Inflation

Using speculative particle physics to avoid fine tuning
We now leave the well-established and understood topics in cosmology in order to discuss something more speculative. The idea in question is **cosmological inflation**, which was invented in 1981 and remains a hot research topic in modern cosmology. Inflation is not a replacement for the Hot Big Bang theory, but rather an extra add-on idea which is supposed to apply during some very early stage of the Universe’s expansion. By the time the Universe has reached the ages we have already discussed, inflation is supposed to be long since over and the standard Big Bang evolution restored, in order to preserve the considerable successes we have already discussed, such as the microwave background and nucleosynthesis.
Problems with the Big Bang Model

The Horizon Problem:
Why is the universe so homogeneous (CMB)? And why, if it’s so homogeneous, is it just inhomogeneous enough, and on just the right scale (CMB anisotropies), to form galaxies and clusters? This is a fine tuning problem.

The Flatness Problem:
We are very close to critical density today, and the curvature parameter decreases with time less rapidly than the matter/radiation density parameters. Why was the Universe created with density so astronomically close to critical density? This is another fine tuning problem.

The Monopole Problem:
Where are the heavy particles predicted by GUTs?
The Monopole (or heavy particle) Problem

Another mystery arises from combining the Hot Big Bang model with modern ideas of particle physics. One of the curious things about the Universe is that it remained radiation dominated for so long, until an age of at least 1000 years. That is unexpected because the radiation density reduces with expansion as $1/a^4$, much faster than any other type of matter. If the Universe starts with just a very small amount of non-relativistic matter, then its slower reduction in density will rapidly bring it to prominence.

In fact, the particles in the Standard Model of particle interactions don’t lead to any problems, because they interact strongly with radiation and thermalization stops them becoming too prominent. But modern particle physics throws up other particles. The most crucial in originally motivating inflation was a type of particle known as a magnetic monopole. Such particles are an inevitable consequence of models of unification of fundamental forces, the so-called Grand Unified Theories, and it is predicted that they were produced with a high abundance at a very early stage in the Universe. They are predicted to be extraordinarily massive; the Grand Unified Scale is thought to be around $10^{16}$ GeV, in comparison to the proton’s puny 1 GeV or so. Such particles would be non-relativistic for almost all the Universe’s history, giving them plenty of time to come to dominate over radiation. Since we know the Universe is not dominated by magnetic monopoles now, theories predicting them are incompatible with the standard Hot Big Bang model. This is further explored in Problem 13.5.

While magnetic monopoles were the relic particle thought most important at the time inflation was conceived, there are now several other kinds of relic particle also speculated to exist which would cause similar problems, going under such elaborate names as gravitinos and moduli fields.
The flatness problem is the easiest one to understand. We have learned that the Universe possesses a total density of material, $\Omega_{\text{tot}} = \Omega_0 + \Omega_\Lambda$, which is close to the critical density. Very conservatively, it is known to lie in the range $0.5 \leq \Omega_{\text{tot}} \leq 1.5$. In terms of geometry, that means that the Universe is quite close to possessing the flat (Euclidean) geometry.

We have seen that the Friedmann equation can be rewritten as an equation showing how $\Omega_{\text{tot}}$ varies with time. Adding modulus signs to equation (7.4), this is

$$|\Omega_{\text{tot}}(t) - 1| = \frac{|k|}{a^2 H^2}.$$  \hspace{1cm} (13.1)

We know from this that if $\Omega_{\text{tot}}$ is precisely equal to one, then it remains so for all time. But what if it is not?
Let's consider the situation where we have a conventional Universe (matter or radiation dominated) where the normal matter is more important than the curvature or cosmological constant term. Then we can use the solutions ignoring the curvature term, equations (5.15) and (5.19), to find

\[
a^2 H^2 \propto t^{-1} \quad \text{radiation domination; (13.2)}
\]
\[
a^2 H^2 \propto t^{-2/3} \quad \text{matter domination. (13.3)}
\]

So we have

\[
|\Omega_{\text{tot}} - 1| \propto t \quad \text{radiation domination; (13.4)}
\]
\[
|\Omega_{\text{tot}} - 1| \propto t^{2/3} \quad \text{matter domination. (13.5)}
\]

In either case, the difference between $\Omega_{\text{tot}}$ and 1 is an increasing function of time. That means that the flat geometry is an unstable situation for the Universe; if there is any deviation from it then the Universe will very quickly become more and more curved. Consequently, for the Universe to be so close to flat even at its large present age means that at very early times it must have been extremely close to the flat geometry.
An alternative way to see this is to remember that the densities of matter and radiation reduce with expansion as $1/a^3$ and $1/a^4$ respectively. These are both faster reductions than the curvature term $k/a^2$. So if the curvature term is not to totally dominate in the present Universe, it must have begun much smaller than the other terms.

