambda \sim a^0$) we have $\Omega_M \simeq 1$ in the cosmic past, and $\Omega_\Lambda \simeq 1$ in the future. Thus, the present epoch is very special—the only period when they are comparable.[30] Then the question is why? How do we understand this requirement of fine tuning the initial values in order to have $\Omega_M \simeq \Omega_\Lambda$ now?

[30] Closely related to this is the puzzle of the respective amounts of decelerating matter and the accelerating dark energy so that their effects cancel each other, leaving the age of the universe very close to that of an empty universe (the Hubble time).

A finite dodecahedral universe: A cautionary tale

It cannot be emphasized too much that the recent spectacular advances in cosmology have their foundation in the ever-increasing amount of high precision observational data. Ultimately any cosmological theory will stand or fall, depending on its success in confronting experimental data. In this context we offer the following cautionary tale.

An inspection of the CMB power spectrum in Fig. 11.13 shows that a few data points in the large angle (low l) region tend to be lower than the theoretical curve based on the standard cosmological model outlined above.

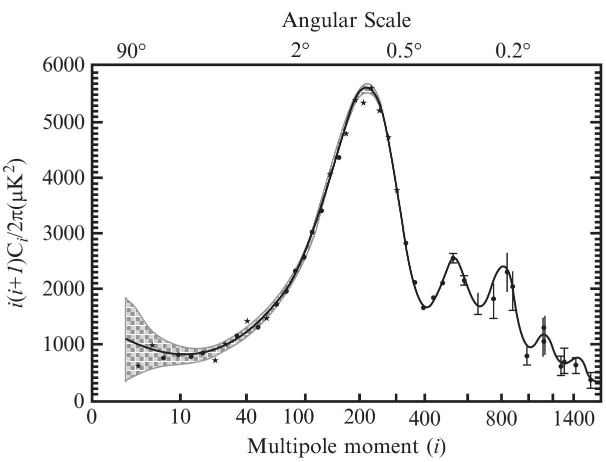


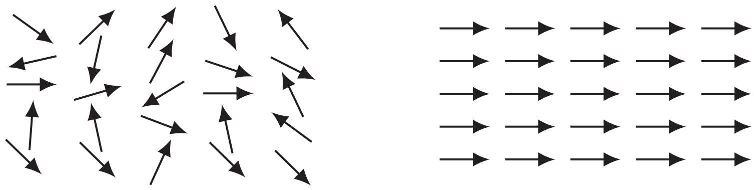
Fig. 11.13 The angular power spectrum of CMB temperature anisotropy. The dots are the first-year data-points from WMAP. The theoretical curve follows from inflationary model (having cold dark matter) with parameters given in Table 11.1. The fan-shaped shaded area at low multiple moments reflects the uncertainty due to cosmic variance, cf. (10.89).

This does not concern most cosmologists because they are still in the shaded area corresponding to the statistical uncertainty called cosmic variance cf. (10.89). Nevertheless, it is possible to interpret these low data points as potential signature of a finite universe. The weakness of the quadrupole ($l = 2$) and octupole term ($l = 3$) can be taken as a lack of temperature correlation on scales greater than 60°. Maybe the space is not infinite and the broadest waves are missing because space is not big enough to accommodate them. Our discussion above has shown the evidence for the space being locally homogenous and isotropic. However, local geometry constrains, but does not dictate, the shape of the space. Thus, it is possible that the topology of the universe is nontrivial. Luminet *et al.* (2003) constructed just such a model universe based on a finite space with a nontrivial topology (the Poincaré dodecahedral space). It has a positive curvature (closed universe) with $\Omega_0 = 1.013$, which is compatible with observation as of 2003. One of the ways to study the shape, or topology, of the universe is based on the idea that if the universe is finite, light from a distant source will be able to reach us along more than one path. This will produce matching images (e.g. circles) in the CMB anisotropy. A search for such matching circles failed to find such features (Cornish *et al.,* 2004). Thus, this finite universe model may, in the end, be ruled out by observation.

