Chapter 11

A Free Lunch Called "The Universe"

Key features of the Big Bang are derived in an elementary manner: the critical density, the expansion rate, the time-temperature relation and the cosmic microwave background temperature. The horizon problem and the flatness problem are described.

Last update: 12/2/12

1. Introduction

I have a confession to make. For a very long time I refused to take cosmology seriously. My misguided position was that extrapolating so far from known and validated physics was preposterous. I changed my mind only when a period of convalescence gave me time to read Weinberg's *The First Three Minutes*, which had been sitting on my bookshelf for years. It is just not true that one must extrapolate so very wildly from secure theory. It rather depends upon when you wish to start the clock running. What I had not appreciated was that certain key features of the Big Bang follow from the most elementary, and hence compelling, of arguments. The aim of this chapter is to reprise these elementary arguments, eschewing both rigour and relativity in favour of simplicity. Cosmologists may squirm as much as they like, I am unrepentant. Chapter 23 will make up for these shortcomings with the relativistic story and also a discussion of dark energy which will be ignored here.

Inevitably the story starts in the middle. It is not yet possible to start the story of the universe at the beginning, though it may be one day. Many of the world's most able physicists toil to extend physical theory to apply even at the moment of creation. To disguise their hubris, this endeavour is given the code name "quantum gravity". Our purpose is far more modest. We shall pick up the story, say, one-hundredth of a second after the big bang. The universe was then simpler than it had been before and simpler than it will be again.

What exactly *was* the Big Bang? It will suffice for now to say that the Big Bang was the origin of the universe in an extremely hot, dense state and in which the constituents of the universe were flying away from each other at very high speeds. The idea seems arbitrary at first, but we shall see that it has an inevitability about it. Of course the Big Bang hypothesis is supported by a substantial body of observational evidence, not least the observed Hubble flow, the light-element constitution of the universe, and the cosmic microwave background. But for me the road to Damascus was provided by the elementary theoretical arguments presented below.

2. The Critical Density

Envisage a region, R, large compared with galactic superclusters, sufficiently large that the mean density of matter enclosed (ρ) is representative of the universe as a whole. In addition to the rest-mass density (of both ordinary matter and any 'dark matter'), this ρ also includes the mass equivalent of the 'positive energies'. The 'positive energies', are the radiation field energies (photons, neutrinos) plus the kinetic energy of the massive particles. What is excluded from the positive energies is the (negative) gravitational potential energy. This is negative because gravity is attractive, so the gravitational energy associated with two bodies in proximity must be less than when they are far apart. But the latter is zero, so the former is negative. Hence, we shall take the galactic masses plus the positive energies as constituting our total mean density, ρ. All this acts as a source of gravity. Thus, within a sphere of radius R, the gravitating mass is $M = (4\pi/3)R^3\rho$. The non-relativistic gravitational potential energy associated with a galaxy of mass *m* situated at the edge of the sphere of radius *R* is therefore,

$$P.E. = -\frac{GMm}{R} = -G\frac{4\pi}{3}mR^2\rho \qquad (1)$$

If the region is expanding then *R* is a function of time and the expansion corresponds to a "velocity" $v = \dot{R}$ (the reason for the inverted commas will emerge shortly). The Hubble parameter is defined as,

$$H(t) = v(t)/R(t) \qquad (2)$$

where the time dependence of all quantities is displayed explicitly. As defined here the parameter $H(t)$ is specific to the chosen region of size *R* at the specified epoch, *t*. However the homogeneity and isotropy of the universe on a large scale means that the same $H(t)$ will apply to any sufficiently large region at the specified time assuming the homogeneity is to persist.

Using (1) and (2), the total non-relativistic energy of our chosen galaxy is thus,

$$E = K.E. + P.E. = \frac{mv^2}{2} + P.E. = mR^2\left[\frac{1}{2}H^2 - \frac{4\pi}{3}G\rho\right] \qquad (3)$$

Note that although we have called this the total energy, it is only the sum of the non-relativistic kinetic energy and potential energy. There is no rest-mass energy included in (3). Hence it does not relate to the universe's density if divided by c^2. Nevertheless we shall see in Chapter 23 that a relativistic treatment leads to essentially this same expression (in the absence of dark energy). In the relativistic treatment, this expression is also a constant of the motion but, rather than being interpreted as energy, it is the curvature of space.

Equation (3) assumes that the Hubble velocity, $v = HR$, is the only significant motion of the typical galaxy. In addition to the tendency to move with the 'Hubble flow', stars and galaxies will have their own local (peculiar) motions superimposed. These additional local movements have been assumed to be slow compared with the Hubble speed. Dominance of the Hubble flow over the peculiar motion is inevitable so long as a large enough distance, *R*, is considered simply by virtue of $v = HR$.