The equations for $|\Omega_{\text{tot}} - 1|$ derived above stop being valid once the curvature or cosmological constant terms are no longer negligible, since we used the $a(t)$ solutions for the flat geometry to derive them. But they are fine to give us an approximate idea of what the problem is. For extra ease let's assume that the Universe always has only radiation in it. Using the equations above, we can ask how close to one the density parameter must have been at various early times, based on the constraint today ($t_0 \simeq 4 \times 10^{17}$ sec).

- At decoupling ($t \simeq 10^{13}$ sec), we need $|\Omega_{\text{tot}} - 1| \leq 10^{-5}$.
- At matter-radiation equality ($t \simeq 10^{12}$ sec), we need $|\Omega_{\text{tot}} - 1| \leq 10^{-6}$.
- At nucleosynthesis ($t \simeq 1$ sec), we need $|\Omega_{\text{tot}} - 1| \leq 10^{-18}$.
- At the scale of electro-weak symmetry breaking, which corresponds to the earliest known physics ($t \simeq 10^{-12}$ sec), we need $|\Omega_{\text{tot}} - 1| \leq 10^{-30}$. 
Written out in long hand, that means we know that at nucleosynthesis, an era we are supposed to understand very well indeed, the density parameter must have lain within the range $0.9999999999999999 \leq \Omega_{\text{tot}} \leq 1.00000000000000000001$!! Out of all the possible values that it might have had, this seems a very restrictive range. Any other value would lead to a Universe extremely different to that which we see.

The easiest way out of this dilemma is to suppose that the Universe must have precisely the critical density. But on the face of it there seems no reason to prefer this choice over any other. What would be nice would be an explanation of such a value.

Regardless of whether or not we understand the physical origin of these numbers, they are an observed fact. One useful thing they tell us is that the Universe is very close to spatial flatness at decoupling and at nucleosynthesis, which means that it is always a good approximation to set $k = 0$ in the Friedmann equation when describing those phenomena.
The horizon problem is the most important problem with the Hot Big Bang model, and refers to communication between different regions of the Universe. The crucial ingredient is that the Universe has only a finite age, and so even light can only have travelled a finite distance by any given time. As I have remarked, the distance which light could have travelled during the lifetime of the Universe gives rise to a region known as the observable Universe. This is the region we can actually see, and is always finite regardless of whether or not the Universe as a whole is finite or infinite.

One of the most important properties of the microwave background is that it is very nearly isotropic. That is, light seen from all parts of the sky possesses, to very great accuracy, the same temperature of 2.725 K. Being at the same temperature is the characteristic of thermal equilibrium, and so this observation is naturally explained if different regions of the sky have been able to interact and move towards thermal equilibrium. Unfortunately, the light we see from opposite sides of the sky has been travelling towards us since decoupling, close to the time of the Big Bang itself. Since the light has only just reached us, it can’t possibly have made it all the way across to the opposite side of the sky. Therefore there has not been time for two regions on opposite sides of the sky to interact in any way, and so one cannot claim that the regions have the same temperature because they have interacted and established thermal equilibrium. This is illustrated in Figure 13.1.
WMAP satellite’s map of CMB

http://wmap.gsfc.nasa.gov/media/121238/index.html
Figure 13.1  An illustration of the horizon problem. We receive microwave radiation from points A and B on opposite sides of the sky. These points are well separated and would not have been able to interact at all since the Big Bang — the dotted lines indicate the extent of regions able to influence points A and B by the present — far less manage to interact by the time the microwave radiation was released. So in the Hot Big Bang model it is impossible to explain why they have the same temperature to such accuracy.
In fact the problem is even worse, because the microwaves have been travelling uninterrupted since decoupling. Regions would have had to interact and thermalize even before then, by which time light could only have travelled a very short distance indeed — the observable Universe is much smaller at early times as light could have travelled much less far. So it transpires that even regions which appear quite close to each other on the sky (any points more than about a degree or two apart — see Problem 13.4) would not have been able to interact before decoupling to establish thermal equilibrium.