Our purpose in reporting this particular episode in the cosmological study is to remind ourselves of the importance of keeping an open mind of alternative cosmologies. This example showed vividly how drastically different cosmological pictures can be based on cosmological parameters that are not that different from each other. Thus, when looking at a result such as $\Omega_0 = 1.03 \pm 0.03$, as known then in 2003, we should refrain from jumping to the conclusion that data has already shown a $\Omega_0 = 1$ flat universe. This shows the importance of acquiring high precision data, which will ultimately decide which model gives us the true cosmology. On the other hand, while a slight change of one or two parameters may favor different cosmological models, it is the overall theoretical consistency, the ability to account for a whole array of data in cosmology and robust in its cross-checks that ultimately allows us to believe that the current concordant picture has a good chance to survive future experimental tests.

11.6 Appendix C: False vacuum and hidden symmetry

In Section 11.2.2 we discussed the theoretical suggestion that the cosmological inflationary epoch is associated with a “false vacuum” of an inflation/Higgs field. This involves the concept of a “spontaneous breakdown of a symmetry,” also described as a “hidden symmetry”—even though a theory is symmetric, its familiar symmetry properties are hidden. Namely, a symmetric theory somehow ends up having asymmetrical solutions. This can happen, as we shall see, when there are “degenerate ground states” —an infinite number of theoretically possible states (related to each other by symmetry transformations)



(a) (b)

Fig. 11.14 (a) Ground state with zero magnetization $\vec{\mathcal{M}}_0 = 0$ for randomly oriented dipoles, (b) asymmetric ground state with $\vec{\mathcal{M}}_0 \neq 0$.

all having the same lowest energy. But the physical vacuum is one of this set, and, by itself, it is not symmetric because it singles out a particular direction in the symmetry space. In this Appendix, we illustrate this phenomenon by the example of the breakdown of rotational symmetry in a ferromagnet near the Curie temperature.

Hidden rotational symmetry in a ferromagnet A ferromagnet can be thought of as a collection of magnetic dipoles. When it is cooled below a certain critical temperature, the Curie temperature T_c, it undergoes spontaneous magnetization: all its dipoles are aligned in one particular direction (a direction determined not by dipole interactions, but by external boundary conditions). Namely, when $T > T_c$ the ground state has zero magnetization $\vec{\mathcal{M}}_0 = 0$ because the dipoles are randomly oriented; but below the critical temperature $T < T_c$, all the dipoles line up, giving arise to a nonzero magnetization $\vec{\mathcal{M}}_0 \neq 0$ (Fig. 11.14). This can happen even though the underlying dynamics of dipole–dipole interaction is rotationally symmetric—no preferred direction is built into the dynamics, that is, the theory has rotation symmetry.

Ginzburg and Landau description of spontaneous symmetry breaking For a mathematical description we shall follow the phenomenological theory of Ginzburg and Landau. When $T \approx T_c$, the rotationally symmetric free energy $\mathcal{F}(\vec{\mathcal{M}})$ of the system can be expanded in a power series of the magnetization $\vec{\mathcal{M}}$:

$$\mathcal{F}(\vec{\mathcal{M}}) = \left(\nabla_i \vec{\mathcal{M}}\right)^2 + \underbrace{a(T)(\vec{\mathcal{M}} \cdot \vec{\mathcal{M}}) + b(\vec{\mathcal{M}} \cdot \vec{\mathcal{M}})^2}_{V(\vec{\mathcal{M}})}. \tag{11.49}$$

