PHYSICAL REVIEW D

PARTICLES AND FIELDS

THIRD SERIES, VOLUME 28, NUMBER 4 15 AUGUST 1983

Spontaneous creation of almost scale-free density perturbations in an inflationary universe

James M. Bardeen

Department of Physics, University of Washington, Seattle, Washington 98195

Paul J. Steinhardt

Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Michael S. Turner

Astronomy and Astrophysics Center, The University of Chicago, Chicago, Illinois 60637

(Received 31 March 1983)

The creation and evolution of energy-density perturbations are analyzed for the "new inflationary universe" scenario proposed by Linde, and Albrecht and Steinhardt. According to the scenario, the Universe underwent a strongly first-order phase transition and entered a "de Sitter phase" of exponential expansion during which all previously existing energy-density perturbations expanded to distance scales very large compared to the size of our observable Universe. The existence of an event horizon during the de Sitter phase gives rise to zero-point fluctuations in the scalar field ϕ, whose slowly growing expectation value signals the transition to the spontaneous-symmetry-breaking (SSB) phase of a grand unified theory (GUT). The fluctuations in ϕ are created on small distance scales and expanded to large scales, eventually giving rise to an almost scale-free spectrum of adiabatic density perturbations (the so-called Zel'dovich spectrum). When a fluctuation reenters the horizon (radius$\simeq H^{-1}$) during the Friedmann-Robertson-Walker (FRW) phase that follows the exponential expansion, it has a perturbation amplitude $\delta\rho/\rho\,|_H = (4 \text{ or } \frac{2}{5})H\Delta\phi/\dot{\phi}(t_1)$, where H is the Hubble constant during the de Sitter phase (H^{-1} is the radius of the event horizon), $\dot{\phi}(t_1)$ is the mean value of $\dot{\phi}$ at the time (t_1) that the wavelength of the perturbation expanded beyond the Hubble radius during the de Sitter epoch, $\Delta\phi$ is the fluctuation in ϕ at time t_1 on the same scale, and 4 ($\frac{2}{5}$) applies if the Universe is radiation (matter) dominated when the scale in question reenters the horizon. Scales larger than about $10^{15}-10^{16}M_\odot$ reenter the horizon when the Universe is matter dominated. Owing to the Sachs-Wolfe effect, these density perturbations give rise to temperature fluctuations in the microwave background which, on all angular scales $\gg 1°$, are $\delta T/T \simeq (\frac{1}{5})H\Delta\phi/\dot{\phi}(t_1)$. The value of $\Delta\phi$ expected from de Sitter fluctuations is $O(H/2\pi)$. For the simplest model of "new inflation," that based on an SU(5) GUT with Coleman-Weinberg SSB, $\dot{\phi}(t_1) \ll H^2$ so that $\delta T/T \gg 1$—in obvious conflict with the large-scale isotropy of the microwave background. One remedy for this is a model in which the inflation occurs when $\dot{\phi}(t_1) \gg H^2$. We analyze a supersymmetric model which has this feature, and show that a value of $\delta\rho/\rho\,|_H \simeq 10^{-4}-10^{-3}$ on all observable scales is not implausible.

I. INTRODUCTION

Guth proposed the original inflationary-universe scenario[1] in an attempt to resolve certain puzzles concerning the conditions of the very early Universe that arise in the standard hot big-bang model. Guth showed that the phase transition associated with the spontaneous symmetry breaking (SSB) of the grand unified theory (GUT) can have a profound influence on the evolution of the Universe. If the transition is strongly first order, the Universe can become trapped in the metastable, symmetric GUT phase which has a vacuum energy density $O(M_G{}^4)$, where M_G is the mass scale associated with the SSB of the GUT. As the Universe cools in the metastable phase to a temperature T less than M_G, the vacuum energy dominates the energy density of the Universe and the scale factor S grows exponentially; $S \propto \exp(t/t_G)$ where $t_G \simeq H^{-1} \simeq M_{\rm Pl}/M_G{}^2$ is the expansion time scale,

$$G^{-1/2} = M_{\rm Pl} = 1.2\times 10^{19}\ \text{GeV} = \text{the Planck mass}\ ,$$

