Fermi National Accelerator Laboratory

FNAL-CONF-90/195A
September, 1990

# First-order inflation*

Edward W. Kolb

*NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory*
*Batavia, Illinois 60510-0500*
*and*
*Department of Astronomy and Astrophysics, Enrico Fermi Institute,*
*The University of Chicago, Chicago, Illinois 60637*

In the original proposal, inflation occurred in the process of a strongly first-order phase transition. This model was soon demonstrated to be fatally flawed. Subsequent models for inflation involved phase transitions that were second-order, or perhaps weakly first-order; some even involved no phase transition at all. Recently the possibility of inflation during a strongly first-order phase transition has been revived. In this talk I will discuss some models for first-order inflation, and emphasize unique signatures that result if inflation is realized in a first-order transition. Before discussing first-order inflation, I will briefly review some of the history of inflation to demonstrate how first-order inflation differs from other models.

*To appear in the proceedings of Nobel Symposium #79, *The Birth and Early Evolution of the Universe*, Gräftåvallen, Östersund, Sweden, June 11-16, 1990.

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

# I. INTRODUCTION

## *Pre-history*

The years before the birth of the inflationary Universe contained a rich pre-history of work in cosmology investigating the cosmological consequences of a Universe dominated by vacuum energy.[1] Vacuum energy is interesting in cosmology because it acts as a cosmological constant, and will drive the Universe in exponential expansion. Recall that the expansion of the Universe is determined by the Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G_N}{3}\rho, \tag{1.1}$$

where $a(t)$ is the cosmic scale factor, $\rho$ is the energy density of the Universe, and the constant $k$ is $\pm 1$ or 0 depending upon the spatial curvature. If the contribution of the vacuum energy density $\rho_V$ dominates, then $\rho$ is a constant (it does not decrease with $a$), and for $k = 0$ the solution to the Friedmann equation is

$$a(t) = a(0)\exp(Ht); \qquad H \equiv \frac{\dot{a}}{a} = \left(\frac{8\pi G_N}{3}\rho_V\right)^{1/2} = \text{ const.} \tag{1.2}$$

Such a rapid expansion may solve several cosmological problems, including the flatness/age problem, the homogeneity/isotropy problem, the problem of the origin of density inhomogeneities, and the monopole problem.[2]

The possibility of a Universe dominated by vacuum energy became much more relevant with the realization that the Universe may have undergone a series of phase transitions associated with spontaneous symmetry breaking. The work of Kirtzhnits and Linde[3] showed that symmetries that are spontaneously broken today should have been restored at temperatures above the energy scale of spontaneous symmetry breaking, and as the Universe cooled below some critical temperature, denoted as $T_C$, there should have been a phase transition in which the symmetry was broken. Thus, phase transitions associated with spontaneous symmetry breaking might offer a mechanism whereby the early Universe may be dominated by vacuum energy for some period of time.[4]

## *The Classical Era of Old Inflation*

Although there was a rich pre-history, the classical era of inflation crystallized with the paper of Guth.[5] In this classical picture, the Universe underwent a strongly first-order phase transition associated with spontaneous symmetry breaking of some Grand Unified Theory (GUT). Whether the phase transition is first order or higher order depends upon the details of the "Higgs" potential for the scalar field whose vacuum expectation value is responsible for symmetry breaking. This theory is now usually referred to as "old" inflation.

The generic shape of the zero-temperature Higgs potential for a first-order phase transition is shown in Fig. 1a. The crucial feature is the barrier separating the symmetric high-temperature minimum, here located at $\sigma = 0$, from the low-temperature true vacuum located at $\sigma \neq 0$. If the transition is strongly first order, the transition from the high-temperature to the low-temperature minimum occurs by the quantum-mechanical process of nucleation of bubbles of true vacuum. These bubbles of true vacuum expand at the velocity of light, converting false vacuum to true.[6]

The bubble nucleation rate ("per volume" will always be understood when discussing the bubble nucleation rate) depends upon the shape of the potential. More details will be given in Sec. III, but in general, it is written as

$$\Gamma = A \exp(-B), \tag{1.3}$$

where $A$ is a parameter with mass dimension 4, and $B$, the bounce action, is dimensionless. A complete explanation of $A$ and $B$ will be given below. For now, let us simply assume that $A \sim \sigma_0^4$, where $\sigma_0$ is the mass scale of spontaneous symmetry breaking (SSB).

In the classical picture, the energy density of the Universe became dominated by the false-vacuum energy of the Higgs field and the Universe expanded exponentially. Sufficient inflation was never a real concern; the problem with the classical picture is in the termination of the false-vacuum phase; usually referred to as the graceful exit problem.

Inside the true vacuum bubble is just what one expects—vacuum. For successful inflation it is necessary to convert the vacuum energy to radiation. The way this is accomplished in a first-order phase transition is through the process of collision of vacuum bubbles. In bubble collisions the energy density tied up in the bubble walls may be converted to entropy. Thus, if a first-order phase transition is to have a graceful exit, there must be many bubble collisions. The decline of the classical era began with the realization that bubbles of true vacuum do not percolate[7] and fill the Universe;[8] i.e., there is no graceful exit. The basic reason is that the exponential expansion of the background space overwhelms the bubble growth. To see this, consider the expression for the *coordinate* (or *comoving*) radius of the bubble. Assume that the bubble is nucleated at time $t_0$ with zero radius, and expands outward at the speed of light. At some time $t > t_0$ after nucleation, the comoving bubble radius is

$$r(t, t_0) = \int_{t_0}^{t} dt'\, a^{-1}(t') = \frac{\exp(-Ht_0) - \exp(-Ht)}{Ha(0)}. \tag{1.4}$$

The *physical* size of the bubble of course is simply $R(t, t_0) = a(t) r(t, t_0)$. Notice that as $t \to \infty$, the comoving bubble size approaches a *finite* value:

$$r(\infty, t_0) = \frac{\exp(-Ht_0)}{Ha(0)}. \tag{1.5}$$

Bubbles nucleated at larger $t_0$ reach a smaller comoving size than bubbles nucleated earlier in the transition. If a bubble is nucleated at time $t_0$, at some later time $t$ the bubble has comoving volume $v(t,t_0)$ and physical volume $V(t,t_0)$ given by

$$\begin{aligned} v(t,t_0) &= \frac{4\pi}{3} r^3(t,t_0) \longrightarrow \frac{4\pi}{3} \frac{\exp(-3Ht_0)}{[Ha(0)]^3} \\ V(t,t_0) &= \frac{4\pi}{3} R^3(t,t_0) \longrightarrow \frac{4\pi}{3} \frac{\exp[3H(t-t_0)]}{H^3}, \end{aligned} \tag{1.6}$$

where in this section arrows indicate the asymptotic values as $t \rightarrow \infty$.

The probability that a point remains in the old (false-vacuum) phase at time $t$ is simply

$$p(t) = \exp\left[ -\int_0^t dt_0 \, \Gamma V(t,t_0) \right] \longrightarrow \exp\left[ -\frac{4\pi}{3} \left( \frac{\Gamma}{H^4} \right) Ht \right]. \tag{1.7}$$

Thus, the probability that a point remains in the false-vacuum phase decreases exponentially in time, just as expected.

Although $p(t)$ decreases exponentially, the volume of space in the false vacuum is increasing exponentially. A measure of whether true vacuum regions will percolate the space is the fraction of physical space in false vacuum:

$$f(t) = p(t)a^3(t) \longrightarrow \exp\left[ -\frac{4\pi}{3} \left( \frac{\Gamma}{H^4} \right) Ht \right] \exp[3Ht]. \tag{1.8}$$

Clearly whether this fraction increases or decreases in time depends upon the competition between the decreasing probability for a point to be in the false vacuum and the increasing volume of space in the false vacuum. A rough estimate of whether $f(t)$ will increase or decrease is the criteria that $\epsilon = \Gamma/H^4$ is much greater or much less than unity. If $\epsilon$ is much less than one the transition will never be completed, while if $\epsilon$ is much greater than one the transition will be completed, but there won't be a sufficient period of inflation. So if $\epsilon$ is small enough to guarantee sufficient inflation, it will be too small for percolation to result.

This graceful exit problem led to the decline of the classical era of inflation and the dawn of the inflationary dark ages.

### *Slow-Rollover Renaissance of New Inflation*

Soon after the demise of the original model, inflation was revived by the realization that it was possible to have an inflationary scenario without recourse to a strongly first-order phase transition. Linde, and Albrecht and Steinhardt proposed that the Universe inflates in the process of the classical evolution of the vacuum.[9] In the classical evolution of the field to its true minimum the field has "kinetic" energy and "potential" energy. If one has a region of the scalar Higgs potential that is "flat," then the velocity of the Higgs field in the evolution to the ground state will be slow, and the potential energy of the Higgs field might dominate the kinetic energy. This can be made more quantitative by writing the classical equation of motion for a spatially homogeneous scalar field $\sigma$ in an expanding Universe under the influence of a potential $V(\sigma)$:

$$\ddot{\sigma} + 3\frac{\dot{a}}{a}\dot{\sigma} + \frac{dV(\sigma)}{d\sigma} = 0. \tag{1.9}$$

If the potential is flat enough that the $\ddot{\sigma}$ term can be neglected, the scalar field will undergo a period of "slow roll." The schematic representation of such a potential is given in Fig. 1b. The energy density contributed by the scalar field is $\rho_\sigma = \dot{\sigma}^2/2 + V(\sigma)$, and in the slow-roll region $V(\sigma) \gg \dot{\sigma}^2$, so the expansion closely approximates the exponential solution. This theory is sometimes referred to as "new" inflation.

The original proposal of slow-rollover inflation was also based upon an $SU(5)$ GUT phase transition. The potential was "flattened" by assuming that it took the Coleman-Weinberg form.[10] However it was soon realized that even this potential was not flat enough. If the scalar potential is approximated by a simple potential of the form $V(\sigma) = \lambda(\sigma^2 - \sigma_0^2)^2$, in order for density fluctuations produced in inflation to be small enough required $\lambda \lesssim \mathcal{O}(10^{-15})$.[11] Clearly such small numbers did not arise naturally in simple unified models, and a successful slow-rollover inflation model must be somewhat more complicated. Unfortunately, it was soon discovered that there is no cosmological upper bound on the complexity of models.

## *The Baroque Era*

It was soon realized that the requirement of a small coupling constant could not easily be accommodated in simple particle physics models. Of course the usual temptation is to modify the Higgs sector by adding more representations than required in the minimal model. In fact a successful model was constructed along these lines.[12]

For a while it was thought that supersymmetric GUTs could hold the key, but they were soon abandoned for a variety of reasons.[13] After supersymmetric models, some very interesting supergravity models emerged.[14] Although many supergravity models raised new problems of their own,[15] some supergravity models were quite successful, and (at least) gave a proof of existence that the inflationary scenario might be implemented in particle models.

All of these Baroque models suffered from a low re-heat temperature as a result of a weakly coupled inflaton. This made baryogenesis problematical, although not impossible. All post-renaissance inflation models involved second-order transitions, and because inflation occurred in a smooth patch of the Universe that originally contained a single correlation region, the observable Universe should contain less than one topological defect produced in the transition. This is good news for the monopole problem, but bad news for cosmic strings and texture. Furthermore, since the inflaton must be very weakly coupled, it cannot be a gauge field responsible for the formation of topological defects. We will return to this question in more detail.

## *Rococo Inflation*

The complexity of inflationary models was again increased as people started modifying the gravitational sector of the theory. In Rococo inflation the identity of the inflaton is up for grabs. There are models where the inflaton is associated with the radius of internal dimensions,[16] with the extra degree of freedom in fourth-order gravity,[17] with the scalar field of induced gravity,[18] etc. Some of these models can be made to work; it might be said that none work naturally.

Perhaps somewhere along the line as more and more detail was added to make the models satisfy all of the constraints, the message, or at least the spirit, of inflation was lost.

*Impressionism*

In response to the excesses of Baroque and Rococo inflation, there grew up around Andrei Linde a Russian school of "Impressionist" inflation. In the impressionist style no serious attempt is made to connect the details of inflaton with any specific particle physics models. In this way the true essence and beauty of the inflationary Universe is realized without any of the cluttering details. The best example of this the the "chaotic" inflation model.[19] In this model the scalar potential is assumed to be simply $V(\sigma) = \lambda\sigma^4$. What a perfect example of impressionism! This potential embodies features common to all scalar potentials without any of the details. Of course it is not "realistic" in the sense that no one would accept the existence of a scalar field whose sole purpose is to make inflation simple, but it can be taken to represent the impressions of every scalar field, while at the same time representing no scalar field.

As Linde has repeatedly emphasized, it is not even necessary to connect inflation with a phase transition. In the $\lambda\sigma^4$ chaotic model the $\sigma$ field is expected to start away from its minimum (at $\sigma = 0$) due to "chaotic" initial conditions. From there, inflation can be analyzed as in slow rollover models.

Despite the seductive beauty of the impressionist approach we must demand more realism. Eventually we want a description of the Universe that has the fine details of the Baroque or Rococo but with the simplicity and spirit of impressionism.

*The Postmodern Era*

One of the most interesting modern developments is postmodernism. The postmodern movement is characterized by an eclectic mixture of classical tradition with some aspect of the recent past. With this definition, it may be said that first-order inflationary cosmologies represent a postmodern trend. The classical tradition is a first-order transition, while the aspect of the recent past will be embodied by the slow rolling of a second scalar field.

The key to first-order inflation is the relaxation of the assumption that $\epsilon \equiv \Gamma/H^4$ is constant in time. There are two ways one might imagine a time dependence for $\epsilon$. The first way is for $H$ to change. Since $H = \sqrt{G\rho_V}$, either the effective gravitational constant $G$ must change or the vacuum energy $\rho_V$ must change.[20] (We will see that in many cases the two possibilities are equivalent representations of the same physics.)

The second way is for $\Gamma$ to change. I will also discuss this possibility. Of course, in general, both $H$ and $\Gamma$ might change.

If $\epsilon$ starts small, much less than one, then there might be a sufficient amount of inflation. If $\epsilon$ grows and eventually becomes much greater than one, then the bubbles of true vacuum will percolate and collisions between the bubble walls might convert the false-vacuum energy into entropy. This is the hope of first-order inflation.

In the next section I discuss a specific model.

## II. EXTENDED INFLATION

The simplest (and original) model of first-order inflation is extended inflation,[21] which is based upon a Jordan-Brans-Dicke (JBD) gravity theory. To illustrate extended inflation, let us consider a gravitational action including the JBD field $\Phi$ (with mass dimension 2), the metric $g$, and a scalar matter field $\sigma$:

$$\begin{aligned} S[\Phi,g,\sigma] &= \int d^4x\,\sqrt{-g} \\ &\quad\times\left[-\Phi\mathcal{R}+\omega g^{\mu\nu}\frac{\partial_\mu\Phi\partial_\nu\Phi}{\Phi}+\frac{1}{2}g^{\mu\nu}\partial_\mu\sigma\partial_\nu\sigma-V(\sigma)\right]. \end{aligned} \tag{2.1}$$

We will assume that $V(\sigma)$ is a potential that leads to a first-order phase transition. Here $\omega$ is a dimensionless parameter. Solar system measurements give a lower bound to $\omega$: $\omega \gtrsim 500$. In the limit $\omega \to \infty$, the theory becomes Einstein gravity with $\Phi = (16\pi G_N)^{-1}$. The field $\sigma$ will play the role of the "inflaton," i.e., it is the scalar field whose vacuum expectation value will drive inflation. We will assume that during inflation $\sigma$ sits quietly anchored at its false vacuum value, $\sigma_{FV}$, and its only effect is to contribute a vacuum energy $\rho_V = V(\sigma_{FV})$.

Before proceeding, let me say something about the choice of conformal frames. If the action contains a term $\mathcal{R}\Phi$, I will say we are in the *Jordan Conformal Frame.* I will soon perform a rescaling of the metric so as to remove the explicit $\Phi$ dependence from the Ricci scalar term. The resulting theory with the usual Einstein-Hilbert form for the Ricci scalar, $-\mathcal{R}/16\pi G_N$, will be the same theory expressed in what I will call the *Einstein Conformal Frame.*

The equations of motion for a spatially homogeneous Universe are

$$\begin{aligned} \left(\frac{\dot a}{a}\right)^2+\frac{k}{a^2} &= \frac{\rho_V}{6}+\frac{\omega}{6}\left(\frac{\dot\Phi}{\Phi}\right)^2-\frac{\dot a}{a}\frac{\dot\Phi}{\Phi}, \\ \ddot\Phi+3\frac{\dot a}{a}\dot\Phi &= \frac{2\rho_V}{3+2\omega}, \end{aligned} \tag{2.2}$$

where a dot denotes $d/dt$. These equations have simple power-law solutions:

$$\Phi(t)=\Phi(0)\,(1+Bt)^2, \qquad a(t)=a(0)(1+Bt)^{\omega+1/2}, \tag{2.3}$$

where

$$\begin{aligned} B &\equiv \frac{1}{\alpha}\left(\frac{\rho_V}{6\Phi(0)}\right)^{1/2}, \\ \alpha^2 &\equiv (3+2\omega)(5+6\omega)/12 \longrightarrow \omega^2 \qquad \text{as } \omega \to \infty. \end{aligned} \tag{2.4}$$

Note that $a(t)$ increases as a (large) power of time, rather than exponential as in old inflation.