[It is perhaps worth pausing to justify this point. Consider these local speeds: the earth's speed in its orbit around the sun is 30 km/s; the sun's speed in its orbit around the milky way is about 250 km/s; the milky way's speed with respect to the centre of the local group of galaxies is about 300 km/s. Suppose these conspire to give a local speed of nearly 600 km/s. Since the Hubble parameter is 71 km/s/Mps = 22 km/s/Mlyr, it follows that the Hubble 'flow' speed will be comparable, i.e., 600 km/s, at a distance of 600 / 22 = 27 Mlyr. The typical diameter of a spiral galaxy is 100 to 150 klyr. The typical size of a cluster of galaxies is ~5 Mlyr, whilst clusters of clusters of galaxies are ~20 Mlyr in size. Finally, the thread-like filaments which comprise the largest non-uniform structure in the universe, the superclusters, are several hundreds of Mlyr in length. Hence, the size scale on which the universe can be regarded as uniform is of the order of hundreds of Mlyr. Thus, if *R* is sufficiently large for the universe to be approximately homogeneous, it will be correct to assume the Hubble motion is dominant. It is no coincidence that the size scale at which the Hubble flow becomes dominant coincides with that at which the universe becomes homogeneous: below this scale the development of structure must have required dominant peculiar motions.] *Check the above numbers*

The energy given by (3) is constant as the universe expands. Note that the net energy, *E*, can be negative if the gravitational potential energy exceeds the kinetic energy in magnitude. If *E* is negative then the *R*-sized region must have a maximum possible size. This is because, if we allowed *R* to tend to infinity then the density would decrease to zero, leaving only the first (kinetic energy) term in (3), which cannot be negative – a contradiction. In fact, from (2), it must be that $H = 0$ at the maximum size, since the velocity must then be zero as the galaxies reverse direction and the universe begins to contract. Hence, for negative *E*, the maximum of *R* follows immediately from (3), i.e.,

For $E < 0$
$$R_{MAX} \sim \frac{GM^2}{|E|} \tag{4}$$

where *M* is the total mass of the *R*-region. The approximate nature of (4) is due to it being strictly invalid to replace our small test mass, *m*, with the mass of the whole *R*-region, *M*. An integration to get the total gravitational potential energy is really required, and this would introduce a factor of 3/5 into (4), see Chapter 57.

Although we have concluded that a maximum size is reached for a specific region, it is valid for any larger region if the universe is homogeneous: the universe as a whole will stop expanding and re-contract at some finite time if $E < 0$.

As the kinetic and gravitational energies become closer to cancellation, i.e., as $E \to 0$, the *R*-region becomes destined to reach an ever greater maximum size. As the cancellation becomes exact, and the net energy of the *R*-region becomes zero, its radius can expand to infinity (with velocities asymptotic to zero). From (3) the critical density at which this condition occurs is,

$$\rho_{critical}(t) = \frac{3}{8\pi G} H(t)^2 \tag{5}$$

For unchanging velocities the age of the universe would be given simply by $t = R/v = 1/H$, so (5) gives,

$$\rho_{critical}(t) = \frac{3}{8\pi G t^2} \tag{6}$$

Just as for the R-region, so for the universe as a whole: if $\rho \le \rho_{critical}$ it will expand forever, if $\rho > \rho_{critical}$ it will stop expanding at some finite time, having reached a minimum density, and re-contract.

Note that the critical density varies with the age, *t*, of the universe, as does the Hubble parameter. Note also that if gravity were weaker (i.e., if *G* were smaller) then the critical density would be larger. This is not surprising, since, if gravity were weaker, it would obviously require a greater density to reverse the universe's expansion.

The value of *H* at the present time is 71 km/sec/Mparsec = 2.3 x 10^{-18} /sec., which is equivalent to the age of the universe being $1/H =$ 13.7 Gyrs. This value for *H* derives from the WMAP/COBE satellite observations of the microwave background and is believed accurate to +/-4%. Since G = 6.67 x 10^{-11} $m^3kg^{-1}sec^{-1}$ we get a critical density of

$$\rho_{critical} = 9.6 \times 10^{-27}\ \text{kg/m}^3 \qquad (7)$$

The mass of one hydrogen atom is 1.67 x 10^{-27} kg. Hence the critical density corresponds to just one hydrogen atom per volume 1.67 x 10^{-27}kg/ 9.6 x $10^{-27}$$m^3$ = 0.17 m^3, or just 5.7hydrogen atoms per cubic metre.

In sharp contrast, the best achievable vacuum in terrestrial physics laboratories is around 10^{11} molecules per cubic metre. Hence, an average cosmic density of material which would constitute a record breaking vacuum by some 10 orders of magnitude on earth would be sufficient in its gravitational effects to eventually halt the expansion of the universe and cause it to contract.

Let us summarises what we have concluded so far,

- If the average cosmic density exceeds a certain time-dependent critical density, then $E < 0$ and the universe will stop expanding and contract instead. The situation is analogous to a projectile reaching the zenith of its trajectory. (This conclusion is not valid if dark energy exists – see Chapter 23).
- Conversely, a universe with an average density less than the critical density has $E > 0$ and is analogous to a projectile whose velocity exceeds the escape velocity: it will expand indefinitely.
- A very simple expression for the critical density has been derived, (6), in terms of the universal gravitational constant, *G*, and the age of the universe alone. At the present epoch (13.7 billion years), the critical density is found to be remarkably small, namely ~5.8 hydrogen atoms per cubic metre.

These remarkable conclusions follow from the simple energy equation (3). It almost seems to be cheating to conclude so much from so little. But we have not yet exhausted the implications of equation (3).