The final twist in the tail, which elevates this to a problem of extreme relevance, is that actually the microwave background is not perfectly isotropic, but instead exhibits small fluctuations (about one part in one hundred thousand) as detected by the COBE satellite. These irregularities are thought to represent the ‘seeds’ from which structure in the Universe grows, as described in Advanced Topic 5. For the same reason that one cannot thermalize separated regions, one also cannot create an irregularity. So in the standard Big Bang theory one cannot have a theory allowing the generation of the seed perturbations — they have to be there already.

Opposite sides of the visible universe haven’t had enough time to thermalize (reach thermal equilibrium) since the big bang - they are only now receiving information about each other. So they certainly could not have thermalized in the time between the big bang and CMB emission… unless conventional matter and radiation isn’t the whole story…
Alan Guth proposed **inflation** in 1981 as a solution to all of these problems. Stripped to its bare bones, inflation is defined as a period in the evolution of the Universe during which the scale factor was accelerating

\[
\text{INFLATION} \iff \ddot{a}(t) > 0.
\]  

(13.6)

Typically this corresponds to a very rapid expansion of the Universe.

Looking at the acceleration equation

\[
\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left( \rho + \frac{3p}{c^2} \right),
\]  

(13.7)

we see immediately that this implies \( \rho c^2 + 3p < 0 \). Since we always assume a positive density, this requires a negative pressure,

\[
p < -\frac{\rho c^2}{3}.
\]  

(13.8)

Fortunately, modern particle physics ideas of symmetry breaking indicate ways in which this negative pressure can be brought about, described later in this chapter.
The classic example of inflationary expansion is a Universe possessing a cosmological constant $\Lambda$. This is equivalent to having a fluid with $p = -\rho c^2$ (see Section 7.2), which satisfies the condition above. We saw in Chapter 7 that the full Friedmann equation, including other matter terms and curvature, becomes

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3}. \tag{13.9}$$

If all the terms on the right-hand side are significant, then this is quite complex. Fortunately, though, the situation quickly becomes more simple, because the first two terms are rapidly reduced by the expansion while the last one remains constant. So after a while, only the cosmological constant term will be significant and we will have

$$H^2 = \frac{\Lambda}{3}. \tag{13.10}$$

Recalling that $H = \dot{a}/a$, this means

$$\dot{a} = \sqrt{\frac{\Lambda}{3}} a, \tag{13.11}$$

which, since $\Lambda$ is a constant, has the solution

$$a(t) = \exp \left( \sqrt{\frac{\Lambda}{3}} t \right). \tag{13.12}$$
Inflation greatly increases the size of a region of the Universe, while keeping its characteristic scale, the Hubble scale, fixed. This means that a small patch of the Universe, small enough to achieve thermalization before inflation, can expand to be much larger than the size of our presently observable Universe, as shown in Figure 13.3. Then the microwaves coming from opposite sides of the sky really are at the same temperature because they were once in equilibrium. Equally, this provides the opportunity to generate irregularities in the Universe which can lead to structure formation.

Another way of expressing the resolution of the horizon problem is to say that, because of inflation, light can travel a much greater distance between the Big Bang and the time of decoupling than it can between decoupling and the present, reversing the usual state of affairs.
Figure 13.3  A schematic illustration of the inflationary solution to the horizon problem, with a small initial thermalized region blown up to encompass our entire observable Universe.
The standard analogy for this solution to the flatness problem is to imagine a balloon being very rapidly blown up, say to the size of the Sun; its surface would then look flat to us. The crucial difference inflation introduces compared to the usual Big Bang case is that the size of the portion of the Universe you can observe, given roughly by the Hubble length $cH^{-1}$ (since $H^{-1}$ is roughly the age of the Universe and $c$ the maximum speed) does not change while this happens. So very quickly you are unable to notice the curvature of the surface. By contrast, in the Big Bang scenario the distance you can see increases more quickly than the balloon expands, so you can see more of the curvature as time goes by.