and $\hbar = c = k_B = 1$. Guth argued that if the Universe remains trapped in the symmetric phase for a time $O(65t_G)$ or greater, sufficient inflation (exponential growth) occurs to resolve the isotropy, homogeneity, and flatness/oldness puzzles. However, as Guth himself realized,[1,2] the original scenario has a fatal flaw: there is no "graceful exit" from the inflationary phase to the usual radiation-dominated Friedmann-Robertson-Walker (FRW) universe. For models in which there is sufficient expansion during the metastable phase to resolve the cosmological puzzles, the nucleation of bubbles of symmetric phase is not rapid enough for the bubbles to coalesce and complete the transition to the SSB phase.[2] The end result is a universe comprised of isolated empty bubbles of SSB vacuum separated by exponentially expanding regions of symmetric metastable vacuum.

Recently, Linde[3] and Albrecht and Steinhardt[4] independently proposed a new inflationary scenario which retains the beneficial features of Guth's original scenario while overcoming the graceful exit problem.[5] They considered GUT's in which the spontaneous symmetry breaking is of the type discussed by Coleman and Weinberg.[6] The so-called "Coleman-Weinberg" potential is scale invariant (up to logarithmic terms) and at a value of $\phi \simeq O(T)$ has a small temperature-dependent barrier between the symmetric minimum ($\phi=0$) and the true minimum ($\phi=\sigma$) whose height is $O(T^4)$. The simplest Coleman-Weinberg model is the SU(5) GUT with one-loop effective scalar potential given by

$$V_T(\phi)=\tfrac{1}{2}B\sigma^4+B\phi^4[\ln(\phi^2/\sigma^2)-\tfrac{1}{2}]+\frac{18T^4}{\pi^2}\int_0^\infty dx\, x^2\ln\{1-\exp[-(x^2+25g^2\phi^2/8T^2)^{1/2}]\}\;. \tag{1.1}$$

Here ϕ is the vacuum expectation value of the adjoint (24) Higgs field which breaks SU(5) to SU(3)×SU(2)×U(1), $\sigma=4.5\times10^{14}$ GeV is the value of ϕ at the SU(3)×SU(2)×U(1) minimum, $B=5625\,g^4/1024\pi^2$, and g is the gauge coupling constant. [In Eq. (1.1) the parameters in the potential have been fine tuned so that, after curvature and fluctuation effects are incorporated,[5] the barrier disappears just as $T\to0$.] Albrecht and Steinhardt[4] argued that as the physical temperature decreases below $\sim10^9$ GeV, the barrier becomes negligibly small and the symmetric phase becomes unstable rather than metastable. Fluctuations cause the unstable universe to spinodally decompose[7] into different "fluctuation regions" inside of which ϕ has a nonzero expectation value; there are many SSB minima [due to the breaking of SU(5) to SU(3)×SU(2)×U(1)] and in different fluctuation regions ϕ evolves toward different SSB minima. Near $\phi=0$ the Coleman-Weinberg potential is very flat, $V'(\phi)=4B\phi^3\ln(\phi^2/\sigma^2)$, and so ϕ evolves slowly, but inevitably, towards an SSB minimum. Within a fluctuation region, the time required for ϕ to evolve from its initial value to $\phi\simeq\sigma$ is much longer than t_G. During most of this time $\phi\ll\sigma$, and the energy density inside the fluctuation region is dominated by the nearly constant potential energy, $V(\phi)\simeq\frac{1}{2}B\sigma^4\simeq M_G{}^4$. Therefore, the scale factor S grows exponentially during the time that ϕ is slowly growing. Because inflation occurs as ϕ is evolving towards the stable SSB phase rather than when the Universe is in the metastable symmetric phase, the problem of completing the phase transition to the SSB state is automatically avoided. Sufficient inflation (exponential growth) occurs so that a single fluctuation region grows to a size much larger than the size of our observable Universe.