In the potential energy function $V(\vec{\mathcal{M}})$ we have kept the higher order $(\vec{\mathcal{M}} \cdot \vec{\mathcal{M}})^2$ term, with a coefficient $b > 0$ (as required by the positivity of energy at large $\mathcal{M}$), because the coefficient a in front of the leading $(\vec{\mathcal{M}} \cdot \vec{\mathcal{M}})$ term can vanish: $a(T) = \gamma(T - T_c)$. With γ being some positive constant, the temperature-dependent coefficient a is positive when $T > T_c$, negative when $T < T_c$. Since the kinetic energy term $(\nabla_i \vec{\mathcal{M}})^2$ is nonnegative, to obtain the ground state, we need only to minimize the potential energy:

$$\frac{dV}{d\vec{\mathcal{M}}} \propto \vec{\mathcal{M}} \left[a + 2b\left(\vec{\mathcal{M}} \cdot \vec{\mathcal{M}}\right)\right] = 0. \tag{11.50}$$

The solution of this equation gives us the ground state magnetization $\vec{\mathcal{M}}_0$. For $T > T_c$, hence a positive a, we get the usual solution of a zero magnetization $\vec{\mathcal{M}}_0 = 0$ (i.e. randomly oriented dipoles). This situation is shown in the plot of

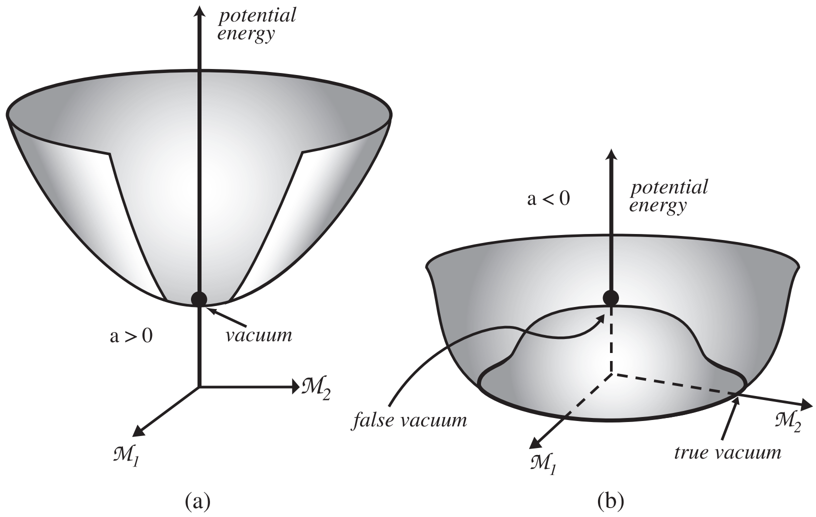


Fig. 11.15 Symmetric potential energy surfaces in the field space: (a) the normal solution, when the ground state is at a symmetric point with $\mathcal{M}_0 = 0$, and (b) the broken symmetry solution, when the energy surface has the shape of a "Mexican hat" with $\mathcal{M} = 0$ being a local maxinum and the true ground state being one point in the trough (thus singling out one direction and breaking the rotational symmetry).

$V(\vec{\mathcal{M}})$ of Fig. 11.15(a), where the potential energy surface is clearly symmetric with any rotation (in the 2D plane[31]) around the central axis. However, for subcritical temperature $T < T_c$, the sign change of a brings about a change in the shape of the potential energy surface as in Fig. 11.15(b). The surface remains symmetric with respect to rotation, but the zero magnetization point $\mathcal{M} = 0$ is now a local maximum. There is an infinite number of theoretically possible ground states at the bottom ring of the wine-bottle shaped surface—all having nonzero magnetization $\mathcal{M}_0 = \sqrt{-a/2b}$, but pointing in different directions in the 2D field space. These possible ground states are related to each other by rotations. The physical ground state, picked to be one of them by external conditions, singles out one specific direction, and hence is not rotationally symmetric. Below the Curie temperature, rotational symmetry in the ferromagnet is spontaneously broken and the usual symmetry properties of the underlying dynamics (in this case, rotational symmetry) are not apparent. We say spontaneous symmetry breaking corresponds to a situation of **hidden symmetry**.

[31] We have simplified the display to the case when $\vec{\mathcal{M}}$ is a 2D vector in a plane having two components $\mathcal{M}_1$ and $\mathcal{M}_2$.

