MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — December 1, 2009

Prof. Alan Guth

Lecture Notes 13
THE NEW INFLATIONARY UNIVERSE

INTRODUCTION:

The new inflationary universe is a scenario in which the mass density of the early universe becomes dominated by the potential energy of a scalar field, in a state which is sometimes called a *false vacuum.* This peculiar form of energy leads to a negative pressure, and hence a repulsive gravitational force, driving the universe into a period of exponential expansion, during which the early universe expands by many orders of magnitude — hence the name "inflationary". The word "new" refers to a modification of my original proposal[1] which was suggested independently by Linde[2] and by Albrecht and Steinhardt.[3] They suggested a new mechanism by which the exponential expansion phase could be ended, solving some crucial problems that were created by my original proposal. The inflationary model is very attractive because it offers possible solutions to the horizon/homogeneity problem, the flatness problem, and the magnetic monopole problem, which were discussed in Lecture Notes 10 and 12. If correct, it would also mean that particle physics mechanisms are responsible for the production of essentially all the matter, energy, and entropy in the observed universe.

SCALAR FIELDS AND THE FALSE VACUUM:

The (original) inflationary universe scenario was developed to solve the magnetic monopole problem, but it quickly became clear that the scenario might solve all three of the problems discussed in Lecture Notes 10 and 12. The scenario contained the basic ingredients necessary to eliminate these problems, but unfortunately the scenario also contained one fatal flaw: the exponential expansion was terminated by a phase transition that occurred by the random nucleation of bubbles of the new phase, very similar to the way that water boils. It was found, however, that this violent boiling would create gross inhomogeneities in the universe. This difficulty — often called the *graceful exit problem* — was summarized in the original paper [1] (in a section credited to Erick Weinberg and Harry Kesten as well as me), and was later discussed in detail by Erick Weinberg and me[4] and by Stephen Hawking,

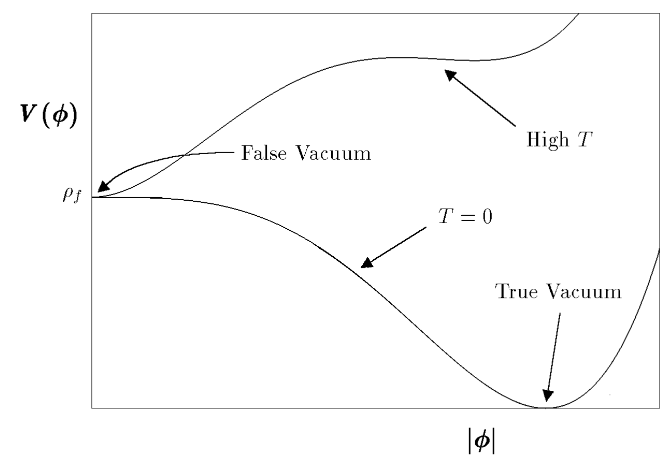
[1] A. H. Guth, *Phys. Rev. D* **23**, 347 (1981).
[2] A. D. Linde, *Phys. Lett.* **108B**, 389 (1982).
[3] A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).
[4] A. H. Guth and E. J. Weinberg, *Nucl. Phys.* **B212**, 321 (1983)

Ian Moss, and John Stewart.[5] This flaw is completely avoided in a variation known as the new inflationary universe, developed independently by Andrei Linde (then at the Lebedev Physical Institute in Moscow, now at Stanford) and by Andreas Albrecht and Paul Steinhardt. (Albrecht and Steinhardt were both at the University of Pennsylvania at the time of their discovery; now Albrecht is at UC Davis, and Steinhardt is at Princeton.)

In order for the new inflationary scenario to occur, the underlying particle theory must contain a scalar field ϕ. The potential energy function $V(\phi)$, which represents the potential energy per unit volume, must have a plateau. This plateau is usually taken to be at $\phi \approx 0$, and $\phi = 0$ is usually assumed to be a local maximum of $V(\phi)$. $V(\phi)$ must be very flat in the vicinity of $\phi \approx 0$. In the example shown below, V is assumed to depend only on $|\phi|$.

To discuss the issues one at a time, I will first discuss the physical properties of a scalar field of the type described in the previous paragraph, and then we will consider the role that such a field might play in the early universe.