If the Universe underwent such an expansion, the isotropy, homogeneity, and flatness/oldness puzzles are resolved for the same reasons as in Guth's original scenario. In addition, the problems associated with the production of various massive topological defects (monopoles and domain walls) are solved, because our observable Universe lies within one domain, within which the Higgs field ϕ is aligned in one SSB minimum.[8] Once ϕ evolves to a value $O(\sigma)$, the Coleman-Weinberg potential becomes very steep, $V'(\phi)=8B\sigma^2(\phi-\sigma)$, and ϕ evolves quickly (on a time scale $\ll t_G$). The rapid variation of ϕ causes the vacuum energy to be efficiently converted to radiation through the creation of particles. All quantum fields which couple to ϕ should be radiated directly and their subsequent decays and interactions should rapidly repopulate the Universe with a thermal distribution of particles and radiation.[9] Particle creation due to the time variation of ϕ smoothly reheats the fluctuation region to a temperature of $\sim10^{14}$ GeV, providing a graceful end to inflation and very likely ensuring that baryogenesis proceeds in the usual way.[9]

In spite of the tremendous successes of the inflationary-universe picture, the inflation appears upon closer examination to be too effective. After a long inflationary epoch, the comoving scales of all inhomogeneities existing before the phase transition become exponentially larger than the comoving scale of the presently observable portion of the Universe. In view of this, whence came all the substructures and inhomogeneities that are so conspicuous in the Universe today—stars, galaxies, clusters of galaxies, etc? The answer to this question lies in a closer examination of the state into which a fluctuation region in the Universe evolves during the long inflationary epoch—a nearly de Sitter vacuum.

The important parameter that characterizes a fluctuation region during the inflationary epoch is the expectation value of the scalar field ϕ which is evolving towards a SSB minimum. Linde,[3] and Albrecht and Steinhardt[4] assumed that ϕ was nearly spatially uniform within a fluctuation region from the beginning of the inflationary epoch, and thus was uniform over the much smaller region that was to develop into our observable Universe (which lies deep inside one fluctuation region). Since the local expectation value of ϕ determines the subsequent local evolution, a spatially uniform value of the expectation value leads to a model in which, after the phase transition, our observable Universe is, to a first approximation, perfectly isotropic and homogeneous. On closer examination, it is actually a smooth background model in which the evolution of small density inhomogeneities that are created during the inflationary epoch can be analyzed without any influence of preexisting inhomogeneities. In fact, it is well known that ϕ is not spatially uniform in a de Sitter vacuum. The presence of an event horizon in de Sitter space leads to a spectrum of zero-point fluctuations in the scalar field with wavelengths $O(H^{-1})$, which can be attributed to the Hawking temperature ($\equiv H/2\pi$) associated with the event horizon.

In this paper we will show that the de Sitter zero-point (quantum) fluctuations of the scalar field within the fluctuation region evolve into a "scale-free" spectrum (up to logarithmic factors)[10] of density perturbations—the so-called Zel'dovich spectrum.[11] Fluctuations are produced during the de Sitter phase (exponential expansion) when their proper length scale is comparable to the Hubble radius (H^{-1}). By the end of the inflation phase the proper length scale of fluctuations is expanded to a length much larger than the Hubble radius (which remains roughly constant during the de Sitter epoch). After the Universe reheats and FRW behavior resumes, the Hubble radius grows until it equals the expanded fluctuation scale and the perturbation then "reenters the horizon" as an acoustic wave or a growing density perturbation, depending upon whether the Universe is radiation dominated or matter dominated by this time. The fractional perturbation amplitude on a given scale after reentering the horizon for the acoustic wave in the former case is found to be

$$\delta\rho/\rho\,|_H = 4H\Delta\phi/\dot{\phi}\ . \tag{1.2a}$$

This is a result relevant to questions of galaxy formation (masses $\lesssim 10^{15}-10^{16}M_\odot$). For the latter case the amplitude when the scale reenters the horizon is

$$\delta\rho/\rho\,|_H = \tfrac{2}{5}H\Delta\phi/\dot{\phi}\ . \tag{1.2b}$$

This is the result relevant to the large-scale isotropy of the microwave background (masses $\gtrsim 10^{15}-10^{16}M_\odot$). $\Delta\phi$ and $\dot{\phi}$ are evaluated at the time when the scale expands beyond the Hubble radius during the de Sitter phase. The quantity $\Delta\phi$ is the fluctuation in ϕ when the scale is of the order of the de Sitter Hubble radius ($H^{-1} \simeq M_{\rm Pl}/M_G{}^2 \simeq 10^{-9}$ $\rm GeV^{-1}$); we find that $\Delta\phi$ is of the order of $H/2\pi$.