Recall that the crucial parameter that determines if a first-order transition will lead to percolation is $\epsilon = \Gamma/H^4$. We will return to the possibility that $\Gamma$ varies in time, but for the moment assume that $\Gamma =$ const. The fact that $\Phi$ increases with time implies that the effective gravitational coupling $16\pi G \sim \Phi^{-1}$ will decrease with time. Therefore $H \sim \sqrt{G\rho_V}$ will decrease in time. This is exactly what happens as can be seen from Eq. 2.3:

$$H \equiv \frac{\dot{a}}{a} = \frac{(\omega+1/2)B}{1+Bt} \longrightarrow \begin{cases} (\omega+1/2)B & Bt \ll 1 \\ (\omega+1/2)/t & Bt \gg 1. \end{cases} \tag{2.5}$$

This is very promising: It is easy to arrange the initial value of $\epsilon$ to be much less than one so that sufficient inflation results, but eventually $\epsilon$, which grows as $t^4$ for $Bt \gg 1$, will become greater than one and the true-vacuum bubbles will percolate. In fact we can estimate the time of percolation of the true-vacuum bubbles by setting $\epsilon = 1$. Of course the end of extended inflation is a somewhat ill-defined time, but this definition is probably adequate for our purposes.

I mentioned previously that whether one considers extended inflation to result from a modification of gravity, or a modification of $\rho_V$ often depends upon one's frame of mind. To illustrate this, consider the theory of Eq. 2.1 expressed in a new frame, related to the original frame by a Weyl rescaling of the metric in terms of a new metric $\overline{g}$, along with a new definition of the "inflaton" degree of freedom $\psi$:

$$\begin{aligned} g_{\mu\nu} &= \Omega^2(x)\overline{g}_{\mu\nu}; \qquad \Omega^2\Phi = (16\pi G_N)^{-1}; \\ \psi &= \psi_0 \ln\left[\Phi(3+2\omega)/\psi_0^2\right]; \qquad \psi_0^2 = (3+2\omega)/16\pi G_N. \end{aligned} \tag{2.6}$$

In this new frame, the Einstein Conformal Frame, the action is

$$\begin{aligned} \overline{S}[\psi,\overline{g},\sigma] = \int d^4x \sqrt{-\overline{g}} \Big[ & -\frac{\overline{\mathcal{R}}}{16\pi G_N} + \frac{1}{2}\overline{g}^{\mu\nu}\partial_\mu\psi\partial_\nu\psi \\ & + \exp(-\psi/\psi_0)\frac{1}{2}\overline{g}^{\mu\nu}\partial_\mu\sigma\partial_\nu\sigma - \exp(-2\psi/\psi_0)V(\sigma)\Big]. \end{aligned} \tag{2.7}$$

We will generally denote quantities in the Einstein Frame by an overbar. With the same assumptions as before, the equation of motion in the new frame are

$$\begin{aligned} \left(\frac{\overline{a}'}{\overline{a}}\right)^2 + \frac{k}{\overline{a}^2} &= \frac{8\pi G_N}{3}\left[\frac{1}{2}\psi'^2 + \exp(-2\psi/\psi_0)\rho_V\right] \\ \psi'' + 3\frac{\overline{a}'}{\overline{a}}\psi' &= \frac{2}{\psi_0}\exp(-2\psi/\psi_0)\rho_V, \end{aligned} \tag{2.8}$$

where prime denotes $d/d\bar{t}$. The solutions to the equations of motion are easily found to be

$$\bar{a}(\bar{t}) = \bar{a}(0)(1 + C\bar{t})^{3/4+\omega/2}, \qquad \psi(\bar{t}) = \psi(0) + \psi_0 \ln(1 + C\bar{t}), \tag{2.9}$$

with

$$C = \frac{1}{\alpha\Phi(0)} \left(\frac{\rho_V}{24\pi G_N}\right)^{1/2} = 2[16\pi G_N \Phi(0)]^{-1/2} B. \tag{2.10}$$

The times in the two frames are related by $(1 + C\bar{t}) = (1 + Bt)^2$.

In the Einstein frame, the solutions with $\sigma = \sigma_{FV}$ and $V(\sigma_{FV}) = \rho_V$ resemble slow-rollover inflation for $\psi$ with an exponential potential $V(\psi) = \rho_V \exp(-2\psi/\psi_0)$. This fact will be of great use when we consider density perturbations.

As advertised, a theory with variable $G$ can be expressed as a theory with constant $G$, but variable microphysics. It is necessary to specify two mass scales before one can address the question of the constancy of either. In the Jordan frame when we say that the gravitational constant changes, of course we mean is that it changes *relative* to some other mass scale, in this case $m_\sigma$. The same is true in the Einstein frame: Although $G$ is said to be constant, it nevertheless changes with respect to $m_\sigma$. All that has been done is that the time dependence has been shifted to the $\sigma$ sector. In reality, there is no observation that can be made that would distinguish between the two frames. It is impossible to distinguish an increase in the gravitational constant from a decrease in the microphysics mass scale.

Soon after the original work it was realized that although it is clear that percolation eventually occurs, the model was not without its own problems. The problem comes about chiefly because of the existence of bubbles nucleated early in the epoch of inflation that grow to become large at the end of inflation.[22]

To illustrate the big-bubble picture, I follow Weinberg's analysis.[22] First, consider the expression for the co-moving size of a bubble in extended inflation [cf. Eq. 1.4)]:

$$r(t, t_0) = \frac{1}{Ba(0)} \left[\frac{1}{[1 + Bt_0]^{\omega-1/2}} - \frac{1}{[1 + Bt]^{\omega-1/2}}\right] \frac{1}{\omega - 1/2}, \tag{2.11}$$

for a bubble at time $t$ that was nucleated at time $t_0$. Now we would like to calculate the volume fraction contained in bubbles greater than some comoving radius, say $r$. This volume fraction will be denoted by $\mathcal{V}_>(r, t_{\rm END})$. If the bubble had comoving size $r$ at $t_{\rm END}$, then the time the bubble was nucleated, denoted by $t_*$, can be found by use of Eq. 2.11 with the substitutions $t \to t_{\rm END}$ and $t_0 \to t_*$. Now bubbles nucleated before $t_*$ will have a comoving size greater than $r$. If $p(t_*) = \exp[-\int_0^{t_*} dt'\, \Gamma V(t_*, t')]$ is the fraction of the Universe in false vacuum at $t_*$, then $\mathcal{V}_>(r, t_{\rm END})$ is given by

$$\mathcal{V}_>(r, t_{\rm END}) = 1 - \exp\left[-\int_0^{t_*} dt'\, \Gamma a^3(t') \frac{4\pi}{3} r^3(t_*, t')\right]. \tag{2.12}$$

This can be easily calculated. It can be shown that for sufficiently large bubbles, $\mathcal{V}_{>}(r, t_{\rm END}) \propto r^{4/\omega}$, or very nearly a scale-invariant distribution. Of course this scaling eventually breaks down, but not until enormous $r$.

Now after inflation if finished, there will be a distribution of different bubble sizes. The "small" bubbles will quickly have their interiors filled by the entropy created in bubble wall collisions. However it takes a finite time for the radiation to fill the vacuum bubble, namely a time $t = R$, where $R$ is the physical size of the bubble. The physical size of the bubble at time $t$ and temperature $T$ after inflation is simply $R(t) = ra(t) = ra(t_{\rm END})T_{\rm END}/T$. Clearly bubbles of size $R(t)$ greater than the horizon, $d_H(t) \sim m_{Pl}/T^2$, could not have thermalized. Therefore Eq. 2.12 can be used to find the volume fraction of the Universe that remains non-thermalized after inflation.

This fraction is shown in Fig. 2 for two values of $\omega$. Clearly the fraction grows with $\omega$. This is not surprising, since in the $\omega \rightarrow \infty$ limit the theory becomes old inflation, which we know does not thermalize the bubbles. The serious problem is that the limits seem to require that $\omega \lesssim 20$, which is in serious disagreement with the solar system $\omega \lesssim 500$ constraints.

So the simple, original model of extended inflation fails. However, like old inflation, it is a very interesting failure. After all, no one really likes JBD with $\omega \gg 500$ as a fundamental theory. However, there are many purportedly fundamental theories that have some of the features of JBD, including, supergravity, Kaluza-Klein, induced gravity, dilatons, and superstrings. In the Section IV I will describe some of the efforts to connect a successful first-order inflation theory to a desirable particle physics (or gravity) model.

However, before going on, in the next section I will discuss some interesting questions regarding the assumption that $\Gamma$ is independent of $t$.

## III. BUBBLE NUCLEATION IN FIRST-ORDER INFLATION

In this section I discuss some fundamental problems that arise in the calculation of the bubble nucleation rate in first-order inflation, and review some recent work. (Some of the work may perhaps even signify progress.)

The fundamental question regards the time dependence of the bubble nucleation rate $\Gamma$. A related question concerns the gravitational corrections to the nucleation rate. First of all, that a potential time dependence to the nucleation rate is potentially important is obvious: after all, the crucial parameter is $\epsilon = \Gamma/H^4$. Why would one expect a time dependence to $\Gamma$? I will illustrate why in two ways. I will also use two ways to illustrate why the present formalisms cannot be used to calculate the decay rate.

First, consider the theory in the Jordan frame, Eq. 2.1. The relevant question here concerns the gravitational corrections to the nucleation rate. Recall the Coleman-De

Luccia result for nucleation in a de Sitter background. If in flat space the bounce action is denoted as $B_0$ (c.f. Eq. 1.3) and the critical bubble size in the absence of gravity as $R_C$, then in the thin-wall limit the gravitational corrections on vacuum decay result in an effective bounce action of[6]

$$B = \frac{B_0}{[1 + (2R_C H)^2]^2}. \tag{3.1}$$

In old inflation the expansion rate $H = (8\pi G_N/3)^{1/2}$ is constant so the gravitational correction is a constant. In the case of extended inflation, $G$ and $H$ change in time. Since $R_C$ is set by microphysics it should be constant. So one would expect the gravitational corrections to the tunnelling rate to also change in time. Equation 3.1 suggests that as $H$ decreases in time, $B$ will grow. Since $\Gamma = A\exp(-B)$, as $B$ grows, $\Gamma$ decreases. Therefore at very early times $B$ should have been larger and bubble nucleation should have been suppressed. This is exactly what we want to get rid of troubling large bubbles nucleated early in inflation. What we have done to arrive at this result is naively to replace $G_N$ by $G(t)$. The reason this is not strictly correct is that gravity consists of more than usual because the JBD field is also part of gravity, and its dynamics must be taken into account. The problem is not so simple as a the replacement $G_N \to 16\pi G = \Phi^{-1}$ because of the effect of the kinetic term for $\Phi$ that induces an effective $\dot{\Phi}$ (or $\dot{G}$).

Now consider the theory as expressed in the Einstein frame, Eq. 2.7. Here the Ricci term in $\overline{S}$ is normal (the statement that gravity is normal is too strong—one can't forget the influence of the JBD field $\psi$), but the inflaton sector is funny because $\psi$ is mixed into both the kinetic and potential terms for the inflaton and will effect the inflaton's dynamics. A visualization of what happens is shown in Fig. 3. For the moment, let's ignore the fact that $\psi$ enters the kinetic term for $\sigma$. Then we might identify $\exp(-2\psi/\psi_0)V(\sigma) = V(\sigma,\psi)$ as the "effective" potential and use it to calculate the tunnelling. (Of course the fact that the kinetic term for $\sigma$ is not canonical makes identification of an effective potential for $\sigma$ risky indeed.) The effective potential depends upon both $\psi$ and $\sigma$. The classical evolution of the system starts from the origin and rolls in the $\psi$ direction. At some point in the evolution the field tunnels in the $\sigma$ direction, but not necessarily orthogonal to the $\psi$ direction. See Fig. 3.

The generalization of the fate of the false vacuum formalism to two fields would be straightforward but for the fact that there is no minimum along the $\psi$ axis. The generalization of the "most probable escape path" of Banks, Bender, and Wu[23] will not apply in this case since there is no false vacuum minimum.

It is easy to see why this presents a problem with the usual boundary conditions imposed on the bounce. For simplicity, first consider the flat-space solution. The Euclidean least-action solution has $O(4)$ symmetry, and is a function of $\rho = (|\vec{x}|^2 + t^2)^{1/2}$. The usual bounce boundary conditions are $\sigma(\infty) = \sigma'(\infty) = 0$, and $d\sigma(0)/d\rho = 0$ (here prime refers to $d/d\rho$). Recall that upon analytic continuation to Minkowski space, $\rho = \infty$ corresponds to the false-vacuum region. For the problem of interest, the Euclidean equations of motion do not admit a solution with the above boundary conditions at $\rho = \infty$. This of course traces to the fact that there is no "static" false-vacuum state. The problem also is present in the usual Coleman-DeLuccia formulation for the gravitational corrections. Clearly the solution must take a different form because the Euclidean action contains a term proportional to $(3a'_E/a_E)\psi'$ (here $a_E$ is the Euclideanized version of the scale factor). The usual solution contains two zeros of $a_E$. The Euclidean action for the usual case remains finite due to the boundary conditions that $\psi' = 0$ at the zeros of $a_E$. However in the Euclideanized version of the JBD theory there is only one zero of $\psi'$! Again we face a problem. Of course the two problems are not unrelated.

It is suspected that the Euclidean formulation of the tunnelling problem is not appropriate for this problem because in fact in the false vacuum state there must be a "real" momentum in the $\psi$ direction. A formulation of fate of the false vacuum appropriate for the rolling and tunnelling problem would be most welcome!

Although the issue is far from settled, some progress has been made.[24,25,26] Here I will discuss the work I have done in collaboration with Rich Holman, Sharon Vadas, Yun Wang, and Erick Weinberg.[25,26] Rather than limit the discussion to the JBD theory, let me start with a slightly more general action, one that will in fact arise in some of the models considered in the next section.

Let us start by considering theories of a JBD field $\Phi$ coupled to an inflaton field $\sigma$ via the following action:

$$S = \int d^4x \sqrt{-g} \times \left[ -\Phi R + \omega g^{\mu\nu} \frac{\partial_\mu \Phi \partial_\nu \Phi}{\Phi} + F(\Phi) \frac{1}{2} g^{\mu\nu} \partial_\mu \sigma \partial_\nu \sigma - G(\Phi) V(\sigma) \right]. \tag{3.2}$$

The simplest coupling functions can be of the form

$$F(\Phi) = (16\pi G_N \Phi)^n, \qquad G(\Phi) = (16\pi G_N \Phi)^m. \tag{3.3}$$

To ensure that our theory reduces to general relativity in the appropriate limit, we require that for $\Phi = \Phi_0 = 1/16\pi G_N$, $F(\Phi_0) = G(\Phi_0) = 1$.

I will elaborate on the method developed by us[25] which allows us to systematically "freeze out" gravitational effects in the bounce, thus enabling us to arrive at approximate expressions for the nucleation rate which reflect the time evolution of the JBD field $\Phi$.

To implement our approximation, we go to the Einstein Conformal Frame. The reason comes from the observation that in the Jordan conformal frame action (i.e., Eq. 3.2), the second term is not the complete kinetic term for $\Phi$, since an integration by parts of the first term will make a contribution to the $\Phi$ kinetic term. Therefore, for semi-classical calculations involving the JBD field, it is more appropriate (and often easier) to use the Einstein Conformal Frame. Then we may transform back to the Jordan frame if we choose. The action of Eq. 3.2 expressed in the Einstein frame becomes

$$\overline{S} = \int d^4x\, \sqrt{-\overline{g}}\, \Big[ -\frac{\overline{\mathcal{R}}}{16\pi G_N} + \frac{1}{2}\overline{g}^{\,\mu\nu}\partial_\mu\psi\partial_\nu\psi$$
$$+ f(\psi/\psi_0)\,\frac{1}{2}\overline{g}^{\,\mu\nu}\partial_\mu\sigma\partial_\nu\sigma - g(\psi/\psi_0)V(\sigma)\Big]. \tag{3.4}$$

where

$$f(\psi/\psi_0) \equiv \exp(-\psi/\psi_0)\, F(\Phi), \qquad g(\psi/\psi_0) \equiv \exp(-2\psi/\psi_0)\, G(\Phi), \tag{3.5}$$

and $\Phi = \Phi(\psi/\psi_0)$ is understood. For the simple couplings of Eq. 3.3, $f(\psi) = \exp[(n-1)\psi/\psi_0]$ and $g(\psi) = \exp[(m-2)\psi/\psi_0]$.