In most theoretical models of this type, one finds that at high temperature T the thermal equilibrium value of ϕ lies at $\phi = 0$. At high temperatures the field will actually fluctuate wildly, but in most theoretical models the ***average*** value is predicted to be zero. A potential energy function of this general form is shown below. The curve labeled "High T" is actually a graph of the free energy per unit volume; it will not be important for us to know exactly what free energy is, but to interpret the graph we should keep in mind that the free energy is minimized in the thermal equilibrium state.



[5] S. W. Hawking, I. G. Moss, and J. M. Stewart, *Phys. Rev. D* **26**, 2681 (1982).

The scalar field ϕ that drives the inflation was originally taken to be the Higgs field of a grand unified theory, but it is now known that this does not work. The Higgs fields are required to have relatively strong interactions in order to induce spontaneous symmetry breaking, which is why the Higgs fields were introduced in the first place. These interactions lead to large quantum fluctuations in the evolution of the field, which in turn lead to unacceptably large inhomogeneities in the mass density of the universe. One must therefore assume the existence of another scalar field, similar to the Higgs field but much more weakly interacting. This field is now frequently called the *inflaton.*

So, in thermal equilibrium at high temperatures, one expects the scalar field to have a mean value around zero. If the system cools, the thermal excitations will disappear, and the scalar field will find itself in a state of essentially zero temperature, with $\phi \approx 0$. This state is called the false vacuum, and its peculiar properties are the driving force behind the inflationary model.

The false vacuum is clearly unstable, as ϕ will not remain forever at a local maximum of $V(\phi)$. However, if $V(\phi)$ is sufficiently flat, then the time that it takes for ϕ to move away from $\phi = 0$ can be very long compared to the time scale for the evolution of the early universe. Thus, for these purposes the false vacuum can be considered metastable. Furthermore, while ϕ remains near zero, the energy density remains fixed near $V(0)$, and cannot be lowered even if the universe is expanding. It is this property that motivates the name, "false vacuum." To a particle physicist, the vacuum is defined as the state of lowest possible energy density. The adjective "false" is used here to mean "temporary," so a false vacuum is a state which temporarily has the property that its energy density cannot be lowered.*

Since the false vacuum has $\phi = 0$ and no other excitations, the mass density has a fixed value which is determined by the potential energy function $V(\phi)$. For a typical grand unified theory, this value can be estimated in terms of the GUT energy scale $E_{\rm GUT} \approx 10^{16}$ GeV by using dimensional analysis:

$$\rho_{\rm f} \approx \frac{E_{\rm GUT}^4}{\hbar^3 c^5} = 2.3 \times 10^{81}\ {\rm g/cm}^3 \ . \tag{13.1}$$

The pressure p of the false vacuum is completely determined by the fact that, on the time scales of interest, its energy density cannot be lowered. To see that a

* Historically, the phrase "false vacuum" was first used to refer to a state in which the scalar field was at a local minimum of the potential energy function, so the state could decay only by quantum mechanical tunneling. Here I have stretched the definition a bit, using the phrase to describe a scalar field which, although still quite stable, is near a local maximum of the potential energy function.

constant energy density implies a negative pressure, remember the conservation of energy equation derived in Problem 2 of Problem Set 6:

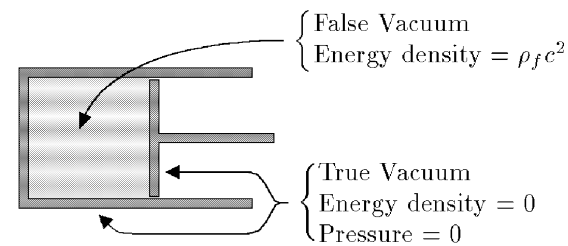
$$\dot\rho = -3\frac{\dot a}{a}\left(\rho + \frac{p}{c^2}\right) \ . \tag{13.2}$$

If $\dot\rho = 0$, this equation implies immediately that

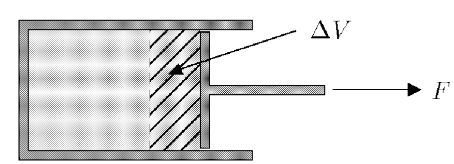
$$\boxed{p = -\rho_{\rm f} c^2 \ .} \tag{13.3}$$

(We used this same argument in Lecture Notes 8, when we were discussing vacuum energy density and the cosmological constant.)