Higgs/inflation field In particle physics we have a system of fields. In particular it is postulated that there are scalar fields (for particles with zero spin) which have potential energy terms displaying the same spontaneous symmetry properties as ferromagnetism near T_c. The magnetization $\vec{\mathcal{M}}$ in (11.49) is replaced, in the case of particle physics, by a scalar field $\phi(x)$. Thus at high energy (i.e. high temperature) the system is in a symmetric phase (normal solution with $a(T) > 0$) and the unification of particle interactions is manifest (cf. Section 11.2.2, see in particular sidenote 6); at lower energy (low temperature) the system enters a broken symmetry phase because of $a(T) < 0$. The ground state of a field system is, by definition, the vacuum. In this hidden symmetry phase we have a nonvanishing scalar field $\phi_0(x) \neq 0$. The relevance to cosmology is as follows: at higher temperature we have a symmetric vacuum. When the universe cools below the critical value, the same state becomes a local maximum and is at a higher energy than, and begins to roll toward, the true vacuum. We say the system (the universe) is temporarily, during the rollover period, in a false vacuum (cf. Fig. 11.3). This semiclassical description indicates the existence of a constant field $\phi_0(x) \neq 0$ permeating everywhere in the universe.

11.7 Appendix D: Quantum vacuum energy as the cosmological constant

We associate the quantum vacuum energy with the dark energy that drives the accelerating universe, while it is the false vacuum energy of a Higgs field that supposedly brings about the primordial inflation. While these two mechanisms may well be related, our presentation assumes that they are separate. In the previous section we discussed the scalar field that can give rise to an inflationary exponential expansion; here we shall concentrate our discussion of the dark energy associated with the accelerating universe. As we shall explain, the zero-point energy of the quantum fields is a natural candidate for such a cosmological constant Λ. However the difficulty of such an association is that the natural size of such a quantum vacuum energy is much too large to account for the observed value of $\Omega_\Lambda \simeq 0.75$. We also briefly note that quantum vacuum energies of boson and fermion fields have opposite signs. Had our universe obeyed supersymmetry exactly with a strict degeneracy of bosonic and fermionic degrees of freedom, their respective contributions to the vacuum energy would exactly cancel, leading to a vanishing vacuum energy and cosmological constant.

Quantum vacuum energy gives rise to a cosmological constant

From the view point of quantum field theory, a vacuum state is not simply "nothingness." The uncertainty principle informs us that the vacuum has an energy because any localization has an associated spread in the momentum value. In fact, quantum field theory pictures the vacuum (defined as the state of lowest energy) as a sea of sizzling activities with constant creation and annihilation of particles. Thus, the cosmological constant Λ, as the energy density of the vacuum, naturally has a non-zero value (Zel'dovich 1968).

The simplest way to see that a quantum vacuum state has energy is to start with the observation that the normal modes of a field are simply a set of harmonic oscillators.[32] Summing over the quantized oscillator energies of all the modes, we have[33] (with the occupation number of the ith state denoted by n_i)

$$E_{\rm b} = \sum_i \left(\frac{1}{2} + n_i\right)\hbar\omega_i, \qquad \text{with } n_i = 0, 1, 2, 3, \ldots \tag{11.51}$$

From this we can identify the vacuum energy (also called the zero-point energy) as

$$E_\Lambda = \sum_i \frac{1}{2}\hbar\omega_i. \tag{11.52}$$

[32] We recall that the Fourier coefficients of, say, an electromagnetic field obey the simple harmonic oscillator equations.

[33] The subscript b stands for "boson;" this bosonic contribution will be compared to that by a fermion field in the last subsection.

At the atomic and subatomic levels, there is abundant empirical evidence for the reality of such a zero-point energy. For macroscopic physics, a notable manifestation of the vacuum energy is the Casimir effect, which has been verified experimentally.