3. The Expanding Universe

Suppose we are dealing with a fixed inventory of mass and no radiant energy. Then the density will be proportional to $1/R^3$. On the other hand, pure radiation would cause the density to vary as $1/R^4$. This is because, as the universe expands, the wavelength of each frequency component of the radiation is effectively stretched in proportion to the expansion of universe. So if the universe expands by a factor X, then the wavelength of each radiation mode also increases by the factor X. This means that the energy associated with each mode decreases by the factor X (because the quantum of energy is inversely proportional to its wavelength). Hence, the total amount of radiant energy decreases according to $1/R$. Dividing through by the volume gives a radiant energy density proportional to $1/R^4$. Hence, at sufficiently early times, $\rho(t)$ will increase at least as fast as $1/R^3$ as we move backwards towards $t = 0$.

Now, since the energy *E* given by (3) is constant it follows that the factor in the bracket must involve a cancellation of the leading divergent term, i.e., that in $1/R(t)^3$ or $1/R(t)^4$, so as to leave this term proportional to $1/R(t)^2$, as required to make (3) constant. It follows that at sufficiently early times, when $R(t) \to 0$, the two terms in the bracket must virtually cancel. This gives, For sufficiently small *t*: $$H(t)^2 \approx \frac{1}{t^2} \approx \frac{8\pi G}{3}\rho(t) \tag{8}$$

But (8) is the same as the expressions (5,6) for the critical density. So we conclude that for sufficiently early times the density of the universe is inevitably very close to the critical density,

For $t \to 0$: $$\frac{\rho}{\rho_{critical}} \to 1 \tag{9}$$

We have already concluded that the density is related to the size scale by,

$$\rho(t) \propto 1/R(t)^n \tag{10}$$

where *n* is 3 or 4 for matter and radiation respectively. Since (8) gives *H* in terms of the density, and (10) gives the density in terms of the radius, we conclude,

$$H(t) \propto 1/R(t)^{n/2} \tag{11}$$

But using (2) with $v = dR/dt$ gives,

$$\frac{dR}{dt} = R(t)^{1-\frac{n}{2}} \tag{12}$$

Integrating (12), and taking the size parameter to be $R = R_0$ at some arbitrary datum time zero, gives

$$R(t)^{n/2} - R_0^{n/2} = At \tag{13}$$

where A is some constant. Consistent with (13) we could claim by *fiat* that the universe started at the finite size scale R_0 at the moment of creation, here given by $t = 0$. However, the derivation of (13) from (12) is just as valid for negative times, for which it gives $R < R_0$. There is nothing physical to prevent this further extrapolation backwards, so claiming that the universe sprang into being with a finite density at $t = 0$ would be no better motivated than claiming that it began at any other arbitrary time. In the absence of any such reason it is natural to assume that that (13) applies all the way back to the time $t = -R_0^{n/2}/A$ when the density becomes singular because $R \to 0$. Having come to this conclusion we can shift our datum zero time to this singularity, i.e., the Big Bang, and (13) becomes simply,

$$R(t) \propto t^{2/n} \tag{14}$$

This has been derived for sufficiently early times only, because (8) has been justified (so far) only for early times. But at sufficiently early times the density of the universe will be dominated by radiation rather than matter, because the former varies as $1/R^4$ and the latter only as $1/R^3$, so the former will dominate as $R \to 0$. Consequently we are justified only in writing (14) as,

Radiation dominance: $$R(t) \propto \sqrt{t} \tag{15}$$

However (14) would hold for arbitrary times if (8) were valid for arbitrary times, in other words if the energy *E*, (3), were zero. There is a temptation to regard the total energy of the universe as zero since it resolves the paradox of how the universe can be created *ex nihilo*. If the net sum of the universe's mass-energy is zero then we no longer have to worry about how something could spring from nothing. Instead it is only nothing that springs from nothing. The universe is therefore the ultimate free lunch, as Alan Guth (1998) has said. However, the concept of the universe's total mass-energy is fraught with difficulty, as explained in Chapter 57. An uncontentious route to concluding that $R(t) \propto t^{2/3}$ in the matter dominated era is provided in Chapter 23 via a relativistic treatment where it is seen to apply when the spacetime curvature, rather than the energy, is zero.

Let us summarise what we have learnt about the cosmic expansion from these very simple considerations,

(i) The size scale, $R(t)$, shrinks to zero at $t = 0$, so that any *finite* region of the universe would have been a point at the Big Bang.

(ii)The rate of the universe's expansion is divergent at $t = 0$, i.e., $dR/dt \propto 1/\sqrt{t}$. In other words, the universe expands divergently rapidly immediately after the Big Bang.

In relation to (i), note that if the universe is infinite then it would have been infinite at $t = 0$ and the Big Bang would have taken place at all points simultaneously.