To understand this result from first principles, think of an imaginary piston that is filled with false vacuum and surrounded by ordinary true vacuum, as shown below:



The true vacuum has zero energy density and zero pressure; it may not be exactly zero (as discussed in Lecture Notes 8), but it is in any case vastly smaller than the values that we are discussing for the false vacuum. Suppose now that the piston is pulled out so that the volume of the chamber increases by ΔV. The energy of the system then increases by $\rho_{\rm f} c^2 \Delta V$, and therefore the agent that moved the piston must have done precisely this amount of work.



$$\Delta W = \rho_f c^2\, \Delta V = -p\, \Delta V$$

Since the pressure on the outside is zero, the agent must be pulling against a negative pressure, which would oppose the motion. Quantitatively, since the work done is $-p\Delta V$, it follows that $p = -\rho_f c^2$, confirming the previous result.

The large negative pressure creates a gravitational repulsion, exactly as we discussed in Lecture Notes 8 in the context of a cosmological constant. The gravitational repulsion can be seen in the second order differential equation for a,

$$\ddot{a} = -\frac{4\pi}{3} G \left(\rho + \frac{3p}{c^2} \right) a \ , \tag{13.4}$$

which implies that both the pressure and the energy density normally contribute to the slowing of the cosmic expansion. For the false vacuum, however, the large negative pressure leads to $\rho + 3p/c^2 < 0$, and it follows that $\ddot{a}$ is ***positive***. The false vacuum creates a gravitational repulsion which causes the growth of the scale factor a to accelerate. It is this repulsion which will drive the colossal expansion of the inflationary scenario. The equations are the same as those for a cosmological constant, except that the false vacuum energy density disappears when the scalar field rolls off the hill in the potential energy diagram, while the vacuum energy associated with a cosmological constant is permanent.

THE NEW INFLATIONARY UNIVERSE:

We can now go through the new inflationary scenario step by step. The starting point of a cosmological scenario is, unfortunately, still somewhat a matter of taste and philosophical prejudice. Some physicists find it plausible to assume that the universe began in some highly symmetrical state. Many others, however, consider it more likely that the universe began in a highly chaotic state, since the number of chaotic configurations is presumably much larger. One advantage of the inflationary scenario, from my point of view, is that it appears to allow a wide variety of starting configurations.

We can begin by discussing what would happen if the early universe were in thermal equilibrium, at least in the sense of having regions of approximately horizon size in which thermal equilibrium held. In that case, inflation could begin if the universe was hot ($kT > 10^{16}$ GeV) in at least some of these regions, and if at least one of these hot regions were expanding rapidly. In the hot regions, thermal equilibrium would imply $<\phi> = 0$, where $<\phi>$ denotes the mean value of the field ϕ as it undergoes its thermal fluctuations. Rapid expansion would cause these regions to cool, and the scalar field would settle down to a cool state in which the field is trapped on the plateau of the potential energy hill. The expansion must be rapid enough so that the cooling of the scalar field occurs before the region recollapses under the influence of gravity.

Thermal equilibrium would make things simple, but we said earlier that the inflaton field must interact very weakly, to avoid generating overly large quantum fluctuations. For such a weakly interacting field, a fairly straightforward calculation of collision rates shows that the mean time between collisions would be long compared to the age of the universe at the onset of inflation. Thus there is no compelling reason to assume thermal equilibrium, although — in the absence of a theory that fixes the initial conditions — one could assume anything one wants. For inflation to start, the minimal assumption would be that there existed at least some regions of high energy density with $<\phi> \approx 0$, and that at least one of these regions was expanding rapidly enough so that ϕ became trapped in the false vacuum.

The above paragraphs describe the new inflationary universe with a hot beginning, but there are certainly other possibilities. Linde has also proposed the idea of chaotic inflation, in which inflation is driven by a scalar field which is initially chaotic but far from thermal equilibrium. In this scenario inflation happens while the scalar field rolls down a gentle hill in the potential energy diagram, so the potential energy diagram need not have a plateau. Alexander Vilenkin (of Tufts University) and Linde have investigated speculative but attractive scenarios in which the universe is created by a quantum tunneling event, starting from a state of absolutely nothing. In these models the universe enters directly into a de Sitter phase. In a similar spirit James Hartle (of the University of California at Santa Barbara) and Stephen Hawking (of Cambridge University) have proposed a unique quantum wave function for the universe, incorporating dynamics which leads to an inflationary era. Later Hawking and Neil Turok (also of Cambridge University) explored a variant of this idea, in which the resulting universe is open rather than closed.