Now we wish systematically to freeze out gravitational effects. The action written as above in the Einstein frame reveals that if we want to freeze out gravitational effects, we must also freeze out the evolution of the $\psi$ field during the bounce. This is due to the fact that we are taking the $G_N \to 0$ limit and $\psi_0 \propto {G_N}^{-1/2}$, thus the second term has the same $G_N$ dependence as the first term. Treating $\psi$ as constant also implies that $\Phi$ must be taken to be constant. This may be alarming since we have to use explicitly the fact that both the scale factor and $\Phi$ are time dependent to make our discussion relevant for inflation. However, we are saved by the observation that the (imaginary-time) bounce configuration used in computing the tunneling action is distinct from the (real-time) background metric and JBD field configuration governing the evolution of the Universe. Thus the latter can remain time *dependent* while we freeze out the time evolution in $\psi$ during the tunneling process.[25]

Corrections to this approximation can also be considered. We expect, using the results of bubble nucleation calculations in standard gravity as a guide, that our approximation will be reliable when the effective Planck mass induced by the JBD field is much greater than the mass scales associated with the $\sigma$ field. In theories where $\Phi$ increases with time, the approximation will work best at late times.

The approximation discussed above yields the following truncated action for the inflaton $\sigma$ in the Euclidean frame:

$$\overline{S}_E = \int d^4x\, \Big[ f(\xi)\frac{1}{2}\partial^\mu\sigma\partial_\mu\sigma + g(\xi)V(\sigma)\Big], \tag{3.6}$$

where $\xi \equiv \exp(\psi/\psi_0)$.

To calculate the bubble nucleation rate (per unit physical three volume)

$$\overline{\Gamma} = A\exp(-B), \tag{3.7}$$

we need to calculate the bounce action, $B$, and the prefactor, $A$. If we rescale the coordinates to

$$\tilde{x}^\alpha = \sqrt{\frac{g(\xi)}{f(\xi)}}\, x^\alpha \tag{3.8}$$

we can rewrite the action of Eq. 3.6 as

$$\overline{S}_E = \frac{f^2(\xi)}{g(\xi)} \int d^4\hat{x} \left[\frac{1}{2}\hat{\partial}^\mu \sigma \hat{\partial}_\mu \sigma + V(\sigma)\right] = \frac{f^2(\xi)}{g(\xi)}\, S_0 \tag{3.9}$$

where $S_0$ is the Euclidean action of the standard theory (i.e., the action of Eq. 3.4 with $\xi = 1$). Clearly, this implies that the bounce configuration $\sigma_B$ is related to the bounce of the theory containing $\hat{\sigma}_B$:

$$\sigma_B(x) = \hat{\sigma}_B(\sqrt{g(\xi)/f(\xi)}\, x). \tag{3.10}$$

The bounce action is

$$B(\xi) = \frac{f^2(\xi)}{g(\xi)} B_0 \qquad [\xi = \exp(\psi/\psi_0)], \tag{3.11}$$

where $B_0$ is the ($\xi$-independent) bounce action calculated for the theory with $\xi = 1$ ($\psi = 0$). The fact that the coupling of $\psi$ into the action of Eq. 3.6 can be factored out by means of coordinate rescaling is essential in enabling us to carry out our calculation.

The prefactor $A$ from Eq. 3.7 is given by[8]

$$A = \left|\frac{\det'[S''_E(\sigma_B)]}{\det[S''_E(\sigma_{FV})]}\right|^{-1/2} \prod_\mu \left(\frac{C_\mu}{2\pi}\right)^{1/2}. \tag{3.12}$$

Here, $\sigma_{FV}$ is the false-vacuum configuration, $\sigma_B$ is the bounce solution, and det$'$ indicates that the functional determinant is to be evaluated in the subspace orthogonal to the four translational zero modes. The $C_\mu$ are normalization factors of the zero modes of the operator $S''_E(\sigma_B)$.

Performing the functional variation of the Euclidean action yields

$$\begin{aligned} A_{DET} &\equiv \left|\frac{\det'[S''_E(\sigma_B)]}{\det[S''_E(\sigma_{FV})]}\right|^{-1/2} \\ &= \left|\frac{\det'[-f(\xi)\partial^2 + g(\xi)V''(\sigma_B)]}{\det[-f(\xi)\partial^2 + g(\xi)V''(\sigma_{FV})]}\right|^{-1/2}. \end{aligned} \tag{3.13}$$

To determine the $\xi$ dependence of the above expression, we observe that if $\Psi_\theta(x)$ is the eigenfunction of the operator $-\hat{\partial}^2 + V''(\hat{\sigma})$ with eigenvalue $\theta$, then

$$\begin{aligned} [-f(\xi)\partial^2 + g(\xi)V''(\sigma)]\Psi_\theta(\sqrt{g(\xi)/f(\xi)}\, x) &= g(\xi)[-\hat{\partial}^2 + V''(\sigma)]\Psi_\theta(\hat{x}) \\ &= g(\xi)\,\theta\,\Psi_\theta(\sqrt{g(\xi)/f(\xi)}\, x), \end{aligned} \tag{3.14}$$

i.e., $\Psi_\theta(\sqrt{g(\xi)/f(\xi)}\, x)$ is the eigenfunction of the operator $-f(\xi)\partial^2 + g(\xi)V''(\sigma_B)$ with eigenvalue $g(\xi)\theta$. Since the primed determinant has four eigenvalues fewer than the unprimed one, we have

$$A_{DET} = \{[g(\xi)]^{-4}\}^{-1/2}\, \hat{A}_{DET} = g^2(\xi)\hat{A}_{DET}. \tag{3.15}$$

The $C_\mu$ are defined so that the properly normalized modes are $C_\mu^{-1/2}\partial_\mu\sigma_B$ ($\mu = 1,\cdots,4$). Thus, $C_\mu = \int d^4x(\partial_\mu\sigma_B)^2$ (no sum over $\mu$ implied), and for an $O(4)$-symmetric bounce, the $C_\mu$ are all equal. The $\xi$ dependence of $C_\mu$ can easily be found:

$$C_\mu = \int d^4x(\partial_\mu\sigma_B)^2 = f(\xi)/g(\xi)\,\widehat{C}_\mu. \tag{3.16}$$

Hence, the nucleation rate in the Einstein frame is

$$\overline{\Gamma}(\bar{t}) = f^2(\xi)\widehat{A}\exp(-B_0 f^2(\xi)/g(\xi)). \tag{3.17}$$

Now we may find the nucleation rate in the Jordan frame. Recall that $\exp(\psi/\psi_0) = 16\pi G_N\Phi$, and that the nucleation rate in the Jordan frame is related to that in the Einstein frame by

$$\begin{aligned}\Gamma &\equiv \frac{dP}{d^4x\sqrt{-g}} = \frac{dP}{d^4x\sqrt{-\bar{g}}}\frac{\sqrt{-\bar{g}}}{\sqrt{-g}} = \Omega^{-4}\overline{\Gamma} \\ &= \xi^2 f^2(\xi)\widehat{A}\exp(-B_0 f^2(\xi)/g(\xi)) \\ &= \widehat{A}F^2(\Phi)\exp\left\{-B_0F^2(\Phi)/G(\Phi)\right\}.\end{aligned} \tag{3.18}$$

$\widehat{A}$ and $B_0$ are $\Phi$ independent and depend only upon the inflaton potential. $B_0$ is dimensionless, while $\widehat{A}$ has mass dimension 4.

For the simple power-law coupling functions in Eq. 3.3, we have

$$\Gamma = \widehat{A}(16\pi G_N\Phi)^{2n}\exp[-B_0(16\pi G_N\Phi)^{2n-m}]. \tag{3.19}$$

In the original extended inflation model $m = n = 0$, and in the Jordan frame the nucleation rate is time independent, although it is time dependent in the Einstein frame (as discussed in Ref. 25). However, in dimensionally reduced theories, the generic form has $m$ and $n$ different from zero.[27] We see from the above equation that if $2n - m \neq 0$ the time dependence of the nucleation probability can be *exponentially* strong through the time dependence of $\Phi$ (or equivalently, $\psi$). If $2n - m = 0$ but $n \neq 0$, the nucleation probability is still time dependent in the Jordan frame, and time dependent in the Einstein frame if $n \neq 1$.

For arbitrary functions $F(\Phi)$ and $G(\Phi)$, we can expect much richer time dependence of the bubble nucleation rate.

To conclude, in generalized extended inflation theories, the bubble nucleation rate acquires an explicit time dependence, even in the limit of freezing out gravitational effects. The time dependence will be *exponentially* strong in the generic case. This remarkable feature of the theories encompassed by our model provides optimistic prospects for the success of percolation, since the time dependence of the percolation parameter $\epsilon$ is enriched through the time dependence of the nucleation probability.

For more details on these calculations, the reader should consult the original papers that are highlighted here. The calculation including the fact that the JBD field is evolving while the inflaton is tunnelling remains an outstanding problem facing extended inflation.

# IV. MODEL BUILDING

Although the original extended inflation model failed, it is a most promising failure. Were it not for the fact that the large-bubble problem required $\omega$ to be much smaller than allowed by observations, extended inflation would be a viable, rich, and interesting model for inflation.

In the eighteen months since the extended inflation model there have been several attempts to construct models that suppress bubble nucleation at early times, yet allow for the inflation to be terminated by bubble collisions. In this section I will review some of the attempts.

Before embarking upon the excursion into model building, it is important to recall again exactly why extended inflation failed. The basic fault was the nucleation of bubbles early in inflation. This could only be solved if $\omega \lesssim 20$, which is at variance with the solar system limit $\omega \gtrsim 500$. The fact that JBD failed should not necessarily be taken as bad. After all, no one really expects JBD to be the fundamental theory. In particular any scalar theory with $\omega \gtrsim 500$ would be called "unnatural."

Somehow a model must be found that can have a small $\omega$ and not be at odds with observation. It is useful to have a rough idea as to why there is the limit $\omega \gtrsim 500$. From Eq. 2.1 it is clear that in the limit that $\omega \to \infty$ the kinetic term for $\Phi$ decouples and $\Phi$ will be constant. So long as $\Phi$ remains constant, we could identify it as $(16\pi G_N)^{-1}$ and the theory would be indistinguishable from GR. This essentially is the reason for the limit $\omega \gtrsim 500$. However there is another way to keep $\Phi$ constant, by giving it a potential to anchor it. The simplest potential one might imagine is a mass for $\Phi$. Most of the extended inflation models are a variation on the theme of JBD theories augmented by a potential for $\Phi$.

It is also important to recall that the motivation for extended inflation was the "small number" problem of slow-rollover inflation, usually characterized by some dimensionless coupling constant having a value less than about $10^{-15}$ or so. We should demand that any successful scenario we find should not be infected by the disease we are trying to cure.

### *Two-field inflation* [28]

Perhaps the simplest realization of first-order inflation is a model in which two scalar fields are coupled together. In this approach it is assumed that there are two scalar fields: one field rolling, and one field trapped in the false vacuum. In keeping with previous notation, I will denote the trapped field as $\sigma$, the inflaton field, and the rolling field as $\psi$. The fundamental difference between this approach and extended inflation is that the kinetic terms for $\sigma$ and $\psi$, as well as gravity, are assumed to be canonical. This model is truly a modification of particle physics, not gravity. This model fits into the first-order inflation class, but not into the extended inflation subclass.

Two-field inflation suffers from the general problem of fine tuning to keep the potential for the rolling field flat. This is similar to the usual fine tuning problem of extended inflation.

Here I illustrate two-field inflation with simple model. First, consider a toy potential for the inflaton field $\sigma$ that will involve a first-order phase transition:

$$V(\sigma) = \frac{\lambda}{4}\sigma^2(\sigma - \sigma_0)^2 - \frac{\lambda}{2}\epsilon\sigma_0\sigma^3 + \Lambda. \tag{4.1}$$

This is the form of the potential illustrated in Fig. 1a. There is a metastable minimum at $\sigma = 0$, and a true minimum at

$$\sigma_{TV} = \sigma_0 \left[3(1+\epsilon) + \sqrt{9(1+\epsilon)2 - 8}\right]/4. \tag{4.2}$$

The constant $\Lambda$ is added to make $V(\sigma_{TV}) = 0$. The nucleation rate of bubbles of true vacuum in the thin-wall approximation is $\Gamma \simeq \sigma_0^4 \exp(-\pi^2/48\lambda\epsilon^3)$.

Now suppose we consider a potential involving both the inflaton field $\sigma$ and a second scalar field, the slow-roll field $\psi$, of the form

$$V(\sigma,\psi) = \frac{\lambda}{4}\sigma^2(\sigma - \sigma_0)^2 - \frac{\lambda}{2}\psi\sigma^3 + \Lambda + V(\psi), \tag{4.3}$$

where $V(\psi)$ is a potential with a minimum at some large value of $\psi$, say $\psi_{TV}$. If we imagine the slow-roll field to be constant, then we recover the original potential with $\epsilon\sigma_0 = \psi$.

In two field inflation the slow-roll field does what it should, slowly rolls. As it does, the bounce action proportional to $\epsilon^{-3}$ will decrease since the effective value of $\epsilon$ is proportional to $\psi$ which is growing. Therefore the nucleation rate $\Gamma \propto \exp(-1/\epsilon^3) \propto \exp(-1/\psi^3)$ will grow as $\psi$ rolls, eventually it is hoped, to grow large enough to trigger percolation of true-vacuum bubbles.

This procedure can be made to work, but at the expense of making the potential $V(\psi)$ flat enough for slow roll, which again requires uncomfortably small dimensionless coupling constants. One also must worry that after extended inflation there will be a period of slow-roll inflation triggered by $V(\psi)$. A final worry is to prevent $V(\psi)$ from dominating the total potential during extended inflation. It seems that the only advantage to the complication of this model compared to rollover inflation is that there are interesting phenomena associated with first-order transitions that will be discussed in the next section.

For more details on two-field models, see Refs. 28.

### *Induced gravity* [30]

The original model of extended inflation failed because the requirement that the production of large bubbles not disturb the isotropy of the Universe led to the constraint $\omega \lesssim 20$, which is at odds with observational constraints $\omega \gtrsim 500$. However the observational constraint holds only in "pure" Brans–Dicke theories, and if the JBD field has a small mass, the constraint will not apply. It is easy to see why: In the limit that $\Phi = \text{const}$, Brans–Dicke is equivalent to Einstein gravity. The coefficient of the $\Phi$ kinetic term is proportional to $\omega$. As $\omega$ becomes large, the kinetic term decouples from the low-energy theory and $\Phi$ becomes constant. In this way large-$\omega$ Brans–Dicke theories are indistinguishable from Einstein. However, there is another way to avoid the observational constraint; to give the JBD field a mass. If there is some potential keeping $\Phi$ anchored at some value (namely $(16\pi G_N)^{-1}$, then the low-energy limit of Brans–Dicke will again resemble Einstein gravity. Models with a potential for $\Phi$ are called "induced gravity" models.