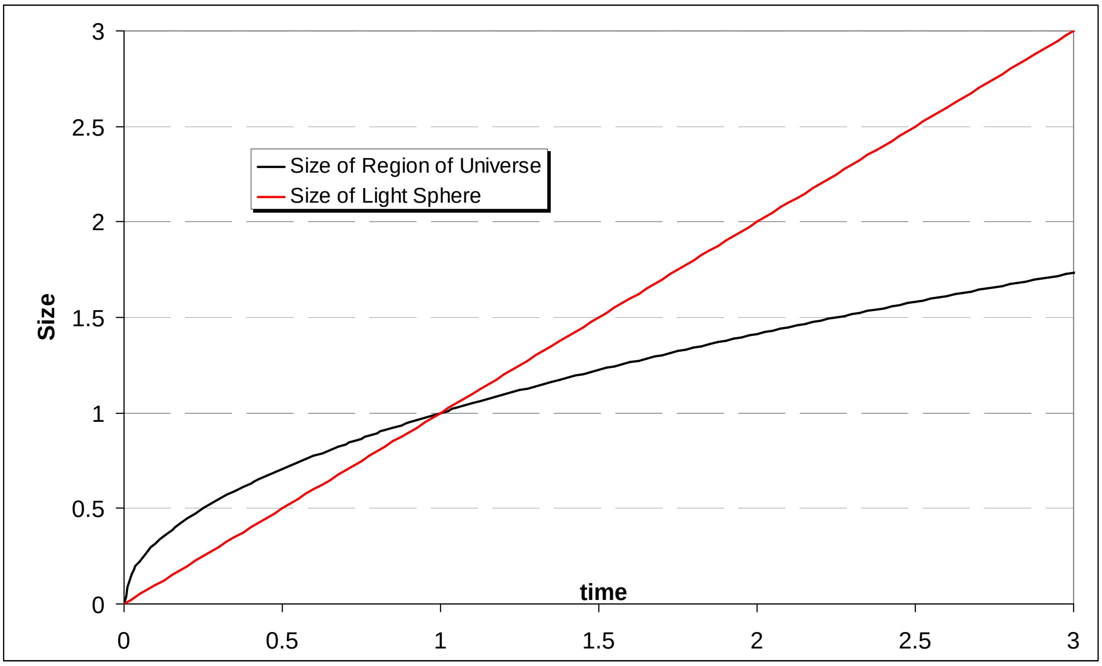
4. The Horizon Problem

The reader will have spotted something odd about (12), or equivalently (15): they imply that as $t \to 0$ or $R \to 0$ the "velocity" $v = dR/dt$ is divergent. But does this not violate the cosmic speed limit, the speed of light? You might reasonably suppose this to be because our treatment has been non-relativistic and that a relativistic analysis would respect $dR/dt \le c$. But this is not so. The general relativistic development in Chapter 23 also results in $dR/dt \to \infty$ near the Big Bang. How can this be?

In effect, the universe expands too quickly for light to keep up in the early stages. The resolution of the apparent conflict with the (special) relativistic requirement that nothing exceeds the speed of light is that dR/dt is not the velocity of any material substance but the expansion of space itself. From this point of view, two so-called 'comoving' particles, that is particles which are simply 'going with the Hubble flow' are as near to stationary as you can get in this universe. Nevertheless the distance between them, R, is increasing at a rate $v = dR/dr = HR$. It is ironical that this general relativistic conundrum would not arise if we had not had special relativity drummed into us so effectively, i.e., if we did not believe in a cosmic speed limit. In some ways, the general relativistic cosmological solutions are conceptually closer to non-relativistic physics than special relativistic physics as a result of the existence of a preferred cosmic time coordinate.

So it seems that we must accept that the universe does expand too rapidly for light to keep up in the early stages. The size scale is increasing as given by (15), $R(t) \propto \sqrt{t}$ and $dR/dt \propto 1/\sqrt{t}$ whereas a pulse of light forms an expanding sphere of radius $r = ct$. The former is inevitably a faster expansion at early times, whilst the latter is inevitably faster at later times. This is illustrated in Figure 1. The normalisation of the time and space axes is arbitrary. The physical region of size R, representing a portion of the universe, can be regarded as defined by a specified inventory of material particles. This is usually called the Hubble volume because we are assuming the matter within it is free to move with the Hubble flow. Suppose we choose the Hubble volume to coincide with the light-sphere at unit time in Figure 1. For all later times the size of the Hubble volume is less than that of the light sphere. However if we look at earlier times the light sphere was *smaller* than the Hubble volume, R. In fact the ratio of the radius of the light sphere to that of the Hubble volume varies as $r/R \propto \sqrt{t}$ and hence shrinks to zero as $t \to 0$.

Figure 1 Illustration of the Horizon Problem



But a causal connection can exist only between regions within the light sphere. So this implies that, at early times, only a vanishingly small proportion of our Hubble volume was in causal contact. The whole Hubble volume would appear to break up into a myriad of regions which are mutually causally unconnected at early times. But in this case, how can the universe have become homogeneous and isotropic? It would seem to demand a remarkable coincidence between independent regions. The point is that homogeneity usually originates from a causally connected region which has permitted, at some time, a thermodynamic equilibrium to be established.

A simpler way to understand this causality issue is to consider a pair of diametrically opposite points on the cosmic horizon. The cosmic horizon is the furthest point which can currently be seen (in principle) due to the limitations imposed by the finite speed of light and the finite age of the universe. A point on the horizon is, by definition, only just able to influence our cosmic locality by the present epoch. So how could diametrically opposite points have any effect upon each other? How can the equilibrium state which is suggested by the homogeneity and isotropy of the universe have arisen? This is the Horizon Problem.