Although a wide variety of scenarios have been proposed to describe the onset of inflation, an important feature of inflation is that all these scenarios lead to similar if not identical predictions. Once inflation starts, the colossal expansion dilutes away the evidence of how it began. Later I will discuss the phenomenon of *eternal inflation,* which carries this idea of dilution to an extreme. We will see that for almost all inflationary models, once inflation starts, it never stops. Instead it goes on producing "universes" forever. This eternal aspect of inflation presumably erases all traces of how inflation began, and it also obviates the question of whether the conditions leading to inflation are likely. As long as the probability that inflation can start is nonzero, it appears (at least to this author) that there are no other questions about initial conditions that need to be answered. An ultimate theory of the origin of the universe would still be very interesting, intellectually, but most likely it would not affect in any way the consequences of inflation.

To continue with the description of the new inflationary scenario, it is easiest to begin by assuming that the region is homogeneous, isotropic, and flat. (Later I will describe what happens when this assumption is dropped.) The region can then be described by the Robertson-Walker flat ($k = 0$) metric

$$ds^2_{\rm ST} = -c^2 dt^2 + a^2(t) d\vec{x}^{\,2} \ , \tag{13.5}$$

and the equation of motion becomes

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho \ . \tag{13.6}$$

The solution is given by

$$\boxed{a(t) = \text{const} \times e^{\chi t} \ ,} \tag{13.7}$$

where

$$\boxed{\chi = \sqrt{\frac{8\pi}{3} G\rho_{\rm f}} \ .} \tag{13.8}$$

This exponential expansion is of course the hallmark of the inflationary model. (For our parameters, $\chi^{-1} \approx 10^{-37}$ sec.) Such a space is called a de Sitter space.

Now let us consider what would happen if the initial region were not homogeneous and isotropic. In that case, one must examine the behavior of perturbations about the Robertson-Walker metric. These perturbations seem to be governed by a "cosmological no-hair theorem", which states that whenever $p = -\rho c^2 = constant$, then any locally measurable perturbation about the de Sitter metric is damped exponentially on the time scale of χ^{-1}. Any initial particle density is diluted to negligibility, and any initial distortion of the metric is stretched (i.e., redshifted) until it is no longer locally detectable. The theorem has been proven only in the context of linearized approximations, but it is believed by many physicists (Stephen Hawking, myself, and others) to be valid in all cases. Thus, a smooth de Sitter metric arises naturally, without any need to fine-tune the initial conditions.

As the space continues to exponentially expand, the mass density of the inflaton field is fixed at $\rho_{\rm f}$. Thus, the total energy of the inflaton field is increasing! If the inflationary model is right, the energy of the inflaton field is the source of essentially all the matter, energy, and entropy in the observed universe.

This creation of energy seems to violate our naive notions of energy conservation, but we must remember that there is also an energy associated with the cosmic gravitational field— the field by which everything in the universe is attracting everything else, thereby slowing down the cosmic expansion. Even in Newtonian

mechanics one can see that the energy density of a gravitational field is ***negative***. To see this, note that the gravitational field is strengthened as one brings masses together from infinity, but the potential energy of the system is lowered. Thus the stronger field corresponds to a lower energy. A good analogy is the electrostatic field, since Coulomb's law is very similar to Newton's law. By calculating how much work needs to be done by pushing charges to create a specified configuration of a static electric field, it is possible to show that the energy density stored in an electric field is given by

$$u_{\rm electrostatic} = \frac{1}{2}\epsilon_0 \left|\vec{E}\right|^2 \tag{13.9a}$$

or

$$u_{\rm electrostatic} = \frac{1}{8\pi} \left|\vec{E}\right|^2 \ , \tag{13.9b}$$

depending on what units you are using. The calculation for Newtonian gravity is essentially identical, giving

$$u_{\rm Newton} = -\frac{1}{8\pi G} \left|\vec{g}\right|^2 \ . \tag{13.10}$$

The sign difference arises from the sign difference in the force law: two positive charges repel, while two positive masses attract. In the context of inflation, the energy stored in the gravitational field becomes more and more negative as the universe inflates, while the energy stored in "matter" (everything except gravity) becomes more and more positive. The total energy remains constant, and very small— perhaps it is exactly equal to zero.