The zero-point energy has the key property of having a density that is unchanged with respect to any volume changes, i.e. an energy density that is constant. The summation of the mode degrees of freedom in Eq. (11.52) involves the enumeration of the phase space volume in units of Planck's constant $\Sigma_i = \int d^3x d^3p\, (2\pi\hbar)^{-3}$, cf. Eq. (11.53). Since the zero-point energy $\hbar\omega_i = E\,(p)$ has no dependence on position, one obtains a simple volume factor $\int d^3x = V$ so that the corresponding energy per unit volume $E_\Lambda V^{-1}$ is a constant with respect to changes in volume. As explained in Section 11.1, this constant energy density corresponds to a negative pressure $(p = -E/V)$ and implies a $-\partial E/\partial x$ force that is attractive, pulling-in the piston in Fig. 11.1. This is the key property of the cosmological constant and is the origin of the Casimir effect—an attractive force between two parallel conducting plates.

Quantum vacuum energy is 10^{120}-fold too large as dark energy

Nevertheless, a fundamental problem exists because the natural size of a quantum vacuum energy is enormous. Here is a simple estimate of the sum in (11.52). The energy of a particle with momentum p is $E\,(p) = \sqrt{p^2c^2 + m^2c^4}$, see (3.37). From this we can calculate the sum by integrating over the momentum states to obtain the vacuum energy/mass density,

$$\rho_\Lambda c^2 = \frac{E_\Lambda}{V} = \int_0^{E_{\rm Pl}/c} \frac{4\pi p^2 dp}{(2\pi\hbar)^3}\left(\frac{1}{2}\sqrt{p^2c^2 + m^2c^4}\right), \tag{11.53}$$

where $4\pi p^2 dp$ is the usual momentum phase space volume factor. The integral in (11.53) would be a divergent quantity had we carried the integration to its infinity limit. Infinite momentum means zero distance; infinite momentum physics means zero distance scale physics. Since we expect spacetime to be quantized at the Planck scale (cf. quantum gravity in Section 8.5.1), it seems natural that we should cut off the integral at the Planck momentum $p_{\rm Pl} = E_{\rm Pl}/c$ as any GR singularities are expected to be modified at the short distance of the Planck length. The Planck momentum being given in Eq. (8.75),

$$p_{\rm Pl} = \frac{E_{\rm Pl}}{c} = \sqrt{\frac{\hbar c^3}{G_{\rm N}}} \simeq 10^{19}\,{\rm GeV}/c. \tag{11.54}$$

the integral (11.53) yields

$$\rho_\Lambda c^2 \cong \frac{1}{16\pi^2}\frac{E_{\rm Pl}^4}{(\hbar c)^3} \simeq \frac{\left(3\times 10^{27}\,{\rm eV}\right)^4}{(\hbar c)^3} \tag{11.55}$$

Since the critical density of (9.17) may be written in such natural units[34] as:

$$\rho_{\rm c} c^2 \cong \frac{\left(2.5\times 10^{-3}\,{\rm eV}\right)^4}{(\hbar c)^3},$$

[34]Since the natural "conversion constant" $\hbar c$ has the unit of length times energy, the combination (energy)$^4/(\hbar c)^3$ has the correct unit of energy per volume.

we have a quantum vacuum energy density ratio $\Omega_\Lambda \equiv \rho_\Lambda/\rho_c$ that is more than a factor of 10^{120} larger than the observed value of dark energy density,

$$(\Omega_\Lambda)_{qv} \simeq 10^{120} \quad \text{vs.} \quad (\Omega_\Lambda)_{obs} \simeq 0.75.$$

Thus if the observed dark energy originates from quantum vacuum energy, it must involve some mechanism to reduce its enormous natural size down to the critical density size. This "cosmological constant problem," in this context, is the puzzle why and how such a fantastic cancellation takes place —a cancellation of the first 120 significant figures (yet stops at the 121st place)!