It is tempting to suppose that the equilibrium was established very early in the life of the universe, only a very short time after the Big Bang. This is indeed thought to be the case. But the difficulty is that it is not consistent with the picture illustrated by Figure 1. As a result of (15) there would be vanishingly little causal connection at early times. The currently dominant hypothesis to resolve this problem is inflation theory which is discussed in simple terms in Chapter 56.

5. The Flatness Problem

In §4 we noted that it is inevitable that the ratio $\Omega = \rho / \rho_{critical}$ tends to unity at early times. However this does not mean that it must be close to unity at later times. Virtually any value of Ω could arise at later times and still have the limiting value $\Omega \to 1$ as $t \to 0$ (as we shall shortly see). From the perspective of the simple theory employed so far, the actual value of Ω at the present epoch appears to be a contingent property of our universe which must be found from observational evidence. Specific cosmological theories such as inflation will constrain, or imply, the value of Ω, but we shall ignore this for now in order to explain the Flatness Problem.

Using the expression (5) for the critical density, the energy equation, (3), can be re-arranged to give,

Matter Dominated Era

$$\Omega = \left|\frac{P.E.}{K.E.}\right| = 1 - \frac{E}{E + \frac{GMm}{R}} \tag{16}$$

Radiation Dominated Era

$$\Omega = \left|\frac{P.E.}{K.E.}\right| = 1 - \frac{E}{E + \frac{Gm\lambda_R}{R^2}} \tag{17}$$

In the matter dominated case the total mass within region R is the constant M. In the radiation dominated case this mass-energy diminishes as the region expands and we have put $M = \lambda_R / R$. In both cases the mass m is merely the test mass whose energy, E, we are considering. Consequently if the size scale is R_0 at some arbitrary time t_0 we can define the constants,

$$\xi = \frac{ER_0}{GMm} \qquad \text{and} \qquad \tilde{\xi} = \frac{ER_0^2}{GM\lambda_R} \tag{18}$$

Hence (16) and (17) become, using (14),

Matter Dominated Era

$$\Omega = 1 - \frac{\xi}{\xi + (R_0 / R)} = 1 - \frac{\xi}{\xi + (t_0 / t)^{2/3}} \tag{19}$$

Radiation Dominated Era

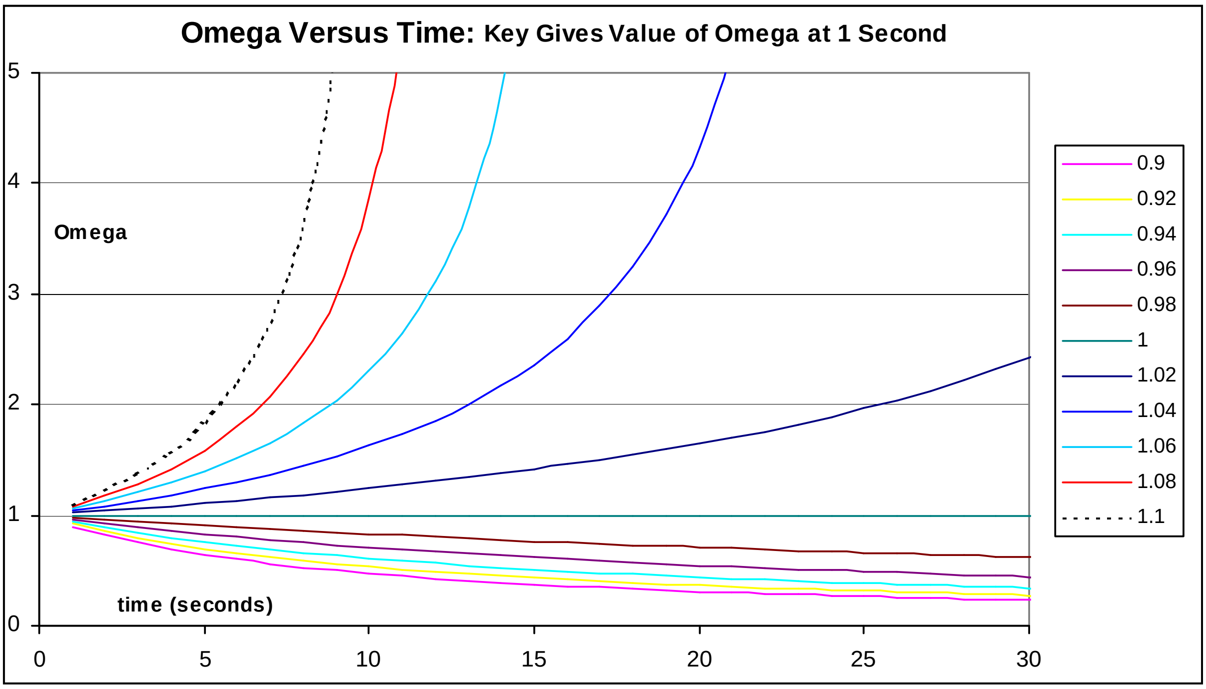
$$\Omega = 1 - \frac{\tilde{\xi}}{\tilde{\xi} + (R_0 / R)^2} = 1 - \frac{\tilde{\xi}}{\tilde{\xi} + (t_0 / t)} \tag{20}$$

Recall that ξ and $\tilde{\xi}$ are constants, so that equations (19) and (20) give the time variation of Ω explicitly. Some qualitative features follow from (19,20) immediately. Consistent with our previous observation, if $E > 0$ so that $\xi, \tilde{\xi} > 0$, then Ω must be less than unity, i.e., a density less than the critical density corresponds to positive energy. However, (19,20) also show that in this case Ω decreases monotonically and is asymptotic to zero.