After the region has undergone exponential expansion for some time, inflation must somehow end, at least in the region that is going to describe our visible universe. The scalar field is in an unstable configuration, perched at the top of the hill of the potential energy diagram shown on p. 2. It will undergo fluctuations due to thermal and/or quantum effects. Some fluctuations begin to grow, and at some point these fluctuations become large enough so that their subsequent evolution can be described by the classical equations of motion. I will use the term "coherence region" to denote a region within which the scalar field is approximately uniform. The coherence regions are irregular in shape, and their initial size is typically of order $c\chi^{-1}$. Note that $c\chi^{-1}$ is only about 10^{-14} proton diameters; the entire observed universe will evolve from a region of this size or smaller.

The scalar field ϕ then "rolls" down the potential energy function shown on p. 2, obeying the classical equations of motion derived from general relativity. As long as the spatial variations in ϕ are small, these classical equations take the form

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} = -\frac{\partial V}{\partial \phi} \ . \tag{13.11}$$

(The derivation of Eq. (13.11) is a straightforward application of general relativity, but it is a little beyond the scope of this course.) If the initial fluctuation is small, then the flatness of the potential for $\phi \approx 0$ will imply that the rolling begins very slowly. Note that the second term on the left-hand-side of Eq. (13.11) is a damping term, helping to slow down the speed of rolling. As long as $\phi \approx 0$, the mass density ρ remains about equal to $\rho_{\rm f}$, and the exponential expansion continues. The expansion occurs on a time scale χ^{-1}, while the time scale of the rolling is much slower. This "slow roll" of the scalar field is the crucial new feature in the ***new*** inflationary universe.

For the scenario to work, it is necessary for the length scale of homogeneity to be stretched from $c\chi^{-1}$ to at least about 10 cm before the scalar field ϕ rolls off the plateau of the potential energy diagram. This corresponds to an expansion factor of about 10^{28}, which requires about 65 time constants (χ^{-1}) of expansion. The expected duration of the expansion depends on the precise shape of the scalar field potential, and models have been constructed which yield much more than the minimally required amount of inflation.

When the ϕ field reaches the steep part of the potential, it falls quickly to the bottom and oscillates about the minimum. The time scale of this motion is a typical GUT time of $\hbar/E_{\rm GUT} \approx 7\times 10^{-41}$ sec, which is very fast compared to the expansion rate. The scalar field oscillations are then quickly damped by the couplings to the other fields, and the energy is rapidly converted into a thermal equilibrium mixture of particles. (From a particle point of view, the scalar field oscillations correspond to a state of spinless particles, just as an oscillating electromagnetic field corresponds to a state of photons. The damping of the scalar field is just the field theory description of the decay of these particles into other kinds of particles.) The release of this energy reheats the region back to a temperature which can be of order $kT \approx 10^{16}$ GeV, or can be much lower, depending on the strength of the interactions.

From here on the standard scenario takes over. The era of inflation has set up precisely the initial conditions that had previously been assumed in standard cosmology. You can check that a region of radius $\approx$ 10 cm, at a temperature $kT \approx 10^{16}$ GeV, will become large enough by the time T falls to 2.7 K to encompass the entire observed universe.

SOLUTIONS TO THE COSMOLOGICAL PROBLEMS:

Let me now explain how the three problems of the standard cosmological scenario discussed in Lecture Notes 10 and 12 are avoided in the inflationary scenario. First, let us consider the horizon/homogeneity problem. The problem is clearly avoided in this scenario, since the entire observed universe evolves from a single

coherence region. This region had a size of order $c\chi^{-1}$ at the time when the fluctuation began to grow classically. This size is much smaller than the sizes that are relevant in the standard model at these times, and the region therefore had plenty of time to come to a uniform temperature before the onset of inflation. The exponential expansion causes this very small region of homogeneity to grow to be large enough to encompass the observed universe.

The flatness problem is avoided by the dynamics of the exponential expansion of the coherence region. As ϕ begins to roll very slowly down the potential, the evolution of the metric is governed by the mass density $\rho_{\rm f}$. Assuming that the coherence region (or a small piece of it) can be approximated by a Robertson-Walker metric, then the scale factor evolves according to the standard Friedmann equation:

$$\left(\frac{\dot a}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2}\ , \tag{13.12}$$

where $k = +1,\ -1$, or 0 depending on whether the region approximates a closed, open, or flat universe, respectively. (There could also be perturbations, but the cosmological no-hair theorem guarantees that they would die out quickly.) In this language, the flatness problem is the problem of understanding why the kc^2/a^2 term on the right-hand-side is so extraordinarily small. But as the coherence region expands exponentially, the mass density ρ remains very nearly constant at $\rho_{\rm f}$, while the kc^2/a^2 term is suppressed by at least a factor of $\left(10^{28}\right)^2 = 10^{56}$. This provides a "natural" explanation of why the value of the kc^2/a^2 term immediately after the phase transition is smaller than that of the other terms by a tremendous factor.