Inflation predicts a Universe extremely close to spatial flatness. If one allows the possibility of a cosmological constant in the present Universe, then a flat Universe requires

$$\Omega_0 + \Omega_\Lambda = 1.$$  \hspace{1cm} (13.16)

Current observations, particularly of cosmic microwave background anisotropies, strongly suggest that this condition is indeed satisfied. So far, then, this simple prediction of inflation stands up well to confrontation with observations.
Recall that we rewrote the Friedmann equation as

$$|\Omega_{\text{tot}}(t) - 1| = \frac{|k|}{a^2 H^2}.$$  \hfill (13.13)

In the Big Bang theory, the problem was that this always increases with time, forcing $\Omega_{\text{tot}}$ away from one.

Inflation reverses this state of affairs, because

$$\ddot{a} > 0 \quad \Rightarrow \quad \frac{d}{dt} (\dot{a}) > 0 \quad \Rightarrow \quad \frac{d}{dt} (aH) > 0.$$  \hfill (13.14)

So the condition for inflation is precisely that which drives $\Omega_{\text{tot}}$ towards one rather than away from one. In the special case of perfect exponential expansion, the approach is particularly dramatic

$$|\Omega_{\text{tot}}(t) - 1| \propto \exp \left( -\sqrt{\frac{4\Lambda}{3}} t \right).$$  \hfill (13.15)

The aim is to use inflation not just to force $\Omega_{\text{tot}}$ close to one, but in fact to make it so extraordinarily close to one that even all the subsequent expansion between the end of inflation and the present is insufficient to move it away again, as shown in Figure 13.2. In the next section we’ll see how much inflation that entails.
Figure 13.2  Possible evolution of the density parameter $\Omega_{\text{tot}}$. There might or might not be a period before inflation, indicated by the dashed line. Inflation then drives $\log \Omega_{\text{tot}}$ towards zero (i.e. $\Omega_{\text{tot}}$ towards 1), either from above or below. By the time inflation ends $\Omega_{\text{tot}}$ is so close to one that all the evolution after inflation up to the present day is not enough to pull it away again. Only some time in the very distant future would it start to move away from one again.
How much Inflation is enough?

We can use the flatness problem to estimate how much expansion is needed from inflation. I’ll make the following simplifying assumptions, all of which could be relaxed for a better calculation.

- Inflation ends at $10^{-34}$ sec.
- The inflationary expansion is perfectly exponential.
- The Universe is perfectly radiation dominated all the way from the end of inflation to the present.
- The value of $\Omega_{\text{tot}}$ near the start of inflation is not hugely different from one.
- For the sake of argument, assume the present value of $|\Omega_{\text{tot}} - 1| \leq 0.1$.

The present age of the Universe is about $4 \times 10^{17}$ sec. During radiation domination

$$|\Omega_{\text{tot}}(t) - 1| \propto t,$$

so

$$|\Omega_{\text{tot}}(t_0) - 1| \leq 0.1 \implies |\Omega_{\text{tot}}(10^{-34} \text{ sec}) - 1| \leq 3 \times 10^{-53}. \quad (13.18)$$
During inflation $H$ is constant, so

$$|\Omega_{\text{tot}}(t) - 1| \propto \frac{1}{a^2}.$$  \hspace{1cm} (13.19)

So the required value at the end of inflation can be achieved provided that during inflation $a$ is increased by a factor of at least $10^{27}$!! Incredibly, by the standards of what comes out of inflation model building this isn’t much at all. Expansion by factors like $10^{10^8}$ are not uncommon!

This can all happen very quickly. Suppose for example that the characteristic expansion time, $H^{-1}$, is $10^{-36}$ sec. Then between $10^{-36}$ sec and $10^{-34}$ sec, the Universe would have expanded by a factor

$$\frac{a_{\text{final}}}{a_{\text{initial}}} \approx \exp[H(t_{\text{final}} - t_{\text{initial}})] = e^{99} \approx 10^{43}.$$  \hspace{1cm} (13.20)

The exponential expansion is so dramatic that very large expansion factors drop out almost automatically.
Resolving the Monopole Problem
(or missing heavy particle problem)

The dramatic expansion of the inflationary era dilutes away any unfortunate relic particles, because their density is reduced by the expansion more quickly than the cosmological constant. Provided enough expansion occurs, this dilution can easily make sure that the particles are not observable today; in fact, rather less expansion is needed than to solve the other problems.