As in the two-field model just discussed, the parameters in any induced extended inflation model must satisfy several constraints. Let's construct a toy model that is Eq. 2.1 with a potential $V(\Phi) = \lambda(\Phi - \Phi_0)^2$, where $\Phi_0 = m_{Pl}^2/16\pi$. The potential enters the equations of motion. For instance the $\ddot{\Phi}$ equation becomes

$$\ddot{\Phi} + 3\frac{\dot{a}}{a}\dot{\Phi} = \frac{2}{3+2\omega}\left[\rho_V + V(\Phi) - \frac{1}{2}\Phi\frac{\partial V(\Phi)}{\partial \Phi}\right]. \tag{4.4}$$

In order for $V(\Phi)$ not to disturb the evolution of the Universe during extended inflation, we must have

$$\frac{1}{2}\Phi\frac{\partial V}{\partial \Phi} \ll \rho_V; \qquad V(\Phi) \ll \rho_V. \tag{4.5}$$

Both constraints will be satisfied for $m_\Phi < M^2/m_{Pl}$, where $m_\Phi^2 = \lambda\Phi_0 \sim \lambda m_{Pl}^2$, and $M \sim \rho_V^{1/4}$ is the GUT mass scale. For $M \sim 10^{14}\text{GeV}$, this implies $m_\Phi \lesssim 10^9\text{GeV}$. This is smaller than the natural mass of $m_{Pl}$. This is evidence for a small $\lambda$. For more details, consult the paper of Accetta and Trester.[30]

### *Hyperextended inflation* [31]

Steinhardt and Accetta have proposed an extended inflation model that they call "hyperextended" inflation. In their model they replace the $-\Phi\mathcal{R}$ term in the Brans–Dicke action by a more general term $-f(\Phi)\mathcal{R}$, where $f(\Phi) = M^2 + \Phi + \beta\Phi^2/M^2 + \cdots$. One might easily imagine a scenario whereby various terms in the gravitational action dominate during different epochs: (a) $\Phi \lesssim M^2$ where the first term dominates; (b) $M^2 \lesssim \Phi \lesssim M^2/\beta$ where the second term dominates, (c) $M^2/\beta \lesssim \Phi$ where the third term dominates, and so on. This is exactly the picture proposed by Steinhardt and Accetta.

It is easy to see why hyperextended inflation works. If $f(\Phi)$ is dominated by one of the terms in the series, it is possible to define a new JBD field $\bar{\Phi} \equiv f(\Phi)$ and the standard JBD action with $\bar{\Phi}$ as the scalar field will be recovered with $\omega = f(\Phi)/2f'(\Phi)$, where $f'(\Phi) = df(\Phi)/d\Phi$. The equations of motion in the hyperextended model are the same as in the original extended inflation model with the addition of a term proportional to $\omega' = d\omega/d\Phi$:

$$\ddot{\bar{\Phi}} + 3\frac{\dot{a}}{a}\dot{\bar{\Phi}} = \frac{1}{3+2\omega}\left[\frac{\rho - 3p}{2} - \omega'\dot{\bar{\Phi}}^2\right]. \tag{4.6}$$

Of course this equation reduces to Eq. 2.2 in the limit that $\omega' = 0$.

In the hyperextended picture extended inflation occurs during regime (b), where $\omega'$ is small and does not change, the usual extended inflation equations of motion obtain. Once $\Phi$ becomes large enough to enter into regime (c), the expansion rate is drastically reduced, and extended inflation ends as $H(t)$ rapidly decreases to a value that will give $\epsilon(t) \gtrsim \mathcal{O}(1)$. Hence in hyperextended inflation, the end of inflation is determined by the parameters of the potential that determine the value of $\Phi$ at the start of regime (c). The constraint that not too many large bubbles are produced again amounts to $\omega_{(b)} \lesssim 20$, where $\omega_{(b)}$ is the effective value of $\omega$ during regime (b).

Now what about after extended inflation? From the above equation for $\ddot{\bar{\Phi}}$ and $\dot{\bar{\Phi}}$ will be constant in the radiation-dominated era. Once the matter-dominated era is reached $\Phi$ will continue to evolve. Steinhardt and Accetta escape the $\omega \lesssim 500$ problem in a clever way. Note that $\omega \propto 1/f'(\Phi)$. If the function $f(\Phi)$ has a maximum where $f'(\Phi) = 0$, $\Phi$ could evolve to that point, the effective value of $\omega$ will diverge decoupling the $\Phi$ kinetic term from the action, and prevent further evolution of $\Phi$. This will effectively ensure that the solar system tests could not rule out the value of the parameters necessary to solve the large-bubble problem.

For more details on this interesting model, see Ref. 31.

### *Overextended inflation* [32]

Holman, Kolb, and Wang[32] showed that problems with extended inflation can be avoided in a new class of models considered by Damour, Gibbons, and Gundlach (DGG).[33] They start with a generalized JBD model in which the JBD scalar field $\Phi$ couples with different strengths to "visible" matter and to "invisible" matter (thus leading to a violation of the weak equivalence principle). Following the line taken by DGG, assume that the inflaton of extended inflation has an "invisible" coupling to the JBD scalar field. Since the identity of the inflaton is unknown, there is no reason to believe that it should couple to the JBD scalar field in the same way as does normal matter.

The procedure is to start with the usual JBD action written in the Jordan frame, and modify the couplings of the JBD field to the inflaton:

$$\begin{aligned} S_I[g_{\mu\nu}, \Phi, \sigma_I] &= \int d^4x \sqrt{-g}\,\Big[\frac{1}{2}(16\pi G_N \Phi)^{1-\beta} g^{\mu\nu}\partial_\mu\sigma_I\partial_\nu\sigma_I \\ &\qquad -(16\pi G_N\Phi)^{2(1-\beta)}V(\sigma_I)\Big], \end{aligned} \tag{4.7}$$

where the subscript $I$ denotes the inflaton sector. Denoting the field content of visible matter by $\sigma_V$, in the Jordan Frame the action for visible matter, $S_V[g_{\mu\nu}, \sigma_V]$, would simply be the action for a minimally coupled field, e.g., Eq. (4.7) with $\sigma_I \to \sigma_V$ and $\beta = 1$. Expressing the action in the Jordan Frame, the parameters of the model are

$\beta$ and $\omega$. Further suppose that visible matter is described as usual via a perfect fluid stress-energy tensor and will play no role in inflation.

Assume for "extended" inflation $\sigma_I = \sigma_0 =$ const, and $V(\sigma_0) = \rho_V$, where the energy density of the false vacuum, $\rho_V$, dominates the total energy density (the subscript $V$ on $\rho_V$ refers to "vacuum" and not "visible"). Setting $\Lambda = 8\pi G_N \rho_V$, the equations of motion for $a(t)$ and $\Phi(t)$ are:

$$
\begin{aligned}
\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} &= \frac{\Lambda}{3}(16\pi G_N \Phi)^{1-2\beta} + \frac{\omega}{6}\left(\frac{\dot{\Phi}}{\Phi}\right)^2 - \frac{\dot{a}}{a}\frac{\dot{\Phi}}{\Phi} \\
\frac{\ddot{\Phi}}{\Phi} + 3\frac{\dot{a}}{a}\frac{\dot{\Phi}}{\Phi} &= \frac{4\beta\Lambda}{2\omega+3}(16\pi G_N \Phi)^{1-2\beta},
\end{aligned}
\tag{4.8}
$$

These equations are most easily solved in terms of the dimensionless field $\chi \equiv 16\pi G_N \Phi$. This system of equations admits power law solutions for $k = 0$, just as in the original extended inflation scenario:

$$
\begin{aligned}
a(t) &= a(0)(1 - Bt)^p; \qquad p = (\omega - \beta + 3/2)/[(2\beta - 1)\beta] \\
\chi(t) &= \chi(0)(1 + Bt)^q; \qquad q = 2/(2\beta - 1).
\end{aligned}
\tag{4.9}
$$

Here $B$ is given by

$$
B^2 = \frac{4\Lambda\beta^2(2\beta-1)^2[\chi(0)]^{1-2\beta}}{(2\omega+3)(6\omega+9-4\beta^2)}.
\tag{4.10}
$$

It can easily be seen that the above results reduce to the usual extended inflation results when $\beta = 1$.

In order for sufficient inflation to occur, $a(t_{\rm END})/a(0) > 10^{27}$, where $t_{\rm END}$ denotes the end of the inflationary period. Following the usual analysis, it is possible to relate $a(t_{\rm END})/a(0)$ to $\Phi(t_{\rm end})/\Phi(0)$ via $a(t_{\rm END})/a(0) = [\Phi(t_{\rm END})/\Phi(0)]^{p/q}$.

Next, consider constraints coming from percolation and thermalization of the phase transition. From the analysis of Section III,

$$
\Gamma(t) = \Gamma_0\, \chi^{2(1-\beta)} = \Gamma_0\, [\chi(0)]^{2(1-\beta)}(1 + Bt)^{4(1-\beta)/(2\beta-1)},
\tag{4.11}
$$

where $\Gamma_0$ is the (constant) tunnelling rate for $\chi = 1$. Thus, as expected, the physical bubble nucleation rate per unit four-volume is *time dependent* in this theory.

As usual, the parameter controlling the percolation properties of the phase transition is $\epsilon \equiv \lambda(t)/H^4(t)$. Here,

$$
\epsilon(t) = \frac{\lambda_0 \chi(0)^{2(1-\beta)}}{p^4 B^4}(1 + Bt)^{4\beta/(2\beta-1)}.
\tag{4.12}
$$

Now turn to the constraint coming from the requirement that the bubble clusters that will comprise the observable universe have enough time to thermalize their energy. Imposing the condition that $\mathcal{V}_>(r, t_{\rm END})$ be less than $10^{-n}$ when the temperature is $T$, the constraint is:

$$\omega + 3/2 < \left\{ 2 + \frac{4[23 + \log_{10}(M/10^{14}\text{GeV}) + \log_{10}(\text{eV}/T)]}{n + \log_{10}\{\ln[p^{-1}(t_{\text{END}})]\}} \right\} \beta^2, \tag{4.13}$$

It is not unreasonable to suppose that $p(t_{\text{END}}) < 1/e$, so that for $M \sim 10^{14}$ GeV and $n \simeq 5$, at recombination ($T \simeq 1/3$ eV), the constraint is $\omega + 3/2 < 20.7\,\beta^2$. Note that setting $\beta = 1$ recovers the standard results. The difference is that whereas before the limit was $\omega \lesssim 20$, the limit now is $\omega/\beta^2 \lesssim 20$, which can easily be satisfied for $\omega > 500$.

The constraints on $(\omega, \beta)$ from sufficient inflation and thermalization of bubbles can be easily satisfied for a wide range of parameters without fine tuning.

### *Kaluza-Klein inflation* [27]

One possibility for successful extended inflation might be multidimensional theories such as Kaluza–Klein[36] theories. After all, the major motivation for the renewed interest in scalar-tensor theories such as Jordan–Brans–Dicke is that an effective low-energy theory of the Jordan–Brans–Dicke form follows naturally in superstring, supergravity, and Kaluza–Klein theories.

Upon reduction to four dimensions, theories originally formulated in higher dimensions take on a Jordan–Brans–Dicke form, with a function of the scale factor of the internal dimensions, $b(t)$, acting as the Jordan–Brans–Dicke field. Thus, it is important to investigate whether these theories can lead to successful models for extended inflation.

Consider a model of higher dimensional gravity coupled to a scalar field $\tilde{\chi}$ with a potential $\tilde{U}(\tilde{\chi})$ allowing for a metastable vacuum state as well as a completely stable one. The action for this theory can be written:

$$\tilde{S} = \int d^{4+D}x\, \sqrt{-\tilde{g}}\, \left[ -\frac{1}{16\pi\tilde{G}}\tilde{R} + \frac{1}{2}\tilde{g}^{MN}\partial_M\tilde{\chi}\partial_N\tilde{\chi} - \tilde{U}(\tilde{\chi}) \right]. \tag{4.14}$$

Here $D$ is the dimension of the internal space (which we take to be a $D$-sphere, $S^D$), and all the quantities with tildes refer to objects living in the full $(4+D)$-dimensional spacetime. We now assume that the spacetime line element $d\tilde{s}^2$ takes the form

$$d\tilde{s}^2 = dt^2 - a^2(t)d\Omega_3^2 - b^2(t)d\Omega_D^2 \tag{4.15}$$

where $d\Omega_3^2$ is the line element corresponding to a maximally symmetric 3-space and $d\Omega_D^2$ is that of a unit $D$-sphere. Denoting by $\tilde{\chi}_0$ the zero mode of $\tilde{\chi}$ (i.e., the part of the harmonic expansion of $\tilde{\chi}$ which is independent of the coordinates $\{y^\alpha\}$ of the $D$-sphere),[37] we can rewrite $\tilde{S}$ as

$$\tilde{S} = \left[ \int d^Dy\, \sqrt{\gamma(y)} \right] S \tag{4.16}$$

where $\gamma_{mn}(y)$ is the metric tensor of the $D$-sphere and $S$ is the effective four-dimensional action, given by:

$$S = \int d^4x \sqrt{-g(x)}\, \Omega_D b^D(t) \left[ -\frac{R}{16\pi\bar{G}} - \frac{D(D-1)}{16\pi\bar{G}} g^{\mu\nu} \frac{\partial_\mu b \partial_\nu b}{b^2} + \frac{\rho_D}{16\pi\bar{G} b^2} + \frac{1}{2} g^{\mu\nu} \partial_\mu \tilde{\chi}_0 \partial_\nu \tilde{\chi}_0 - \tilde{U}(\tilde{\chi}_0) \right]. \tag{4.17}$$

Here $\rho_D b^{-2}$ is the Ricci scalar of the internal space (i.e., constructed from $\gamma_{mn}(y)$ alone). Also $\{x^\alpha\}$ are the coordinates in the 4-dimensional space, and $g_{\mu\nu}(x)$ is the metric on this space. Note that the kinetic term for the $b$ field has the wrong sign. We will now rewrite $S$ using the following definitions:

$$\begin{aligned}
\Omega_D &\equiv \left[ \int d^D y \sqrt{\gamma(y)} \right] = 2\pi^{(D+1)/2}/\Gamma[(D+1)/2] \quad \text{for a } D \text{ sphere} \\
\frac{\Omega_D b_0^D}{16\pi\bar{G}} &\equiv \frac{1}{16\pi G_N} \\
\sigma &\equiv (\Omega_D b_0^D)^{1/2} \tilde{\chi}_0 \\
V(\sigma) &\equiv (\Omega_D b_0^D)\, \tilde{U}\left( \frac{\sigma}{(\Omega_D b_0^D)^{1/2}} \right),
\end{aligned} \tag{4.18}$$

where $G_N$ is the four-dimensional Newton's constant. Note that $\sigma$ is a canonical scalar field and $V(\sigma)$ is its 4-dimensional potential. Finally, defining

$$\Phi \equiv \frac{1}{16\pi G_N} \left( \frac{b}{b_0} \right)^D \tag{4.19}$$

to be the effective JBD field, we have

$$S = \int d^4x \sqrt{-g(x)} \left[ -\Phi R - \omega g^{\mu\nu} \frac{\partial_\mu \Phi \partial_\nu \Phi}{\Phi} + \alpha \Phi^{1-2/D} + (16\pi G_N \Phi) \left( \frac{1}{2} g^{\mu\nu} \partial_\mu \sigma \partial_\nu \sigma - V(\sigma) \right) \right], \tag{4.20}$$

with $\omega \equiv 1 - 1/D$ and $\alpha \equiv \rho_D (16\pi G_N)^{-2/D} b_0^{-2}$. Note that $\alpha$ has mass dimension $2(1+2/D)$. We have thus recast the Kaluza–Klein action into a JBD form as expressed in the Jordan frame.. There are, however, some important differences: (a) $\Phi$ has the "wrong" sign for a standard kinetic term, (b) there is a nontrivial self-interaction term for $\Phi$, namely $\alpha\Phi^{1-2/D}$, and (c) there are also $\Phi$–$\sigma$ cross terms.

We may now use the action of Eq. (4.20) to arrive at the Friedmann–Roberstson–Walker (FRW) equations for this system. Setting $\sigma = \sigma_{FV}$, its value in the false vacuum, and defining $V(\sigma_{FV}) \equiv \rho_V$ and $\Lambda \equiv 8\pi G_N \rho_V$, we have the following equations of motion in the flat space ($k = 0$) limit:

$$\begin{aligned}
\left( \frac{\dot{a}}{a} \right)^2 + \frac{\dot{a}}{a} \frac{\dot{\Phi}}{\Phi} &= -\frac{1}{6} \omega \left( \frac{\dot{\Phi}}{\Phi} \right)^2 - \frac{1}{6} \alpha \Phi^{-2/D} + \frac{\Lambda}{3} \\
\ddot{\Phi} + 3 \frac{\dot{a}}{a} \dot{\Phi} &= -\alpha \Phi^{1-2/D} + \frac{2\Lambda}{1 + 2/D} \Phi.
\end{aligned} \tag{4.21}$$

These equations differ from the extended inflation solutions because of the potential term for $\Phi$. In general, no power-law solutions exists. Numerical integration the the equations give the behavior of the JBD field rapidly being driven to zero before much inflation. Although this is not a promising model for inflation (extended or otherwise), we can learn some things to guide our thinking for the construction of a successful model. First, a potential for $\Phi$ of the form here will not work. There must be a long, flat region in the potential so enough e-folds of inflation may occur before the JBD field is driven to its minimum. Secondly, the potential is sick for small $\Phi$—there is nothing to prevent the extra dimensions from shrinking to zero—the minimum of $\Phi$ is at $\Phi = 0$. Both of these problems have been previously recognized and several solutions have been proposed.