If $E < 0$ so that $\xi, \tilde{\xi} < 0$, and assuming $t < t_0 / |\xi|^{3/2}$ (Equ.19) or $t < t_0 / |\tilde{\xi}|$ (Equ.20), then Ω must be greater than unity, i.e., a density greater than the critical density corresponds to negative energy. However, (19,20) also show that in this case Ω increases monotonically and becomes divergent at $t = t_0 / |\xi|^{3/2}$ (Equ.19) or $t = t_0 / |\tilde{\xi}|$ (Equ.20). This is the time at which the universe reaches its maximum size and the critical density is instantaneously zero.

Hence, despite the density necessarily being very close to the critical density at early times, it will diverge away from the critical density at later times unless the net energy, E, is very close to zero. For a tiny positive energy the density tends to zero as $t \to \infty$ and $R \to \infty$. For a tiny negative energy the density reaches a non-zero minimum after a finite time when R reaches a finite maximum. This behaviour is illustrated in Figure 2. This graph assumes some arbitrary value $\Omega = \Omega_1$ at 1 second. Equs.(19,20) determined the value of Ω thereafter. A range of starting values, Ω_1, from 0.9 to 1.1 are plotted in Figure 2. After only a few seconds Ω has diverged away from its initial value unless it starts extremely close to 1.

Figure 2 Illustration of the Flatness Problem



In the general relativistic interpretation, $\Omega = 1$ corresponds to flatness of the spacetime. Observational evidence, from the WMAP satellite, Jarosik *et al* (2011), and other sources, indicates that Ω equals unity to within ~1% at the present epoch. In view of the above results this implies that the value of Ω at (say) 1 second would have to be equal to unity with extraordinarily high precision. Just *how* fine-tuned to unity would Ω_1 have to be? We can work it out using (19,20).

Take the current age of the universe to be 13.7 billion years (t_{now} = 4.3 x 10^{17} s). We shall also need the time at which the radiation era gives way to the matter dominated era, since the time dependence of Ω is different in the two eras. The radiation era will be taken to end at 75,000 years (t_{eq} = 2.4 x 10^{12} s) for the sake of this illustration.

Suppose $\Omega_{now} = 0.99$ and consider firstly the period t_{eq} to t_{now}. This period is matter dominated and (19) gives, setting $t_0 = t_{now}$,

$$\Omega_{now} = 0.99 = 1 - \frac{\xi}{\xi + (1)^{2/3}} = \frac{1}{1+\xi} \quad \text{hence} \quad \xi = 0.0101$$

Hence,
$$\Omega_{eq} = 1 - \frac{0.0101}{0.0101 + \left(4.3\times 10^{17} / 2.4\times 10^{12}\right)^{2/3}} = 1 - 3.18\times 10^{-6}$$

Now consider the period from 1 second to t_{eq}. This period is radiation dominated and (20) gives, setting $t_0 = t_{eq}$,

$$\Omega_{eq} = 1 - 3.18\times 10^{-6} = 1 - \frac{\tilde{\xi}}{\tilde{\xi} + (1)} = \frac{1}{1+\tilde{\xi}} \quad \text{hence} \quad \tilde{\xi} = 3.18\times 10^{-6}$$

Hence,
$$\Omega(1\,\text{sec}) = 1 - \frac{3.18\times 10^{-6}}{3.18\times 10^{-6} + \left(2.4\times 10^{12} / 1\right)} = 1 - 1.3\times 10^{-18}$$

Thus, to produce a value Ω_{now} = 0.99 we require a value of Ω at 1 second which differs from unity by only ~10^{-18}. The same is found for Ω_{now} = 1.01. And, of course, the use of 1 second as the starting time was arbitrary. Using a small fraction of a second as the starting time increases the required precision still further.

It is one of the triumphs of inflation theory that it provides a mechanistic explanation for the extreme degree of fine tuning required to produce $\Omega \approx 1$ in the present epoch (see Chapter 56).

In the present simplistic development, Equs.(16,17) link $\Omega = 1$ with an exact balance of the kinetic and gravitational potential energies so that the net (non-relativistic) energy, E, is zero. This tempts one to imagine that the universe began as a quantum fluctuation. Ordinarily a quantum fluctuation does not last very long. Specifically, a quantum fluctuation of energy E would be expected to last for a time of about $\hbar / E$. But if the net energy is fine-tuned to be close to zero, a quantum fluctuation could perhaps persist. From this perspective, the flatness of the universe becomes a prerequisite for its longevity. Appealing though this may be, there are theoretical difficulties with this idea, not least that the total energy of the universe is dreadfully ill defined (see Chapter 57).