Observations of the cosmic microwave background severely constrain the amplitude of the Zel'dovich spectrum. On large angular scales ($\gg 1°$) the observed temperature anisotropy is $\lesssim 10^{-4}$. As was shown by Sachs and Wolfe,[12] the predominant anisotropy on these scales is due to perturbations in the "gravitational potential." Their analysis for a matter-dominated background and the Zel'dovich spectrum give

$$\delta T/T \simeq \tfrac{1}{2}(\delta\rho/\rho)\,|_H \cong \tfrac{1}{5}H\Delta\phi/\dot{\phi}(t_1)$$

for angular scales $\gg 1°$. Therefore, the microwave background constrains $H\Delta\phi/\dot{\phi}(t_1)$ to be less than about 10^{-3}, or equivalently $\delta\rho/\rho \lesssim 2\times 10^{-4}$ on the scale of the present horizon.

The type (adiabatic or isothermal) and size of the inhomogeneities required for galaxy formation is a matter of some debate.[13] The nature of the "dark matter," which apparently dominates the mass density of the Universe, whether it is massive neutrinos, gravitinos, or baryons, is certainly an important consideration. If, as the inflationary scenario predicts, Ω is very nearly one, a massive neutrino species (or another relic species) which dominates the mass density is necessary since D is unproduced and ^{4}He is overproduced (relative to their observed abundances) during primordial nucleosynthesis unless $\Omega_{\rm baryon} < 0.2$.[14] In fact, only in a Universe dominated by a light relic species can adiabatic perturbations, whose size is consistent with the isotropy of the microwave background, grow sufficiently to produce galaxies, etc., by the present epoch.[15] Roughly speaking, adiabatic perturbations of size $\delta\rho/\rho \simeq 10^{-4\pm 1}$ when the relevant scales ($\gtrsim$ galactic scale) enter the horizon can account for the present structure. Perturbations on mass scales $\gtrsim 10^{15}$ g (mass of a black hole which would be evaporating in the present epoch) should have amplitudes less than $O(1)$ when they enter the horizon, or else too many black holes would be produced and still be present today. Thus, the Zel'dovich spectrum predicted in inflationary models is viable if the amplitude of the perturbations is $\sim 10^{-4}-10^{-3}$.

In the simplest model for "new inflation," an SU(5) model with Coleman-Weinberg SSB, $\dot{\phi}$ and ϕ^2 are less than $O(H^2)$ during the de Sitter phase and $\delta\rho/\rho \simeq 10$ on the scale of the present horizon. Such a large value is clearly inconsistent with the present state of the Universe and is a devastating blow to the new inflationary scenario. (As Vilenkin and Ford[16] and Linde[17] have recently discussed, these same scalar field fluctuations also lead to an effective negative-mass-squared term in the scalar potential for small values of ϕ, which quickens the evolution of ϕ towards the SSB vacuum and can prevent sufficient exponential expansion. The latter effect might be compensated for by slightly altering the bare mass term in the Lagrangian, but this does not alter the conclusion that the amplitude of the density perturbation is too big.)

These defects can be remedied if the period of slow evolution of the scalar field (during which the Universe grows exponentially in size) takes place when $\phi \gg H$ (and $\dot{\phi} \gg H^2$). Thus, the necessary features of a scalar potential suitable for new inflation seem to be (1) flatness for $\phi \gg H$, to ensure the slow growth of ϕ (and, hence, inflation) and the creation of density perturbations of the desired magnitude, $\delta\rho/\rho\,|_H \sim 10^{-4}-10^{-3}$; (2) strong curvature near the true minimum so that ϕ varies rapidly (compared to H) and sufficient coupling of ϕ to other particles to ensure an efficient conversion of vacuum energy to radiation. Recently it has been proposed that "reverse hierarchy" supersymmetric models lead to inflation without any of the undesirable fine tuning of parameters required in the SU(5) GUT model with Coleman-Weinberg SSB.[18] We will analyze this kind of model, and show that it is easily possible to satisfy the first criterion; the large amplitude of the perturbations found for the SU(5) Coleman-Weinberg model therefore may only be a minor setback for new inflation.