The main problem with this higher-dimensional model in terms of their extended inflation properties is that they cannot be made to inflate *enough*! This is similar to the problem of induced gravity and two-field inflation. Here the potential for the JBD field was set by the curvature of the internal space. There is no way to naturally keep that potential flat enough for sufficient inflation.

For more details the reader should consult Ref. (27).

## V. THE INFLATON SECTOR

It is clear from the discussion in the previous section that it is possible to construct models that satisfy all of the necessary constraints. It is also fair to say that no compelling model has yet emerged. In this section we discuss features of first-order inflation that may be generic and depend only upon the fact that the inflation is terminated by bubble nucleation in a first-order transition.

### *Density Perturbations* [39]

Density perturbations certainly arise as remnants of the bubbles that are nucleated during the phase transition; these perturbations have been addressed elsewhere. While it is possible that the density perturbations that arise due to the bubbles are interesting, it seems uncertain: If bubble nucleation turns on rapidly, there will be very few bubbles of cosmologically interesting size; if bubble nucleation turns on slowly, there will be too many large bubbles to be consistent with the isotropy of the cosmic microwave background radiation (CMBR). Unless the bubble nucleation rate is just so, it is not possible for relic bubbles to be both interesting and consistent with the isotropy of the CMBR. In any case, we will focus on the density fluctuations that arise due to quantum fluctuations in the various fields in the theory during extended inflation. For comparison, in slow-rollover inflation it is these fluctuations in the inflaton field that lead to the dominant density perturbations: scale-invariant (Harrison–Zel'dovich) curvature perturbations (perturbations characterized by constant amplitude at horizon crossing), and that also necessitate a very small coupling constant for the inflaton field. Curvature perturbations arise in extended inflation, but they are not quite scale invariant (they have a power-law spectrum), and they arise due to fluctuations in the Brans–Dicke field (the field whose value controls the value of the gravitational constant). Most importantly, no dimensionless parameter needs to be set to a very small value to ensure that they are of an acceptable size. The amplitude of the fluctuations is determined by the ratio of the unification scale to the Planck scale.

Now we compute these perturbations by a conformal transformation to the Einstein frame. In this frame, extended inflation closely resembles slow-rollover inflation, with the Brans–Dicke field playing the role of the inflaton with an exponential potential. For this reason, the formulas derived for curvature fluctuations and graviton production in slow-rollover inflation are directly applicable.

Recall that in extended inflation there is little variation in $\Phi$ during the matter or radiation-dominated regimes, the value of $\Phi$ at the end of inflation, denoted as $\Phi_E$ is approximately equal to its value today:

$$\begin{aligned} \Phi_E &\simeq (16\pi G_N)^{-1} \equiv m_{Pl}^2/16\pi G_N \simeq \Phi_0 B^2 {t_{\rm END}}^2 \\ &\Rightarrow t_{\rm END} \equiv \frac{m_{Pl}}{\rho_V^{1/2}}\sqrt{\frac{(6\omega+5)(2\omega+3)}{32\pi}}, \end{aligned} \tag{5.1}$$

where the time $t_{\rm END}$ corresponds to the end of extended inflation. Generally we will be interested in the large-$\omega$ limit. Since the factor $\sqrt{(6\omega+5)(2\omega+3)/32\pi\omega^2}$ will enter most of the equations, we will denote it as $q$. We will also set $\rho_V = M^4$. Since the value of $\Phi$ at the end of inflation will affect our results for the amplitude of the density perturbations, one should keep in mind the possibility that in a more realistic model $\Phi$ might evolve significantly after inflation, in which case $\Phi_E$ would be very different from $m_{Pl}^2/16\pi$.

The physical wavelength of a linear perturbation grows with the scale factor of the Universe: $\lambda_{\rm phys} \propto a(t)$. During inflation, a given perturbation begins sub-Hubble sized and then crosses outside the physics horizon; later, during the matter- or radiation-dominated epoch, it crosses back inside the horizon. We will need to know the time $t$ that a fluctuation of present physical wavelength $\lambda$ crossed outside the horizon during extended inflation; in terms of this time, $\lambda$ is given by

$$\lambda = \frac{M}{2.75\,{\rm K}}\;\frac{a(t_{\rm END})}{a(t)}\;H^{-1}(t), \tag{5.2}$$

where the reheat temperature is assumed to be $M$ and $a(t_0)/a(t_{\rm END}) = M/2.75$ K. Writing $\lambda = \lambda_{\rm Mpc}$ Mpc $\simeq \lambda_{\rm Mpc} 10^{38}\,{\rm GeV}^{-1}$ and taking $a(t_{\rm END})/a(t) \simeq (t_{\rm END}/t)^{\omega+1/2}$ it follows that

$$\begin{aligned} \lambda_{\rm Mpc} &= 10^{-25}\frac{q\,m_{Pl}}{M}(t_{\rm END}/t)^{\omega-1/2}; \\ \frac{t_{\rm END}}{t} &= 10^{25/(\omega-1/2)}\left(\frac{M}{q\,m_{Pl}}\right)^{1/(\omega-1/2)} {\lambda_{\rm Mpc}}^{1/(\omega-1/2)}. \end{aligned} \tag{5.3}$$

In the Einstein frame the Brans–Dicke field $\psi$ takes on the appearance of a minimally coupled scalar field with with a potential, $V(\psi) = M^4 \exp(-2\psi/\psi_0)$. The equation of motion for $\psi$ is familiar:

$$\ddot{\psi} + 3\bar{H}\dot{\psi} - \frac{1}{\bar{a}^2}\nabla^2\psi + \frac{dV(\psi)}{d\psi} = 0. \tag{5.4}$$

Assuming that the $\Phi$ field is homogeneous, its evolution is just that of a "slow roller:" $d\psi/d\bar{t} \simeq -(dV/d\psi)/3\bar{H}$.

Because $\psi$ behaves just like an inflaton field and because the gravitational part of the action is just that of general relativity, we can compute the curvature fluctuations that result from quantum fluctuations in $\psi$ by taking advantage of the machinery developed for slow-rollover inflation. When a given scale $\lambda$ crosses back inside the horizon after extended inflation (denoted by "HOR") the amplitude of the fluctuation on that scale is given by

$$\left(\frac{\delta\rho}{\rho}\right)_{\rm HOR} \simeq \frac{\bar{H}^2}{d\psi/d\bar{t}} \simeq -\frac{3\bar{H}^3}{dV/d\psi}; \tag{5.5}$$

where the quantities on the right side of Eq. (5.5) are to be evaluated when the scale crossed outside the horizon *during* inflation. Moreover, well after extended inflation the Jordan and the Einstein frames coincide so that the curvature fluctuations in both frames are the same! That is, the fluctuation amplitude in the Jordan frame—which is what we are interested in—is equal to that computed in the Einstein frame—where the amplitude that is most easily and unambiguously computed.

Remembering that $\bar{H}^2 = 8\pi V/3m_{Pl}^2$ and $dV(\psi)/d\psi = -2V/\psi_0$, it is simple to evaluate Eq. (5.5) for $(\delta\rho/\rho)_{\rm HOR}$:

$$\begin{aligned}\left(\frac{\delta\rho}{\rho}\right)_{\rm HOR} &\simeq 4\pi\sqrt{\frac{2\omega+3}{6}}\left(\frac{M}{m_{Pl}}\right)^2 \frac{m_{Pl}^2}{\Phi}, \\ &\simeq 10^{50/(\omega-1/2)} 4\pi\sqrt{\frac{2\omega+3}{6}}\, q^{-2/(\omega-1/2)} \\ &\quad\times \left(\frac{M}{m_{Pl}}\right)^{(2\omega+1)/(\omega-1/2)} \lambda_{\rm Mpc}{}^{2/(\omega-1/2)}.\end{aligned} \tag{5.6}$$

Note that the power-law spectrum of curvature fluctuations that arise due to quantum fluctuations in $\Phi$, given by Eq. (5.6), becomes flatter as $\omega$ becomes large. The amplitude of these fluctuations is very interesting: for $\omega = 10$ and $M = 10^{14}$ GeV,

$$\left(\frac{\delta\rho}{\rho}\right)_{\rm HOR} \simeq 4 \times 10^{-4} \lambda_{\rm Mpc}{}^{0.21}. \tag{5.7}$$

The associated temperature fluctuations on large angular scales, $\theta \sim 1^\circ$ to $180^\circ$, corresponding to scales $\lambda \sim 100$ Mpc to 1000 Mpc, are given by[13]

$$\left(\frac{\delta T}{T}\right)_{\theta \gtrsim 1^\circ} \simeq 10^{50/(\omega-1/2)}\sqrt{\frac{2\omega+3}{6}}q^{-2/(\omega-1/2)}\left(\frac{M}{m_{Pl}}\right)^{(2\omega+1)/(\omega-1/2)}$$
$$\times 10^{4/(\omega-1/2)}(\Omega_0 h)^{-2/(\omega-1/2)}\left(\frac{\theta}{1^\circ}\right)^{2/(\omega-1/2)}, \quad (5.8)$$

where we use the fact that the comoving scale $\lambda$ corresponds to an angular size of $\theta = 34.4''(\Omega_0 h)\lambda_{\rm Mpc}$ at recombination. For $\omega = 10$ and $M = 10^{14}$ GeV, the temperature fluctuations are certainly too large to be consistent with the current limit to the quadrupole anisotropy, $\delta T/T \lesssim 3 \times 10^{-5}$. Increasing $\omega$ or decreasing $M$ slightly can remedy this problem, while still predicting fluctuations of an interesting size on smaller scales. That bubble nucleation occur rapidly enough so that there are not too many large bubbles requires that $\omega$ must be less than about 20.[6] This fact together with the desire to associate $M$ with a scale of order the GUT scale seems to imply that the fluctuations will be both of an interesting magnitude and *not* exactly scale invariant. The fact that the amplitude of the density perturbations increases with scale may be of some importance in that it boosts the fluctuation amplitude on large scales.

Here I have just discussed the adiabatic density fluctuations. Also expected are fluctuations in $\Phi$, as well as fluctuations in any other effectively massless field during inflation. For details of these fluctuations, as well as details about the calculations outlined here, the reader should consult Ref. 39.

### *Baryogenesis* [40]

One of the most important results in particle astrophysics is the development of a framework that provides a dynamical mechanism for the generation of the baryon asymmetry. The baryon number density is defined as the number density of baryons, minus the number density of antibaryons: $n_B \equiv n_b - n_{\bar{b}}$. Today, $n_B = n_b = 1.13 \times 10^{-5}(\Omega_B h^2)$ cm$^{-3}$. Of course, the baryon number density changes with expansion, so it is most useful to define a quantity $B$, called the *baryon number of the universe*, which is the ratio of the baryon number density to the entropy density $s$. Assuming three species of light neutrinos, the present entropy density is $s = 2970$ cm$^{-3}$, and the baryon number is $B = 3.81 \times 10^{-9}(\Omega_B h^2)$. Primordial nucleosynthesis provides the constraint $0.010 \leq \Omega_B h^2 \leq 0.017$,[41] which implies $B = (3.81$ to $6.48) \times 10^{-11}$. So long as baryon number violating processes are slow compared to the expansion rate and no entropy is created in the expansion, $B$ is constant.

A key feature of inflation is the creation of a large amount of entropy in a volume that was at one point in causal contact. The creation of entropy in inflation would dilute any pre-existing baryon asymmetry, so it is necessary to create the asymmetry after, or very near the end of, inflation. In order for the baryon number to arise after inflation in the usual picture, it is necessary for three criteria to be satisfied: baryon number (B) violating reactions must occur, C and CP invariance must be broken, and non-equilibrium conditions must obtain. There are two standard scenarios for baryogenesis:[2] In the first picture the baryon asymmetry is produced by the "out of equilibrium" B, C, and CP violating decays of some massive particle, while the second scenario involves the evaporation of black holes.

In the out of equilibrium decay scenario, the most likely candidate for the decaying particle is a massive boson that arises in Grand Unified Theories (GUTs). In the simplest models, the degree of C and CP violation is larger for Higgs scalars than for the gauge vector bosons, so we will assume that the relevant boson is a massive Higgs particle. This Higgs is also taken to be different from the inflaton. The Higgs of GUTs naturally violate B. The origin of the C and CP violation necessary for baryogenesis is uncertain. It is practical simply to parameterize the degree of C and CP violation in the decay of the particle. To illustrate such a parameterization, imagine that some Higgs scalar $H$ has two possible decay channels, to final states $f_1$, with baryon number $B_1$, and $f_2$, with baryon number $B_2$. Consider the initial condition of an equal number of $H$ and its antiparticle, $\bar{H}$. The $H$'s decay to final states $f_1$ and $f_2$ with decay widths $\Gamma(H \rightarrow f_1)$ and $\Gamma(H \rightarrow f_2)$, while the $\bar{H}$'s decay to final states $\bar{f}_1$ and $\bar{f}_2$ with decay widths $\Gamma(\bar{H} \rightarrow \bar{f}_1)$ and $\Gamma(\bar{H} \rightarrow \bar{f}_2)$. The decays produce a net baryon asymmetry per $H$-$\bar{H}$ given by

$$\epsilon \equiv \sum_{i=1,2} B_i \frac{\Gamma(H \rightarrow f_i) - \Gamma(\bar{H} \rightarrow \bar{f}_i)}{\Gamma_H}, \tag{5.9}$$

where $\Gamma_H$ is the total decay width. Of course $\epsilon$ can be calculated if one knows the masses and couplings of the relevant particles. Reasonable upper bounds for $\epsilon$ are in the range of $10^{-2}$ to $10^{-3}$, but it could be much smaller. For more details, the reader is referred to Ref. (2).

The non-equilibrium condition is most easily realized if the particle interacts weakly enough so that by the time it decays when the age of the Universe is equal to its lifetime, the particle is nonrelativistic. Then the decay products will be rapidly thermalized, and the "back reactions" that would destroy the baryon asymmetry produced in the decay will be suppressed.

In most successful models of new inflation the reheat temperature is constrained to be rather low. This is due to the fact that new inflation requires flat scalar potentials in order for inflation to occur during the "slow roll" of the scalar field toward its minimum. In order to maintain the flatness of the potential, the inflaton field must be very weakly coupled to all fields so that one-loop corrections to the scalar potential do not interfere with the desired flatness of the potential. The feeble coupling of the inflaton to other fields means that the process of converting the energy stored in the scalar field to radiation ("re"heating) is inherently inefficient. Although it is possible to overcome this difficulty in several ways, it remains a concern for new inflation.

The thermalization process of bubble wall collision at the end of extended inflation provides a natural arena for baryogenesis in the early Universe, as it automatically creates conditions far from thermal equilibrium, exactly as required for B, C, and CP violating GUT processes to produce an asymmetry.

Our only assumption about first-order inflation is that the parameter that determines the efficiency of bubble nucleation, $\epsilon(t) = \Gamma(t)/H^4(t)$, where $\Gamma$ is the nucleation rate per volume and $H$ is the expansion rate of the Universe, has a time dependence that suppresses bubble nucleation early in inflation, then rapidly increases so inflation is brought to a successful conclusion in a burst of bubble nucleation.

In order to keep the discussion as general as possible, consider the salient features of the potential in terms of a few parameters that can be easily identified with any scalar potential that undergoes spontaneous symmetry breaking. The parameters of the potential are assumed to be:

1. $\sigma_0$, the energy scale for SSB, i.e., the VEV of the scalar field.

2. $\lambda$, a dimensionless coupling constant of the inflaton potential. We will assume that the potential is proportional to $\lambda$.

3. $\xi$, a dimensionless number that measures the difference between the false and the true vacuum energy density via $\rho_V = \xi\lambda\sigma_0^4$. $\xi$ must be less than unity for sufficient inflation to occur.

From these few parameters it is possible to find all the information required about the bubbles formed in the phase transition. For instance, an important parameter is the size of bubbles nucleated in the tunnelling to the true vacuum. In the thin-wall approximation, the size of a nucleated bubble is given by $R_C \sim 3(\xi\lambda^{1/2}\sigma_0)^{-1}$. Bubbles smaller than this critical size will not grow, and it is exponentially unlikely to nucleate bubbles larger than this critical size. We will assume that all the true-vacuum bubbles are initially created with size $R = R_C$.