One important proviso though is that the decay of the cosmological constant which ends inflation must not regenerate the troublesome particles again. This means that the temperature which the Universe is at after inflation must not be too high, in order to make sure there is no new thermal production.
So we have saved the Big Bang Model from fine tuning, but we have added some very speculative physics...

After some amount of time, inflation must come to an end, with the energy in the cosmological constant being converted into conventional matter. One should think of this as a decay of the particles acting as the cosmological constant into normal particles. The Big Bang can then proceed just as before. Provided all this happens when the Universe was extremely young, none of the successes of the Hot Big Bang model are lost. In typical models the Universe is extremely young indeed when inflation is supposed to occur, perhaps around $10^{-34}$ sec which is about the time appropriate to the Grand Unification scale of $10^{16}$ GeV — see equation (11.11).
The way I've discussed inflation, defining it as a period of accelerated expansion and showing that, for example, a cosmological constant can give such behaviour, is fine for developing an understanding of what inflation is and why it can solve the various cosmological problems. However, simply postulating a cosmological constant and claiming that it is able to decay away after having done its work is clearly a very *ad hoc* approach. A true model of inflation should contain a reasonable hypothesis for the origin of the cosmological constant, and a natural way of bringing inflation to an end.

To find such a model, we have to search the realms of particle physics. Remember that we must not spoil nucleosynthesis, so the very latest that inflation could have happened was when the Universe was one second old. We saw in Chapter 11 that this already corresponds to temperatures of over $10^{10}$ K, and in fact typical inflation models happen at much earlier times, and hence hotter environments, than that. In order to describe such extreme physical conditions, in which violent particle collisions are the norm, fundamental particle physics is required, and in particular theories of the fundamental interactions. Inflation is assumed to be driven by a new, as-yet-undiscovered, form of matter required by such theories.

A key idea is that of phase transitions. A phase transition corresponds to a dramatic change in the properties of a physical system as it is heated or cooled. Familiar examples are the freezing of water into ice, the lining up of domains in a cooled ferromagnet, or the onset of superconductivity or superfluidity at low temperatures. It is believed that the Universe itself will have undergone a series of phase transitions as it cooled, an example being when quarks first condensed to form hadrons.
A phase transition is a particularly dramatic event in the history of the Universe, a time when its properties change substantially. Phase transitions are controlled by an unusual form of matter known as a scalar field. Depending on the precise nature of the transition, scalar fields can behave with a negative pressure, and can satisfy the inflationary condition $\rho c^2 + 3p < 0$. That is, they behave like an effective cosmological constant. Once the phase transition comes to an end, the scalar field decays away and the inflationary expansion terminates, hopefully having achieved the necessary expansion by a factor of $10^{27}$ or more.

Inflation is currently a very active research field, and most of the study is carried out under the general hypothesis that inflation is driven by a scalar field. The hope is that eventually some specific particle physics phase transition can be identified which is likely to be the one giving inflation. Early work focussed on the Grand Unification phase transition, where the strong nuclear force first obtains an identity distinct from the electro-weak force (which will itself later split into the weak nuclear force and the electromagnetic interaction). This is believed to have happened at the very high energy of $10^{16}$ GeV, when the Universe was only $10^{-34}$ sec old, and was the example I used in working out the amount of inflation required.
More recently, attention has focussed on a different idea known as supersymmetry, already invoked in Chapter 9 to give a dark matter candidate. Supersymmetry postulates that every fundamental particle we know about, such as photons, electrons and quarks, has a partner particle with similar properties but with a higher mass. This higher mass makes them very difficult to create using particle accelerators, which is why they have yet to be seen in experiments (apart from the obvious possibility that they haven’t been seen because they are a figment of particle physicists’ imaginations). In the early Universe, the particles and their partners would have had very similar properties, and then a phase transition would lead to their present, more separate, identities. Currently, supersymmetric theories of particle physics appear the best prospect for creating models for the inflationary expansion. However, there are now a very large number of different models of inflation, and one of the goals of cosmology is to narrow this down to a favoured model or, alternatively, to disprove the inflation theory.