Partial cancellation of boson and fermion vacuum energies

We should note that in the above calculation, we have assumed that the field is a boson field (such as the photon and graviton fields), having integer spin and obeying Bose–Einstein statistics. The oscillator's creation and annihilation operators obey commutation relations, leading to symmetric wavefunctions. On the other hand, fermions (such as electrons, quarks, etc.) have half-integer spins, and obey Fermi–Dirac statistics. Their fields have normal modes behaving like **Fermi oscillators**. The corresponding creation and annihilation operators obey anticommutation relations, leading to antisymmetric wavefunctions. Such oscillators have a quantized energy spectrum as (see, for example, Das, 1993)

$$E_f = \sum_i \left(-\frac{1}{2} + n_i\right) \hbar\omega_i, \qquad \text{with } n_i = 0 \text{ or } 1 \text{ only.} \tag{11.56}$$

For a fermion field, the zero point energy is negative! Therefore, there will be a cancellation in the contributions by bosons and fermions.[35] Many of the favored theories to extend the Standard Model of particle physics to the Planck scale incorporate the idea of **supersymmetry**. In such theories the bosonic and fermionic degrees of freedom are equal. In fact, the vacuum energy of systems with exact supersymmetry must vanish (i.e. an exact cancellation). However, we know that in reality supersymmetry cannot be exact[36] because its implication of equal boson and fermion masses $m_f = m_b$ in any supersymmetric multiplet is not observed in nature. If the supersymmetry is broken, we expect only a partial cancellation between bosons and fermions, even though there are equal numbers of bosonic and fermionic degrees of freedom. The first-order fermion and boson contributions in Eq. (11.53) would lead to a result that modifies the boson Eq. (11.55) as

$$\left(\rho_\Lambda c^2\right)_{susy} \cong \frac{1}{16\pi^2} \frac{E_{Pl}^4}{(\hbar c)^3} \left(\frac{\Delta m^2 c^4}{E_{Pl}^2}\right), \tag{11.57}$$

where Δm^2 is the fermion and boson mass difference $m_f^2 - m_b^2$. The fact we have so far not observed any superparticles means that such particles must at least be heavier by $\Delta m^2 \gtrsim (10^2 \text{ GeV}/c)^2$, which can only produce a suppression factor $(\Delta m^2 c^4/E_{Pl}^2) = 10^{-36}$ at the most—thus still some 80 to 90 orders short of the required $O(10^{-120})$. Clearly, something fundamental is missing in our understanding of the physics behind the dark energy.

[35] The vacuum energy is, of course, the sum totaling up the contributions from all quantum fields (gravitons, gauge bosons, leptons, and quarks, etc.).

[36] For example, we do not see a spin-zero particle, a " selectron," having the same properties, and degenerate in mass, as the electron; similarly we have not detected the photon's superpartner, a massless spin-$\frac{1}{2}$ particle called the "photino," etc. A plausible interpretation is that supersymmetry is broken, and the superpartners (selectrons, photinos, etc.) of the known particles (electron, photons, etc.) are much more massive and are yet to be produced and detected in our high energy laboratories.

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 274 of 276 September 24, 2009 17:51