Another interesting parameter is the thickness of the bubble wall separating the true-vacuum region inside from the false-vacuum region outside the bubble. For the potential described above, the bubble wall thickness is

$$\Delta \sim (\lambda^{1/2}\sigma_0)^{-1}. \tag{5.10}$$

Note that the ratio of the bubble thickness to its size is $\Delta/R_C \sim \xi$; as advertised, if $\xi \ll 1$, the thin-wall approximation is valid. We note here that the results are (probably) valid even in the absence of the thin-wall approximation. Finally, the energy per unit area of the bubble wall is $\eta \sim \lambda^{1/2}\sigma_0^3$.

It is necessary to have some idea of the size of bubbles at the end of inflation, when bubbles of true vacuum percolate, collide, and release the energy density tied up in the bubble walls. The bubbles of true vacuum are nucleated with size $R = R_C$. After nucleation the bubble will grow until it collides with other bubbles.

As discussed in the first section, bubbles nucleated at late time will have little growth in coordinate radius, and any increase in the physical size of such a bubble is due solely to the growth in the scale factor between the time the bubble is nucleated and the end of inflation.

The physical size of a bubble nucleated at time $t_{\rm NUC}$ is related to its coordinate size by $R(t_{\rm NUC}) = r(t_{\rm NUC})a(t_{\rm NUC}) = R_C$. If there is negligible growth in the coordinate size of the bubble between the $t_{\rm NUC}$ and end of inflation $t_{\rm END}$, then at the end of inflation the bubble will have a physical size $R(t_{\rm END}) \equiv R = r(t_{\rm NUC})a(t_{\rm END}) = R_C[a(t_{\rm END})/a(t_{\rm NUC})]$. Assume that the burst of bubble nucleation at the end of inflation leads to bubbles all of the same size, $R = \alpha R_C$, where $\alpha \equiv a(t_{\rm END})/a(t_{\rm NUC})$.

Now we have the picture of the Universe at the end of extended inflation. To a good approximation the Universe is percolated by bubbles of true vacuum of size $R = \alpha R_C$, with all the energy density residing in the bubble walls. The next step is to examine how the release of energy from the bubble walls into radiation via bubble wall collisions takes place.

Now concentrate on a single bubble of radius $R = \alpha R_C$. The collisions of the bubble walls produce some spectrum of particles, which are subsequently thermalized. We need to estimate the typical energy of a particle produced in these collisions. When a bubble forms, the energy of the false vacuum has been entirely transformed into potential energy in the bubble walls, but as the bubbles expand, more and more of their energy becomes kinetic and the walls become highly relativistic. A simple calculation shows that if the bubble has expanded by a factor of $\alpha$ since nucleation as discussed in the previous section, then only $1/\alpha$ of its energy remains as potential energy. The numerical simulations of bubble collisions by Hawking, Moss, and Stewart[44] demonstrate that during collisions the walls oscillate through each other, and it seems reasonable that the kinetic energy is dispersed at an energy related to the frequency of these oscillations (see their discussion of phase waves). The kinetic energy is presumably dispersed into lower energy particles, and does not participate in baryogenesis. We are more interested in the fate of the potential energy. The bubble walls can be imagined as a coherent state of inflaton particles, so that the typical energy of the products of their decays is simply the mass of the inflaton. This energy scale is just equal to the inverse thickness of the wall. Note that by the time the walls actually disperse, most of the kinetic energy has been radiated away,[44] so the walls are probably no longer highly relativistic.

The probable first step in the reheating process is converting this coherent state of Higgs into an incoherent state. The next step would be the conversion of the incoherent state of Higgs into other particles either through decay of the Higgs, or through inelastic scattering. We are assuming that baryon-number violating bosons $H$ will be produced in the process. The $\sigma$ field is typically in the adjoint representation of the gauge group, while $H$ is typically in the fundamental or some other representation. It is possible to envision some symmetry forbidding a direct $\sigma$–$H$ coupling, or that the coupling is very small compared to other couplings. If this is the case, production of $H$ relative to other particles will be suppressed by some power of the small coupling constant. However in the generic case where all couplings are of the same magnitude there will be no suppression. Of course the ultimate answer is model dependent but calculable.

As discussed earlier, bubbles do not grow substantially before percolation in our idealized extended inflation model. Hence $\alpha$ remains not too far from 1, although a growth by a factor of 1000 even will not necessarily rule out the model. The bubble wall collisions yield a significant amount of the original false-vacuum energy in the form of potential energy, giving rise to high energy particles. The potential energy in the bubble walls is given by $M_{\rm POT} = 4\pi\eta R^2 \sim 4\pi\lambda^{1/2}\sigma_0^3 R^2$. Taking the mean energy of a particle produced in the collisions to be of the order of the inverse thickness of the wall, $\langle E\rangle \sim \Delta^{-1}$, the mean number of particles produced in the collisions from the wall's potential energy is

$$\langle N\rangle \simeq M_{\rm POT}/\langle E\rangle \sim 4\pi\Delta\lambda^{1/2}\sigma_0^3 R^2. \tag{5.11}$$

In general, the bubble collisions will produce all species of particles, at least all species with masses not too large compared to $\langle E\rangle$. In the following we will assume that this is the case for the baryon-number violating Higgs particles. If the Higgs mass exceeds $\Delta^{-1}$ by a significant amount, we can expect some suppression, presumably exponential, in the number of Higgs formed. This possibility will be discussed at the end of this section. For now, we simply parameterize the fraction of the primary annihilation products that are supermassive Higgs by a fraction $f_H$, which in general will depend on the masses and couplings of a particular theory in question. The typical number of Higgs particles produced per bubble is

$$\langle N_H\rangle \sim f_H\langle N\rangle \sim 4\pi f_H\Delta\lambda^{1/2}\sigma_0^3 R^2. \tag{5.12}$$

Now assume that the only source of the supermassive Higgs is from the primary particles produced in the bubble-wall collisions. This will be true if the reheat temperature, $T_{RH}$, is below the Higgs mass.

The Higgs particles produced in the wall collisions decay, producing a net baryon asymmetry $\epsilon$ per decay, where $\epsilon$ is given in Eq. (5.9). Hence, the excess of baryons over antibaryons produced from a single bubble, $N_B = N_b - N_{\bar b}$, is given by

$$N_B = \epsilon\langle N_H\rangle \sim 4\pi\epsilon f_H\sigma_0^2 R^2, \tag{5.13}$$

where we have substituted in for the bubble thickness from Eq. (5.10). This results in a baryon number density of

$$n_B = N_B/(4\pi R^3/3) = 3\epsilon f_H\sigma_0^2 R^{-1}. \tag{5.14}$$

Now calculate the entropy generated in bubble-wall collisions. As stated above, the potential energy of a bubble is $M_{\rm POT} = 4\pi\sigma_0^3\lambda^{1/2}R^2$. Including the (possibly dominant) kinetic energy contribution, the total mass of the bubble is $M_{\rm BUB} = 4\pi\sigma_0^3\lambda^{1/2}R^2\alpha$. Thermalization of the mass in the bubble walls will redistribute this energy throughout the bubble, resulting in a radiation energy density

$$\rho_R \sim M/(4\pi R^3/3) \sim 3\lambda^{1/2}\sigma_0^3\alpha/R = \xi\lambda\sigma_0^4, \tag{5.15}$$

which is just the false vacuum energy. The reheat temperature is related to the radiation energy density via $\rho_R = (g_*\pi^2/30)T_{RH}^4$, where $g_*$ is the effective number of degrees of freedom in all the species of particles which may be formed in the thermalization process. From this we obtain the entropy density, $s$, produced by the thermalization of the debris from bubble-wall collisions:

$$s = \frac{2\pi^2}{45} g_* T_{RH}^3 \sim g_*^{1/4} \xi^{3/4} \lambda^{3/4} \sigma_0^3. \tag{5.16}$$

From Eqs. (5.14) and (5.16) we can calculate the baryon asymmetry $B$ as

$$B \equiv \frac{n_B}{s} = \epsilon f_H \, \alpha^{-1} g_*^{-1/4} \lambda^{-1/4} \xi^{1/4}. \tag{5.17}$$

Provided the mass of the Higgs is less than $T_{RH}$, one might conjecture that $f_H$ is given simply by $g_H/g_*$, where $g_H$ is the number of Higgs degrees of freedom; that is, all suitably light particles are produced equally. In general the situation will be more complex, and the fraction of Higgs produced will depend on the various couplings in the theory. This introduces a model dependence into the picture, though in fact one can always regard $\epsilon f_H$ as a single unknown parameter. For simplicity, we assume here that all particles are indeed produced equally. Substituting this gives the final result $B = \epsilon g_H \, \alpha^{-1} g_*^{-5/4} \lambda^{-1/4} \xi^{1/4}$. This allows us to make numerical estimates of $B$ based on sample values of these parameters. Notice that the dependence on both $\lambda$ and $\xi$, which are the two parameters on which the inflaton's potential depends, is very weak. The important contributions are the degree of asymmetry in CP violating Higgs decays, the number of particle species available for production in the wall collisions and the factor $\alpha$ by which bubbles expand before colliding. Numerical estimates for $B$ based upon this expression will be made in the concluding section.

It is also interesting to note the possibility of isothermal perturbations arising from the thermalization process. While we have assumed throughout this paper that at percolation all the true vacuum bubbles have the same size, the full picture is somewhat more complicated, as bubbles formed earlier in inflation will grow to larger sizes than those formed right at the end. While homogeneity of the microwave background requires large bubbles to be suppressed, one would still expect to see a range of sizes of small bubbles, and hence spatial variations in the ratio of baryon number density to entropy density from point to point.

In conclusion then, we have seen that the result of the first-order phase transition bringing extended inflation to an end is an environment well out of thermal equilibrium. In such conditions baryogenesis via the decay of baryon number violating Higgs particles can proceed, and we have demonstrated a means by which the baryon number can be estimated. The mechanism has further been shown to work for a large range of model parameters and to have the capability of predicting a baryon asymmetry of the required magnitude.

For more details on baryogenesis, the reader is referred to the original paper of Barrow, Copeland, Kolb, and Liddle.[40]

### *Black Holes* [45]

There are three possible sources for the formation of small primordial black holes after extended inflation. Holes may form via the gravitational instability of inhomogeneities formed during the thermalization phase; there is the possibility of trapped regions of false vacuum (within their Schwarzschild radii) caught between bubbles of true vacuum;[46] and there is the possibility that black holes are formed in the collision process.[44]

Unfortunately, the technical details of even estimating the typical number density and mass of the black holes formed by these processes are quite difficult. Some progress in this direction was made by Hawking, et al.,[44] in the context of the original inflationary scenario, and more recently Hsu[47] has examined black hole production from false vacuum regions in extended inflation. In order to keep the discussion on a more general footing, for now simply assume that some fraction $\beta$ of the energy after collisions is in black holes, while the remaining $1-\beta$ is in radiation,[48] and later consider the various outcomes implied by the differing values of $\beta$.

The total energy density at the end of extended inflation is partitioned between the energy density of radiation, $\rho_R$, and black holes, $\rho_{BH}$:

$$\begin{aligned}
\rho(t_{\rm END}) &= \rho_R(t_{\rm END}) + \rho_{BH}(t_{\rm END}) \\
\rho_R(t_{\rm END}) &= (1-\beta)\rho(t_{\rm END}) = \frac{\pi^2}{30} g_* T_{RH}^4 \\
\rho_{BH}(t_{\rm END}) &= \beta\rho(t_{\rm END}) = M_0 n_{BH}(t_{\rm END}),
\end{aligned} \tag{5.18}$$

where $T_{RH}$ is the reheat temperature, $M_0$ is the initial mass of the black holes formed (for convenience we will assume that they all have the same mass), and $n_{BH}$ is the number density of black holes. The time $t_{\rm END}$ can also be expressed in terms of $\rho(t_{\rm END})$:

$$t_{RH}^2 \equiv \left(\frac{3}{32\pi}\right) \frac{m_{Pl}^2}{\rho(t_{\rm END})}. \tag{5.19}$$

(For matter domination, the factor $3/32\pi$ is replaced by $1/6\pi$.) From $H_{\rm END}$ and $\rho$ we also define a "horizon mass" at the end of inflation:

$$M_{\rm HOR} = \frac{4\pi}{3}\rho(t_{\rm END})(2t_{RH})^3 = \left(\frac{3}{32\pi}\right)^{1/2} \frac{m_{Pl}^3}{\rho^{1/2}(t_{\rm END})}. \tag{5.20}$$

(The right hand side is the same in the matter dominated case.) $M_{\rm HOR}$ represents the mass within the "physics horizon," at the end of inflation, and plays the same role as the mass within the horizon in the standard FRW model.

Once formed, the black holes evaporate at a rate given by

$$\dot{M}_{BH} = -\frac{g_*}{3}\frac{m_{Pl}^4}{M_{BH}^2}, \tag{5.21}$$

which leads to a time dependence of the black hole mass of

$$M_{BH}^3(t) = M_0^3 - g_* m_{Pl}^4 (t - t_{\rm END}). \tag{5.22}$$

It is convenient to define a black hole lifetime,

$$\tau \equiv M_0^3 / g_* m_{Pl}^4, \tag{5.23}$$

and the expression for the mass as a function of time becomes $M(t) = M_0[1 - (t - t_{\rm END})/\tau]^{1/3}$. The evaporation ends at time $t_{BH} = t_{RH} + \tau$.

Black holes radiate as blackbodies with temperature $T_{BH} = m_{Pl}^2/8\pi M_{BH}$. This allows us to calculate what is, for our purposes, the most important quantity—the number of particles emitted during the course of the evaporation. Let us first calculate the number of particles emitted while the black hole is between the temperatures $T$ and $T+dT$. The change in mass of the black hole, $dM$, which is the amount of energy radiated as particles, is given by

$$dM = \frac{m_{Pl}^2}{8\pi}\left(\frac{1}{T} - \frac{1}{T+dT}\right). \tag{5.24}$$

Each emitted particle has energy $3T$ (the mean energy of a particle in a Maxwell-Boltzmann distribution at temperature $T$), so the number of particles emitted between those temperatures is just

$$dN = \frac{m_{Pl}^2}{24\pi T}\left(\frac{1}{T} - \frac{1}{T+dT}\right) = \frac{m_{Pl}^2}{24\pi T^3}dT. \tag{5.25}$$

Integrating this, we find that the number of particles emitted as the black hole temperature increases from its initial temperature $T_0$ to $\infty$ is

$$N = \frac{4\pi M_0^2}{3m_{Pl}^2}. \tag{5.26}$$

Note that this gives the total number of particles emitted.

It is interesting to consider the possibility that amongst the particles radiated are Higgs bosons, again denoted as $H$, whose decay can lead to the baryon asymmetry. Again, $B$ will depend upon the fraction of the particles emitted as $H$, denoted as $f_H$. To determine the appropriate form for $f_H$, the initial temperature of the black hole at formation may be important. If it is less than the mass of the Higgs boson, $m_H$, then the thermal spectrum in the initial phase of the evaporation will not include Higgs as the typical energy is not high enough to produce so massive a particle. Only when the black hole temperature has increased to $m_H$ will the thermal radiation include a significant fraction of Higgs. This can lead to an overall suppression in the number of Higgs produced during the complete course of the evaporation. Once the temperature is high enough to radiate Higgs, we expect that the energy of radiated particles will be distributed evenly amongst all radiated species, so that $f_H$ is a constant given by $g_H/g_*$.

Black hole evaporation affects the evolution of both components of the total mass density. Since the hole mass is decreased by evaporation, the evolution of the black hole energy density, which in the absence of evaporation would be that of nonrelativistic matter ($\rho_{NR} \propto a^{-3}$, where $a$ is the scale factor), is altered. The production of radiation from the hole evaporation also modifies the evolution of radiation energy density, which normally scales as $a^{-4}$. Of course, the departure of the energy densities from the normal evolution is most pronounced around the time $t \sim t_{RH} + \tau$. An exact treatment of this effect is given in Ref. (45), where a network of equations is derived describing the evolution of the different components of the energy density and also the evolution of the baryon asymmetry. In order to understand the general results, let us for the moment ignore the complication resulting from the decrease of the hole mass.

Two different situations arise, depending on whether black holes or radiation dominate the energy density of the Universe at the time the holes evaporate.[48] If $\beta < 1/2$, then the evolution of the scale factor is that appropriate to a radiation-dominated Universe, i.e., $a(t) \sim t^{1/2}$, and the energy density of black holes goes as $a^{-3} \propto t^{-3/2}$, while that of radiation goes as $a^{-4} \propto t^{-2}$. Therefore, provided their lifetime is sufficiently long, black holes will come to dominate the Universe at a time $t_* = t_{\rm END}(1-\beta)^2/\beta^2$, and hence if $\tau > t_* - t_{\rm END}$, they will come to dominate before their evaporation. If $\beta > 1/2$, black holes dominate even initially. If the black holes dominate before evaporation, then their evaporation produces not only the baryons, but also the entropy.