However there is another issue which even an approximate equality of the potential and kinetic energies assists in explaining. Why were the contents of the universe flying away from each other at the start – why did the Big Bang have the characteristic of an explosion? Why is the universe expanding rather than static? A static universe suggests being gravitationally bound. But by the virial theorem this would require the kinetic energy to be half the magnitude of the potential energy (and the opposite sign). Zero net energy therefore indicates an excess of kinetic energy – with the result being universal expansion. In this sense a Big Bang scenario seems inevitable if we link zero net energy to the possibility of *ex nihilo* creation (or flatness to inflation). However these thoughts are merely heuristic and should not be regarded as secure.

6. The Time-Temperature Relationship in the Early Universe

During the first ~14 seconds or so there are still large numbers of electrons and positrons in the Big Bang fireball due to the very high temperatures. However, after ~14 seconds, and whilst the universe is radiation dominated, the universe's density is due to photons and neutrinos. The density of the photons is that of black body radiation given in terms of Stefan's constant, $\sigma = 5.67 \times 10^{-8}$ $Wm^{-2}K^{-4}$, by ,

$$\rho^{photons} = \frac{4\sigma T^4}{c^3} \tag{21}$$

where T is the absolute temperature. In the Appendix to this Chapter the contribution of the neutrinos to the density are calculated and shown to equal 68% of the photon contribution. Hence the total density in the radiation era is,

$$\rho^{radiation} = \frac{6.72\sigma T^4}{c^3} \tag{22}$$

But we know how density relates to the Hubble parameter from (5), giving,

$$T^4 = 0.01775\frac{c^3 H^2}{\sigma G} \tag{23}$$

The Hubble parameter is related to the age of the universe to a reasonable approximation by $H \approx 1/t$ at later times. However this is not accurate in the radiation dominated period. Here we have, from (15) that $R = \tilde{A}\sqrt{t}$ for some constant $\tilde{A}$. This gives,

Radiation era:
$$H \approx \frac{1}{2t} \tag{24}$$

Hence (23) becomes,

$$T^4 = 0.00444\frac{c^3}{\sigma G t^2} \tag{25}$$

Substituting numerical values for the universal constants gives,

Radiation era:
$$T = \frac{1.33 \times 10^{10}}{\sqrt{t}} \tag{26}$$

where temperature is in K and time in seconds, and (26) is valid for $14s < t < t_{eq}$. So, at ~100 seconds the temperature of the Big Bang fireball is about 1.3 billion K. The Appendix shows that the annihilation of the electron-positron pairs at ~14 seconds causes the temperature of the fireball to increase by a factor of 1.4. So (26) also implies that the temperature in the first 14 seconds is $T \approx \frac{0.95 \times 10^{10}}{\sqrt{t}}$. The temperature is thus ~10 billion K at 1 second.

It still amazes me that this can be deduced with nothing in the way of observational data required in the calculation. Equ.(25) involves only universal constants in the time-temperature relation.

7. The Cosmic Microwave Background Temperature

Can we push our luck and also derive an expression for the temperature of the universe in the matter dominated era? Well, no, not really, because once matter and radiation have decoupled there is no longer a unique temperature applicable to the whole universe. It is no longer in thermal equilibrium. The matter content is destined to undergo gravitational collapse and to give rise to a wide range of temperatures in the form of stars and gas clouds, etc. However, the radiation from the original Big Bang fireball has now decoupled from the matter and retains its own independent temperature. Over the eons its temperature has reduced as the universe has expanded. This is what we now detect as the cosmic microwave background (CMB). Its temperature has been very accurately measured, 2.725K. Can we extend our temperature estimate to cover the CMB?

The temperature of the radiation in the matter dominated era will continue to reduce in proportion to $1/R$. But in the matter era $R \propto t^{2/3}$, so $T_{rad} \propto t^{-2/3}$ for $t > t_{eq}$ where t_{eq} is the time when the radiation era gives way to the matter era. We shall simply assume that this is 75,000 years. The temperature at the end of the radiation era is thus, from (26),

$$T_{eq} = \frac{1.33 \times 10^{10}}{\sqrt{t_{eq}}} \tag{27}$$

Using $T_{rad} \propto t^{-2/3}$ after this period we deduce that the temperature of the CMB now should be,

$$T_{CMB} = \frac{1.33 \times 10^{10}}{\sqrt{t_{eq}}} \left(\frac{t_{eq}}{t_{now}} \right)^{2/3} = 1.33 \times 10^{10} \, \frac{t_{eq}^{1/6}}{t_{now}^{2/3}} \tag{28}$$

Inserting the numerical values t_{eq} = 2.4 x 10^{12} s and t_{now} = 4.3 x 10^{17} s provides our estimate of the CMB temperature of 2.70K. This is remarkably close to the measured 2.725K given the simplicity of the estimate.

However, we have cheated. We have smuggled the time of radiation/matter equality, t_{eq}, into the calculation. To derive t_{eq} we must know the matter content of the universe, and this must include dark matter as well as ordinary (baryonic) matter. This can be deduced from observational evidence. However, in practice, the accuracy of the measurements of the CMB temperature are such that these are used to constrain estimates of the matter density of the universe (often expressed as the photon:baryon ratio). Despite the confession of having cheated, note that (28) is fairly insensitive to t_{eq} due to the power $1/6$, and t_{eq} can be estimated from the observed matter content of the universe without appeal to the CMB. For example, doubling the estimate of t_{eq} only changes the calculated CMB temperature to 3.0K. So it is rather impressive that the very simple theory presented above can predict the CMB temperature so well – after all this time!