Review questions

1. Use the first law of thermodynamics to show that the constancy of a system's energy density (even as its volume changes) requires this density to be equal to the negative of its pressure.
2. A vacuum energy dominated system obeys Newton's equation $\nabla^2\Phi = -\Lambda c^2$, where Λ is a positive constant. What is the gravitational potential $\Phi(r)$ satisfying this equation? From this find the corresponding gravitational field $\mathbf{g}(r) \equiv -\nabla\Phi(r)$.
3. From the Friedmann equation $1 - \Omega(t) = -kc^2/\dot{a}(t)R_0{}^2$ and the fact that the universe has been matter-dominated since the radiation–matter equality time with redshift $z_{\mathrm{RM}} = O(10^4)$, show that the deviation of energy density ratio Ω from unity at t_{RM} must be a factor of 10 000 times smaller than that at the present epoch t_0:
$$[1 - \Omega(t_{\mathrm{RM}})] = \left[1 - \Omega(t_0)\right] \times 10^{-4}.$$
Use this result (and its generalization) to explain the flatness problem.
4. What is the horizon problem? Use the result that the angular separation corresponding to one horizon length at the photon decoupling time is about one degree (for a flat universe) to explain this problem.
5. Use a potential energy function diagram to explain the idea of a phase transition in which the system is temporarily in a "false vacuum." How can such a mechanism be used to give rise to an effective cosmological constant?
6. Give a simple physical justification of the rate equation obeyed by the scale factor $\dot{a}(t) \propto a(t)$ in a vacuum energy dominated universe. Explain how the solution $a(t)$ of such a rate equation can explain the flatness and horizon problems.
7. How does the inflationary cosmology explain the origin of mass and energy in the universe as well as the origin of the cosmic structure we see today?
8. The CMB power spectrum can be divided into three regions. What physics corresponds to each region?
9. How can the observed temperature anisotropy of the CMB be used to deduce that the average geometry of the universe is flat?
10. The age of a flat universe without the cosmological constant is estimated to be $\frac{2}{3}t_{\mathrm{H}} \approx 9$ Gyr. Why can an accelerating universe increase this value?
11. What is dark energy? How is it different from dark matter? How is Einstein's cosmological constant related to such energy/matter contents? Do cosmic neutrinos contribute to dark energy?
12. Give two reasons to explain why type Ia Supernovae are ideal "standard candles" for large cosmic scale measurements.
13. Why should the accelerating universe lead us to observe the galaxies, at a given redshift, to be dimmer than expected (in an empty or decelerating universe)?
14. Why is the observation of supernovae with the highest redshifts (> 0.7) to be in the decelerating phase taken to be convincing evidence that the accelerating universe interpretation of SNe data ($z \;\; 0.2 - 0.7$) is correct?
15. What is the cosmic coincidence problem?
16. What is the standard Λ CDM cosmology? What is the spacetime geometry in this cosmological model? How old is the universe? What is the energy/matter content of the universe?

Problems

11.1 **Another form of the expansion equation** Use either the Friedmann equation or its quasi-Newtonian analog to derive (11.25).

11.2 **The epoch-dependent Hubble's constant and $a(t)$** Use (10.7) to replace the curvature parameter k in the Friedmann equation (10.1) to show the epoch dependence of Hubble constant through its relation to the density parameters as in (11.38).

11.3 **Luminosity distance and redshift in a flat universe** Knowing the redshift-dependence of the Hubble's constant from Problem 11.2 in a flat universe with negligible $\Omega_{R,0}$, show that the Hubble curve $d_L(z)$ can be used to extract the density parameters Ω_M and Ω_Λ from the simple relation

$$d_L(z) = c(1+z)\int_0^z \frac{c\,dz'}{H_0\left[\Omega_{M,0}\left(1+z'\right)^3+\Omega_\Lambda\right]^{1/2}}.$$

11.4 **Negative Λ and the "big crunch"** Our universe is spatially flat with the dominant component being matter and positive dark energy. Its fate is an unending exponential expansion. Now consider the same flat universe but with a negative dark energy $\Omega_\Lambda = 1 - \Omega_{M,0} < 0$, which provides a gravitational attraction, cf. (11.7). Show that this will slow the expansion down to a standstill when the scale factor reaches $a_{max} = (-\Omega_\Lambda/\Omega_{M,0})^{1/3}$. The subsequent contraction will reach the big crunch $a(t_*) = 0$ at the cosmic time $t_* = \frac{2}{3}\pi t_H(-\Omega_\Lambda)^{-1/2}$.

11.5 **Estimate of matter and dark energy equality time** Closely related to the deceleration/acceleration transition ("inflection") time is the epoch when the matter and dark energy components are equal. Show that the redshift result $z_{M\Lambda}$ obtained in this way is comparable to that of (11.48). Estimate the cosmic time $t_{M\Lambda}$ when the matter-dominated universe changed into our present dark energy-dominated universe.

978–0–19–957364–6 11-Cheng-c11 Cheng (Typeset by SPi, Chennai) 276 of 276 September 24, 2009 17:51