For the details of the calculations the reader is referred to Ref. (45). Here I shall simply summarize the results.

First consider the case where black hole evaporation occurs before domination. This corresponds to small $\beta$ and initially light black holes. Since the black holes never dominate, the Universe expands like a radiation-dominated Universe, with $a \propto t^{1/2}$. If the black holes evaporate before domination, their radiation will not significantly change the background entropy density.

In this case the final baryon asymmetry is

$$B_A \equiv \frac{n_B}{s} = \frac{1}{2}\epsilon f_H \left(\frac{45\pi}{g_*}\right)^{1/4} \left(\frac{M_0}{m_{Pl}}\right)^{1/2} \left(\frac{M_0}{M_{\rm HOR}}\right)^{1/2} \frac{\beta}{(1-\beta)^{3/4}}, \tag{5.27}$$

where we have used Eq. (5.20). Note that the penultimate factor gives the initial black hole mass as a fraction of the horizon mass.

Now consider the second possibility, that holes evaporate after they dominate the energy density. This divides into two further sub-cases; in the former, black holes come to dominate at time $t_*$ as defined earlier, while in the latter black holes dominate immediately after formation.

In the first of these sub-cases, once $t > t_*$ the scale factor evolves as appropriate for a matter-dominated Universe, $a(t) \sim t^{2/3}$, and so $\rho_{BH}(t) = \rho_{BH}(t_*)(t_*/t)^2$ and $\rho_R(t) = \rho_R(t_*)(t_*/t)^{8/3}$, with the energy densities equal at $t_*$.

As before, the evaporation of a single black hole gives a baryon number $n_B = \epsilon f_H N\ n_{BH}(t_{BH})$. This time, though, the entropy is also determined by the other black hole evaporation products, as they provide the dominant contribution. The result for the baryon number is

$$B_{B1} = \frac{1}{2}\epsilon f_H \left(\frac{45\pi}{g_*}\right)^{1/4} \left(\frac{M_0}{m_{Pl}}\right)^{1/2} \left(\frac{M_0}{M_{\rm HOR}}\right)^{1/2} (1-\beta)^{1/4} \left(1+\frac{\tau}{t_{RH}}\right)^{-1/2}. \tag{5.28}$$

This expression is very similar to that obtained in the "evaporation before domination" scenario; in particular the black hole mass appears in the same functional form, and the prefactors are all the same with the exception of the $\beta$ term, which naturally has changed as we move to a different physical situation. The last factor demonstrates how a long black hole lifetime dilutes the baryon asymmetry obtained; if $\tau$ is very small this factor is just equal to one, while for $\tau \gg t_{RH}$ we get a reduction in the baryon asymmetry by a factor of about $\sqrt{M_0^3/M_{\rm HOR} m_{Pl}^2 g_*}$. Clearly, this factor can be important for long-lived (initially massive) black holes. These are also exactly the type of holes that one might expect to survive long enough to come to dominate even if $\beta$ is originally substantially less than 1/2.

We now examine the second sub-case of black hole domination—that in which the black holes dominate even initially. The black hole energy density is now given by $\rho_{BH}(t) = \rho_{BH}(t_{RH})(t_{RH}/t)^2$, and

$$B_{B2} = \frac{1}{2}\epsilon f_H \left(\frac{45\pi}{g_*}\right)^{1/4} \left(\frac{M_0}{m_{Pl}}\right)^{1/2} \left(\frac{M_0}{M_{\rm HOR}}\right)^{1/2} \beta^{1/4} \left(1+\frac{\tau}{t_{RH}}\right)^{-1/2}, \qquad (5.29)$$

which is just Eq. (5.28) multiplied by $(\beta/(1-\beta))^{1/4}$. This factor represents the dilution of the black hole energy density up to domination. As expected, Eqs. (5.28) and (5.29) match in the case of marginal domination where $\beta = 1/2$. The $\beta$ dependence in Eq. (5.29) simply reflects the fraction of the horizon mass contributed by black holes. It differs from Eq. (5.28) because here there is no evolution in the initial radiation-dominated phase, hence no era of dilution before domination. In the case of Eq. (5.28) an extra multiplier of $[(1-\beta)/\beta]^{1/4}$ is needed to account for the evolution in the initial radiation-dominated phase.

This completes the set of results for the different regions of domination, and is summarized in Table I. Many more details are to be found in the paper of Barrow, Copeland, Kolb, and Liddle.[45]

Table I. Results for the baryon number produced by black hole evaporation depend upon $\beta$ (the fraction of the energy of the Universe in black holes at $t = t_{\rm END}$, where $t_{\rm END}$ is taken to be the end of inflation), $t_*$ (the time at which the black holes dominate the mass of the Universe), and $\tau = M_{BH}^3/g_* m_{Pl}^4$ (the lifetime of a black hole of mass $M_{BH}$).

| $\beta$ | $\tau$ | $B \equiv n_B/s$ |
|---|---|---|
| $\beta < 1/2$ | $\tau < t_* - t_{\rm END}$ | Eq. (5.27) |
| $\beta < 1/2$ | $\tau > t_* - t_{\rm END}$ | Eq. (5.28) |
| $\beta > 1/2$ | independent of $\tau$ | Eq. (5.29) |

*Topological Defects* [50]

I have already discussed the generation of adiabatic density fluctuations during extended inflation. However there might very well be a different mechanism for the formation of structure after extended inflation, namely the formation of topological defects in the inflaton field formed as it passes through the phase transition. Calculations of the false-vacuum decay rate made so far consider the evolution from a false-vacuum state to a *unique* true-vacuum state. However, the inflaton is far more likely to have degenerate minima, especially if it is part of a grand-unified Higgs sector.

Recall the standard picture of defect formation in a smooth second-order phase transition.[51] At early times the universe was very hot and the fields describing interactions were in a highly symmetric phase. However as the universe expanded and cooled, symmetry breaking occurred, which may have left behind remnants of the old symmetric phase, possibly in the form of strings, domain walls or monopoles. Here, we concentrate on strings.

Strings appropriate to galaxy formation are required to have a line density of $G\mu \sim 10^{-6}$, where $\mu \sim \sigma_0^2$, corresponding to a breaking scale of $10^{-3}$ Planck masses. Unfortunately, generic new and chaotic inflationary scenarios occur at or below this energy scale, and hence the strings form before or early in the inflationary epoch and are rapidly inflated away. It has been demonstrated that the universe cannot be made to reheat after inflation to sufficiently high temperatures as to restore the symmetry of the string-forming field and allow a new phase of string formation after inflation.[52,53] This leads to the incompatibility of cosmic strings with new or chaotic inflation. These arguments apply whether the inflaton and the cosmic string fields are the same field or different ones (in chaotic inflation the inflaton field can never be identified with the cosmic string field as the symmetry is broken even initially). In the case where the cosmic string field is distinct from the inflaton field, models have been proposed which resolve the conflict. The model of Vishniac, Olive, and Seckel[53] couples the inflaton and the string field in a particular way, but the only motivation for doing this is to solve the strings-inflation problem, so their solution appears unnatural. More recently, Yokoyama[54] has suggested that a non-minimal coupling to gravity of the string field can hold it in its symmetric phase during inflation, and allow strings to form at the end of inflation.

Now consider the picture of string formation in extended inflation, where the fact that the transition is first order has crucial consequences. As the Universe cools from high temperatures, a complex scalar field is trapped in a false-vacuum state and the Universe enters a phase of rapid power-law expansion. Bubbles of true vacuum then begin to nucleate and grow at the speed of light. Due to the presence of event horizons in the inflating Universe they grow to a constant comoving volume which depends on their time of formation. The important ingredient to our scenario is that each bubble forms independently of the rest, and so there is no correlation between the choice of true vacuum made in each bubble from the selection of degenerate true vacua. Eventually the bubbles grow and collide, finally percolating the Universe and bringing the inflationary era to an end.

At the end of inflation, the collision of bubble walls (in which all the energy is held) produces particles and causes thermalization of the energy. However, because the scalar field is only correlated on the scale of a bubble, we can expect topological defects to be present. The usual arguments state that there is typically of order one cosmic string per correlation volume of the scalar field, and hence we expect roughly one string per mean bubble size at the end of inflation.

This model for the formation of strings allows for the existence of large voids, which would be a consequence of the rare large bubbles. Although the typical string separation at the end of inflation is $\xi_{\rm eff}$, extended inflation allows for the possibility of rare large bubbles, formed by quantum tunnelling early in inflation. The true vacuum formed inside bubbles contains no matter (any matter originally in that volume is assumed to be inflated away while the scalar field dominates the energy density). All the energy of the Universe after inflation is contained in the walls of the expanding bubbles which collide to form matter and to cause thermalization of the energy density. After collisions, matter will flow back into the void, though as it cannot travel faster than light, we can calculate the minimum time the bubble will require to thermalize. A large bubble will have a coherent scalar field vacuum and hence no strings will be formed within it—we can thus expect the interior of the bubble to evolve into a large region void of strings. If cosmic strings are to provide the seeds for galaxy formation, then we can expect to see few or no galaxies within the void. The presence of voids is an additional property of this model which may help explain observed large-scale structure.

In fact, at the time of percolation the bubbles may have a range of sizes, which can lead to the formation of an initial string network differing from the usual one. As the correlation length is essentially just the bubble size, and because there would appear to be no *a priori* reason why bubbles everywhere should be *exactly* the same size (at small sizes the assumption of a scale-invariant bubble size distribution would seem more reasonable), the strings will be formed with a randomly spatially varying correlation length. This will presumably lead to higher densities of strings in some regions than others, which again may have implications for structure formation, depending on how much the effects of the initial string distribution might be wiped out by the future evolution and decay of strings. One desirable effect of a more dilute string network would be to avoid the uncomfortable bounds from gravitational wave production from small string loops.[55] The fact that the correlation length will generically be greater (and in some models perhaps much greater) than that of the Kibble mechanism may also have important implications, though perhaps not as great as one might naively suppose if the small strings rapidly disappear from the network once string evolution commences.

These formation arguments can be equally well applied to the cases of domain walls and monopoles, again giving rise to an estimate of order one defect per bubble at the time of bubble collision. In the case of domain walls this will give rise to an excessive number, and will be disallowed on cosmological grounds. Hence, any extended inflation model featuring a potential with domain wall solutions (i.e., a disconnected vacuum manifold) can be ruled out. The situation is less clear for monopoles, because the correlation length may well be substantially greater than that of the Kibble mechanism and hence proportionally fewer monopoles are expected. However, standard estimates of the cosmological monopole abundance[56] give values of perhaps twenty orders of magnitude in excess of the Parker limit,[57] so the correlation length would have to be increased by seven or eight orders of magnitude before being within experimental limits—such an increase seems very unlikely.

If we consider the unification to be part of a grand-unified theory, the problem of monopole overproduction must be addressed, as any breaking to the symmetry of the standard model must produce monopoles at some stage. The simplest method is to arrange for monopoles to be formed in a partial symmetry breaking and then later inflated away in a second transition.

For more details, the reader should see the original paper of Copeland, Kolb, and Liddle.[50]

### *Gravity Waves* [58]

One of the most interesting new features of a completed first-order phase transition is the observation by Turner and Wilczek that a significant amount of gravity waves might be produced in the reheating process.[58]

The beauty of this observation is that it is largely independent of the details of the particular extended inflation model. In the picture of reheating I have been describing here, bubbles of size $R$ and mass $M_{\rm BUB}$ collide. Furthermore, the bubbles are most likely relativistic, or at least semi-relativistic, when they collide. The luminosity in gravity waves emitted in such a close encounter can be estimated from the quadrupole formula:

$$L_{GW} \sim G_N \left(\frac{d^3Q}{dt^3}\right)^2 \sim G_N \frac{M_{\rm BUB}^2}{R^2}. \tag{5.30}$$

Thus a bubble collision releases an energy $E_{GW}$ given by

$$E_{GW} \sim RL_{GW} \sim G_N \frac{M_{\rm BUB}^2}{R}, \tag{5.31}$$

in the form of gravitational radiation with wavelength $R$.

Of course it is most useful to compare this energy with the total energy released in the bubble collision. Since the total mass of the bubble, $M_{\rm BUB}$, is eventually released into radiation, then the ratio of the gravitational wave energy density to the radiation energy density is

$$\epsilon_{GW} \equiv \frac{E_{GW}}{M_{\rm BUB}} \sim G_N \frac{M_{\rm BUB}}{R}. \tag{5.32}$$

If this is true after extended inflation, then the present ratio would be approximately $g_*(\text{today})/g_*(T_{RH}) \sim 0.01$ times this value. Since the contribution to $\Omega$ in radiation is today about $3 \times 10^{-5} h^{-2}$, and $\rho_{GW}$ and $\rho_R$ both decrease in expansion as $a^{-4}$, this implies that today $\Omega_{GW} h^2 \sim 10^{-5} \epsilon_{GW}$.

The wavelength of the gravitational radiation today would simply be the wavelength at creation, $\lambda(T_{RH}) \sim R$, redshifted by the expansion of the Universe:

$$\lambda(\text{today}) = R[a(\text{today})/a(T_{RH})] \sim RT_{RH}/2.7\,\text{K} \sim 4\times 10^{26} R(M/10^{14}\text{GeV}), \quad (5.33)$$

where again we have assumed that the re-heat temperature is comparable to the mass scale of symmetry breaking $M$.

Now the question is what to use for $R$. Turner and Wilczek make the reasonable assumption that the size of the bubble is the particle horizon at the end of extended inflation. If this is true, then $G_N M_{\text{BUB}}/R$ is about unity, $\epsilon_{GW} \sim 0.01$, and $\Omega_{GW}h^2 \sim 10^{-5}$. The fact that $\epsilon_{GW}$ is about unity in this case is easy to understand: masses the size of the horizon are moving about with velocities of about the velocity of light! This choice for $R$ also predicts $R \sim H^{-1} \sim m_{Pl}/M^2 \sim 2 \times 10^{-19}(10^{14}\text{GeV}/M)^2$cm, which leads to a present wavelength for the gravity waves of $\lambda(\text{today}) = 8 \times 10^3(10^{14}\text{GeV}/M)$ cm. This is quite interesting because it is within the sensitivity and wavelength range of LIGO II and other large second-generation interferometric detectors.

However it might be equally possible that the bubbles are much smaller. The smallest they might be is $R_C$, their critical size. Let's take the pessimistic view that $R \sim M^{-1}$. If this is true, then $\epsilon_{GW} \sim G_N M_{\text{BUB}}/M = M_{\text{BUB}}M/m_{Pl}^2$. If the bubble size is $M^{-1}$ and the false-vacuum energy is of order $M^4$, then $M_{\text{BUB}} \sim M$, and $\epsilon \sim M^2/m_{Pl}^2 \sim 10^{-10}(M/10^{14}\text{GeV})^2$. This would lead to a present value of $\Omega_{GW}h^2 \sim 2 \times 10^{-15}(M/10^{14}\text{GeV})^2$. Another price to be paid is that the present wavelength of the gravity waves would be much smaller: $\lambda(\text{today}) = 8 \times 10^{-2}$ cm. This is too small in magnitude and wavelength for interferometric detectors.

Clearly the correct answer is model dependent. The latter assumption is most likely far too pessimistic, while the former assumption may turn out to be somewhat optimistic.

## V. CONCLUSIONS

It is clear that rumors of the death of first-order inflation were premature. It offers rich and beautiful possibilities for cosmology. It is also clear that no beautiful and compelling model has emerged. Both modifications to the gravity sector or to the microphysics sector seem to lead to possible models. There is much work to be done in model building.

One outstanding problem that seems interesting in its own right is the problem of two-field tunnelling discussed in Section III. There must be a new formalism developed here. This formalism will have applications outside of cosmology.

Even without the guidance of a definite model it is possible to say that interesting new phenomena are predicted in a completed first-order phase transition. Mentioned in this review are several: density perturbations, baryogenesis, black hole formation, generation of topological defects, and gravity waves. The rough outlines of these phenomena have been considered, but much work remains to be done. Even estimating the spectrum and number of black holes produced in bubble collisions seems difficult. The details of bubble collisions are important for reheating, baryogenesis, and gravity wave production. Probably some serious numerical work is required. The first step in this direction was taken by Hawking, Moss, and Stewart, but it was indeed only a first step.

First-order inflation is an attractive alternative to the usual slow rollover inflation models.