With the benefit of twenty-twenty hindsight it is surprising that no one set out deliberately to detect the CMB, leaving it instead for Penzias and Wilson (1966) to stumble upon serendipitously. Given the simplicity of the calculations leading to the above estimate of the CMB temperature it was evident that it should be sufficiently intense to be detectable. The detection of the CMB at about the expected temperature marked the point at which cosmology, and the Big Bang hypothesis in particular, came of age as a science in its own right.

Appendix - Neutrino Numerology and Electron-Positron Annihilation

The neutrinos cease to interact significantly with the rest of the universe at about one second after the big bang. The neutrinos are no longer in thermal equilibrium with the rest of the universe after this time. This does not cause any immediate difference in temperature between the neutrinos and everything else, but it will do so only ~13 seconds later. By 14 seconds the electrons and positrons have mostly annihilated, producing large numbers of very energetic gamma rays (photons). These gamma rays quickly become ‘thermalised’ in the very dense conditions that prevail. In other words, the energy released by annihilation of the electron/positron pairs heats up the remaining contents of the universe – at least, those constituents which interact sufficiently to enable the energy transfer to take place. Thus, the temperature of the photons, and of the remnant electrons, protons and neutrons, are almost instantaneously increased. The neutrinos, however, do not partake of this free hand-out of energy. They cannot do so because they are no longer interacting with anything. Hence, from ~14 seconds until the present day, the neutrino background has been cooler than the microwave background. But by how much?

It is essentially book-keeping to evaluate the temperature increase of the photons due to electron-positron annihilation. The trick to the book-keeping exercise is to consider the entropy of the photon and e^- / e^+ radiations. The entropy of a thermalised (equilibrium) particle species with N_s spin states and N_a distinct antiparticles is, per unit volume,

$$S = fN_a N_s \frac{2\pi^2 k_B}{45}\left(\frac{k_B T}{\hbar c}\right)^3 \tag{29}$$

where $f = 1$ for bosons (e.g. photons) but $f = 7/8$ for fermions (e.g., electrons). Consequently, the leading factor fN_aN_s before the e^- / e^+ annihilation is,

$$fN_aN_s = \frac{7}{8}\times 2\times 2 + 1\times 1\times 2 = \frac{7}{2} + 2 = \frac{11}{2} \tag{30}$$

where the first term accounts for the electrons and positrons, and the second term for the photons. After the annihilation of the electrons and positrons, this factor is just 2, i.e., the photons alone. Hence, if the temperature remained constant, Equ.(29) would imply that the entropy had decreased by a factor of 4/11. But entropy cannot decrease. Thus, the temperature must increase sufficiently to counteract the decrease in the fN_aN_s factor, giving,

$$\text{photon temperature increases by a factor of } \left(\frac{11}{4}\right)^{1/3} = 1.401 \tag{31}$$

Hence, the neutrino temperature today will be the photon temperature divided by 1.401, i.e. 2.725^oK / 1.401 = 1.945^oK. Note that the neutrinos themselves have played no part in the derivation of this temperature. It has been based solely on the measured photon temperature, and on the numerology of the photons and electrons/positrons. Unfortunately the cosmic neutrino background is undetectable due to the extremely weak interaction cross-sections of neutrinos, especially at such low energies.

The energy (and hence mass) density of a blackbody field is proportional to the absolute temperature to the power 4. However, the factor of fN_aN_s in Equ.(29) also occurs in the energy density. In the standard theory of weak interactions, neutrinos occur in just one spin state, despite being spin ½. Only negative helicity neutrinos and positive helicity antineutrinos take part in weak processes. Whether this means that positive helicity neutrinos and negative helicity antineutrinos do not exist is not certain, but because they do not interact at all, except by gravity, we do not know of them. If they do exist then the neutrino density will be double what we assume here.

Hence, for a given type of neutrino (say, the electron neutrino), fN_aN_s = 7/8 x 2 x 1 = 7/4. But there are three types of neutrino, the electron neutrino, the mu neutrino, and the tau neutrino. Hence, fN_aN_s = 21/4 in total. Compared to the photon field (fN_aN_s = 2), this factor is therefore 21/8 larger. But, each of the neutrino fields is at a temperature which is only $(4/11)^{1/3}$ times that of the photons, from Equ.(31). Hence, the three neutrino fields contribute to the total energy density of the universe as follows,

$$\frac{\textit{energy density of all neutrinos}}{\textit{energy density of photons}} = \frac{21}{8}\left(\frac{4}{11}\right)^{4/3} = 0.681 \tag{32}$$

Consequently the total energy density of the photon and neutrino fields is 1.681 times that for the photons alone.

References

A.H.Guth (1998) "The Inflationary Universe", Vintage Press

N. Jarosik et al (2011) "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results", The Astrophysical Journal Supplement Series, **192**:14 (15pp), 2011 February.

A.A.Penzias and R. W. Wilson (1965). "A Measurement Of Excess Antenna Temperature At 4080 Mc/s". Astrophysical Journal Letters **142**: 419–421.