Except for a very narrow range of parameters, this suppression of the curvature term will vastly exceed that required by present observations. This leads to the prediction that the kc^2/a^2 term of Eq. (13.12) should remain totally negligible until the present era, and even far into the future. This implies that the value of Ω today is expected to be equal to one with a high degree of accuracy.

The inflationary prediction that $\Omega = 1$ seemed to be at odds with observation until the last few years. Astronomers have never found enough matter to make up a critical mass density, although there were some measurements of the large-scale flows in the universe which found velocities that were high enough to indicate a critical density in some kind of unclumped matter; that is, matter that does not clump on the scale of galaxy clusters or smaller. These observations, however, did not stand the test of time. Some inflationary theorists constructed versions of inflation that could lead to an open universe; this could be arranged by choosing the parameters to that inflation proceeds for just long enough to solve the flatness problem, but not so long that it flattened the universe completely.

But the situation changed dramatically in 1998 with the Supernova Type Ia measurements, which indicated the presence of a cosmological constant or a very slowly evolving scalar field that could simulate a cosmological constant. In either case, the total energy in this new component of the universe is about what is needed to complete the inventory for a flat universe. The best current estimate of Ω is based on the WMAP satellite data for the anisotropies of the cosmic microwave background radiation, giving $\Omega = 1.02 \pm 0.02$.

Finally, we turn to the monopole problem. Recall that in the standard scenario, the tremendous excess of monopoles was produced by the disorder in the Higgs field (i.e., by the Kibble mechanism). There is no known way to prevent the Kibble mechanism from operating, but as long as inflation occurs after or during the process of monopole formation, the monopoles will be diluted enormously. During inflation the volume of the coherence region increases by a factor of about $\left(10^{28}\right)^3 = 10^{84}$, which is enough to convert the monopole glut into a situation where no monopoles will be seen.

ETERNAL INFLATION:

We will not have time to fully discuss the mind-boggling implications of this feature, but the basic facts are pretty straight-forward. As the scalar field rolls off the potential energy plateau shown on p. 2, we must remember that in a full quantum mechanical treatment there will always be some probability that the scalar field will remain at the top of the hill. Approximate calculations show that this probability falls off exponentially with time, with a time constant that is similar to, but maybe a factor of 100 slower than, the time constant of the exponential expansion. This means that if an observer stayed at any one point of the inflating region, it is highly probable that she would see inflation end in a very short amount of time, perhaps 10^{-35} second. However, if we were to calculate how the total volume of false vacuum changes with time, we would find that the growing exponential of the expansion dominates over the falling exponential of the decay, so the total volume of false vacuum grows exponentially in time! Once inflation starts we expect it never to stop, but instead it will continue forever. The decay of the false vacuum (the transition of the scalar field to the true vacuum value) does not happen globally, but instead pieces of the false vacuum undergo the decay and produce huge regions of inhabitable space that can be called *pocket universes.* An infinite number of pocket universes are produced. Can we see these other universes? No. Is this discussion physics or metaphysics? That's debatable, but in my opinion it is physics, albeit very speculative physics at this stage. First, it seems to be an almost unavoidable consequence of inflation, which itself makes a number of testable predictions. Second, I believe that ultimately the full predictions of inflation will have to be understood in a statistical way, where the properties of the statistical ensemble will be determined by the equilibrium properties of the eternally inflating region.

CHAOTIC INFLATION:

While I have described the new inflationary model, because I think it is the simplest version to understand, there are now many variants of inflationary models. One very important variant is known as *chaotic inflation*, invented by Andrei Linde. Linde realized that in fact inflation does not require a plateau in the potential energy diagram, but can in fact happen with a potential energy function as simple as

$$V(\phi) = \frac{1}{2} m^2 \phi^2 \ , \tag{13.13}$$

which in fact describes a non-interacting particle of mass m. If the field ϕ is started at a large enough value, then sufficient inflation can occur as the scalar field rolls towards $\phi = 0$. Linde initially proposed that the scalar field could start at a large value in some places due to "chaotic" initial conditions. Later he showed that quantum fluctuations cause these models to also undergo eternal inflation, so the question of initial conditions is perhaps irrelevant.