## ACKNOWLEDGEMENTS

I would like to acknowledge my collaborators on the work reported here: John Barrow, Ed Copeland, Rich Holman, Andrew Liddle, David Salopek, Sharon Vadas, Michael Turner, Yun Wang, and Erick Weinberg. All of the work reported here is a result of collaborative efforts with them. I have also benefitted from conversations with Paul Steinhardt, Ian Moss, and Mark Hindmarsh. This work has been supported by the Department of Energy and NASA (grant #NAGW-1340).

## REFERENCES

1. For a history, see D. Lindley, Fermilab Report, unpublished (1985).

2. See, e.g., E. W. Kolb and M. S. Turner, *The Early Universe* (Addison-Wesley, Redwood City, CA., 1990).

3. D. A. Kirtznits, *JETP Lett.*, **15**, 745 (1972); D. A. Kirtznits and A. D. Linde, *Phys. Lett.* **42B**, 421 (1972).

4. E. W. Kolb and S. Wolfram, *Ap. J.* **239**, 428 (1980).

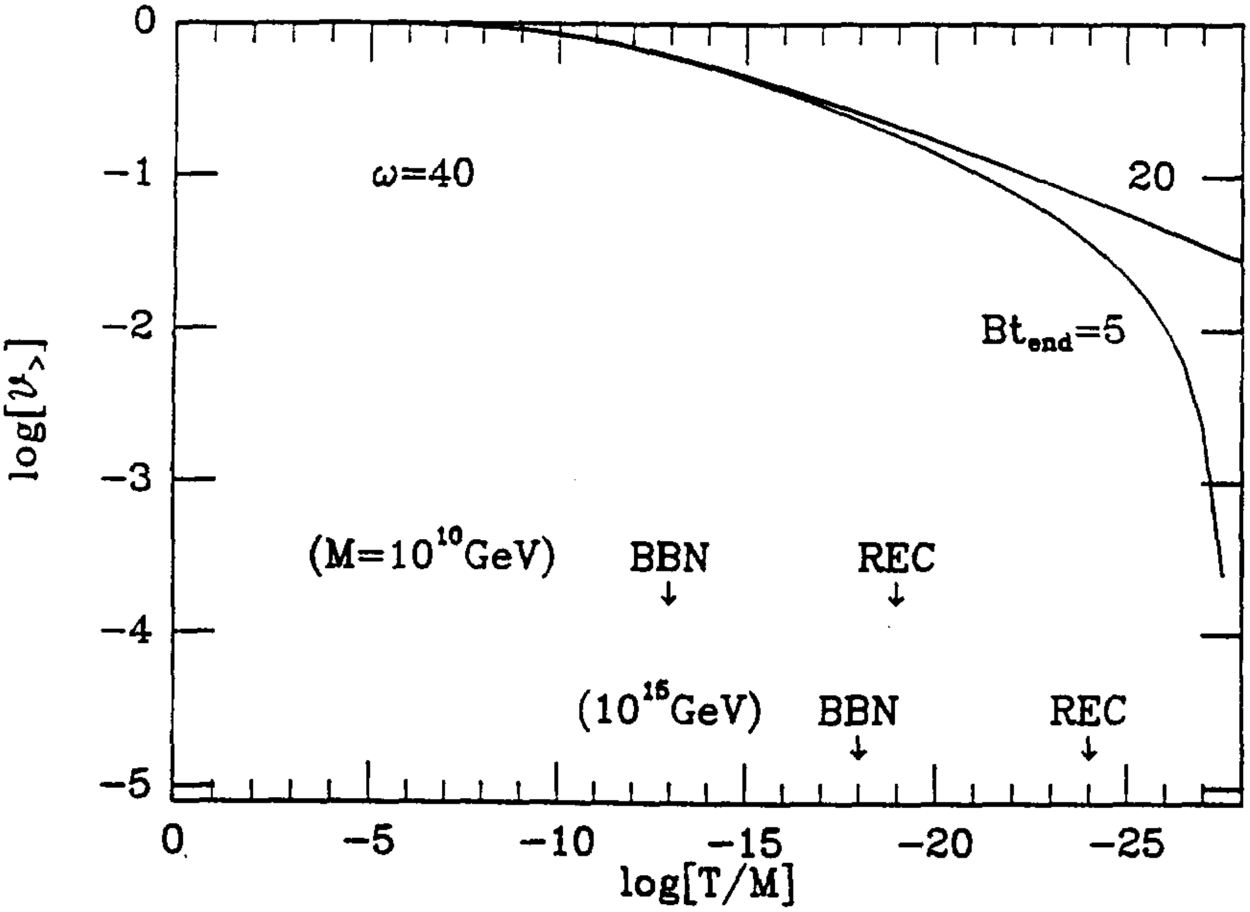
5. A. H. Guth, *Phys. Rev. D* **20**, 30 (1980).

6. S. Coleman, *Phys. Rev. D* **15**, 2929 (1977); C. G. Callan and S. Coleman, *Phys. Rev. D* **16**, 1762 (1977); S. Coleman and F. Du Luccia, *Phys. Rev. D* **21**, 3305 (1980).

7. By percolation it is meant that there are infinite connected regions of true vacuum. If percolation of the Universe by true vacuum bubbles occurs, it is possible that there will be a true-vacuum volume large enough to contain our visible Universe, and it resulted from enough bubbles that the energy produced by bubble wall collisions could thermalize and result in a smooth radiation-dominated Universe.

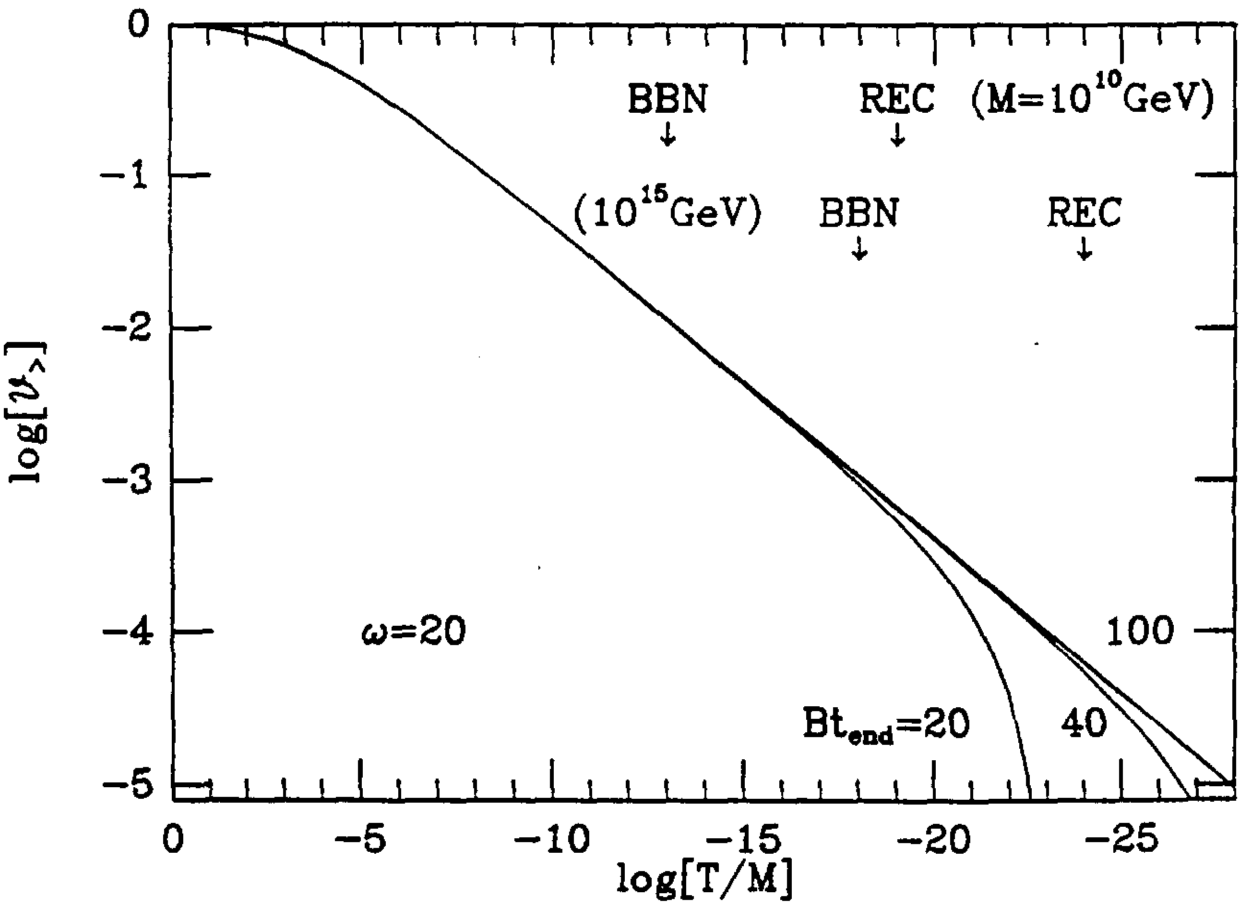
8. A. H. Guth and E. J. Weinberg, *Nucl. Phys.* **B212**, 321 (1983).

9. A. D. Linde *Phys. Lett.* **129B**, 177 (1983); A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48**, 1220 (1982).



ω=40
20
Bt$_{end}$=5
(M=10$^{10}$GeV)
BBN
REC
(10$^{15}$GeV)
BBN
REC
log[$\mathcal{V}_>$]
log[T/M]
0
−1
−2
−3
−4
−5
0
−5
−10
−15
−20
−25


(2a)



BBN
REC
(M=10$^{10}$GeV)
(10$^{15}$GeV)
BBN
REC
ω=20
100
Bt$_{end}$=20
40
log[$\mathcal{V}_>$]
log[T/M]
0
−1
−2
−3
−4
−5
0
−5
−10
−15
−20
−25


(2b)

27. R. Holman, E. W. Kolb, S. L. Vadas, and Y. Wang, *Phys. Rev. D* (1990).

28. A. D. Linde, CERN preprint CERN-TH.5806/90 (1990); F. Adams and K. Freese, MIT preprint (1990).

29. L. Kofman and A. D. Linde, *Nucl. Phys.* **B282**, 555 (1987).

30. F. S. Accetta and J. J. Trester, *Phys. Rev. D* **39**, 2854 (1989).

31. P. J. Steinhardt and F. S. Accetta, *Phys. Rev. Lett.* **64**, 2740 (1990).

32. R. Holman, E. W. Kolb, and Y. Wang, *Phys. Rev. Lett.* **65**, 17 (1990).

33. T. Damour, G.W. Gibbons, and C. Gundlach, *Phys. Rev. Lett.* **64**, 123 (1990).

34. J. Scherk and J. H. Schwarz, *Nucl. Phys.* **B81**, 118 (1974); K. Maeda, *Mod. Phys. Lett. A* **3**, 243 (1988).

35. R. Holman, E. W. Kolb, S. L. Vadas, and Y. Wang, in preparation (1990)

36. T. Kaluza, *Preus. Acad. Wiss.* **K1**, 966 (1921); O. Klein, *Zeit. Phys.* **37**, 895 (1926) and *Nature* **118**, 516 (1926); M. J. Duff, B. E. W. Nilsson, and C. N. Pope, *Phys. Rep.* **130**, 1 (1986).

37. A. Salam and J. Strathdee, *Annals. Phys.* **141**, 163 (1982).

38. D. La and P. J. Steinhardt, *Phys. Rev. Lett.* **62**, 376 (1989).

39. E. W. Kolb, D. S. Salopek, and M. S. Turner, to be published, *Phys. Rev. D* (1990).

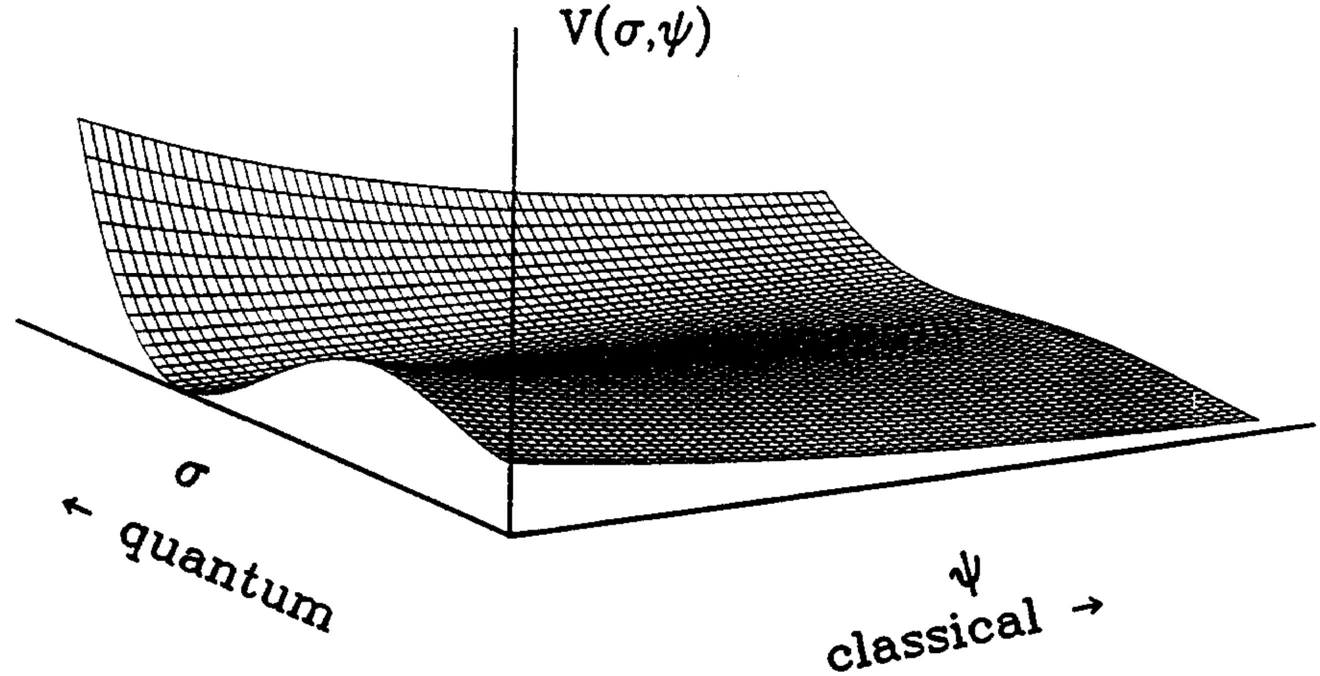
40. J. D. Barrow, E. J. Copeland, E. W. Kolb, and A. R. Liddle, Fermilab Report FNAL-PUB-90/62A.

41. See, e.g., K. A. Olive, "Big Bang Nucleosynthesis: Theory and Observations", Univ. Minnesota preprint (1990); B. E. J. Pagel, in *Evolutionary Phenomena in Galaxies*, eds. J. E. Beckman and B. E. J. Pagel (Cambridge University Press, Cambridge, 1989).

42. J. D. Barrow, *Mon. Not. R. astron. Soc.* **192**, 427 (1980); J. D. Barrow and G. Ross, *Nucl. Phys.* **B181**, 461 (1980).

43. L. F. Abbott and M. B. Wise, *Nucl. Phys.* **B244** 541 (1984); J. D. Barrow, *Phys. Lett.* **B187**, 12 (1987); F. Lucchin and S. Matarrese, Phys. Rev. D **32**, 1316 (1985); J. D. Barrow and P. Coles, *Mon. Not. Roy. astron. Soc.* **244**, 188 (1990).

44. S. W. Hawking, I. G. Moss, and J. M. Stewart, *Phys. Rev. D* **26**, 2681 (1982). It should be noted that the collision of two bubbles of equal size may not at all approximate the process of reheating at the end of extended inflation.




(3)

## FIGURE CAPTIONS

**Figure 1:** Generic forms for the zero-temperatre potentials for a first-order phase transition (1a) and a higher-order transition (1b). The broken curve in (1a) indicates the high-temperature limit of the potential. The high-temperature minimum of the potential becomes a false-vacuum state of the zero-temperature potential.

**Figure 2:** The fraction of the volume of the Universe that has not thermalized as a function of temperature $T$ in units of the mass scale of the potential $M$ for Jordan–Brans–Dicke extended inflation models with the indicated values of $\omega$. The temperature of big-bang nucleosynthesis (BBN) and recombination (REC) are indicated for two different choices of $M$.

**Figure 3:** In the Einstein conformal frame the evolution of the two scalar fields is determined by a potential of this form. The classical behavior of the system is for the the JBD field $\psi$ to slowly roll. The quantum behavior is for the inflaton field $\sigma$ to tunnel from $\sigma = 0$ to its true-vacuum value.