In Sec. II of this paper we will set up the formalism for discussing the evolution of scalar field perturbations in the inflationary phase and their conversion to radiation-density perturbations during reheating. Our starting point will be the general gauge-invariant formalism of Bardeen,[19] adapted to deal with a scalar field rather than a matter- and/or radiation-dominated fluid. There are several alternative formulations of the perturbation equations, based on different ways of slicing the spacetime. The dynamics of the scalar field during the inflationary epoch are particularly simple when viewed from comoving hypersurfaces (on which the energy flux vanishes), but uniform Hubble constant hypersurfaces are best for following the perturbations through reheating. Both independent approaches will be followed. Much of the discussion in this section is highly technical, but in Sec. II A we have provided a qualitative discussion of the evolution of the perturbations to explain why the spectrum of perturbations in the new inflationary scenario is nearly scale invariant. In Sec. III we apply the results of Sec. II to the simple SU(5) model with Coleman-Weinberg SSB, a model for which the effective mass must be fine tuned in order to achieve sufficient inflation; only by artificially adjusting the coupling constant can we obtain a scale-free spectrum of density fluctuations of an acceptable magnitude. Finally, we turn to a semirealistic supersymmetric potential which can be adjusted (rather than fine tuned) to yield the desired results. We briefly discuss and summarize our results in Sec. IV.

II. EVOLUTION OF DENSITY PERTURBATIONS IN AN INFLATIONARY UNIVERSE

A. Heuristic derivation of the scale-invariant spectrum

Although the quantitative analysis of the evolution of the energy-density perturbations is highly technical and gauge dependent, the most important qualitative feature, the scale invariance of the spectrum of perturbations, can be explained intuitively. We believe the following heuristic discussion will serve as a useful guide to the detailed computations that follow.

The Hubble radius ($\equiv H^{-1}$) represents an important scale in the analysis of the creation and evolution of energy-density perturbations. During the de Sitter or inflationary phase, H is roughly constant,

$$H^2 \simeq 8\pi V(0)/3M_{\rm Pl}{}^2 \simeq M_G{}^4/M_{\rm Pl}{}^2 \ .$$

During the subsequent FRW phase $H \propto t^{-1}$ (t=age of the Universe). In either case, the scale factor S undergoes one e-folding in a time interval $O(H^{-1})$. For this reason, microphysics can only operate coherently on proper length scales less than $O(H^{-1})$.

The perturbation spectrum can be characterized by the amplitude of the Fourier components of the density perturbation as a function of their wave number: Determining the perturbation on a given scale signifies computing the amplitude of the Fourier component with an inverse wave number equal to that scale. To each scale can be associated a proper (or physical) wave number k/S which decreases as the Universe expands and a comoving wave number k which remains constant as the Universe expands because the effect of the growing scale factor has been divided out. The evolution of a perturbation is best described as a function of the ratio of the physical wave number to the Hubble constant k/SH.

All scales we observe in our Universe have a value of k/SH that is much greater than unity when the inflationary epoch begins; the horizon problem, for example, is solved in the inflationary scenario because our observable Universe lies within a causal horizon volume [with radius $O(H^{-1})$] before the inflationary phase begins. Microphysics (quantum fluctuations, etc.) can only affect perturbations on scales for which $k/SH > 1$. As inflation proceeds, the physical scale of a perturbation grows (k/SH decreases) until $k/SH \simeq 1$, and then onward until $k/SH \ll 1$. Once the physical scale of the perturbation grows such that $k/SH < 1$, microphysics cannot act coherently on that scale and alter the amplitude of the physical perturbation. When the Universe reheats and FRW behavior resumes, the Hubble radius (H^{-1}) increases ($\propto t$); k/SH begins to grow until $k/SH \simeq 1$ once again (now in the FRW phase) and the scale of the perturbation is said to reenter the horizon. The evolution during the period when $k/SH \ll 1$ is essentially kinematic in character, given by the classical evolution equations for energy-density perturbations in an expanding universe.

Clearly, the calculation of the perturbation spectrum is divided into two parts: First, the effects of microphysics must be estimated to determine the amplitude of the perturbations (characterized by comoving wave number k) generated while $k/SH > 1$ up to the time when dynamics "freezes out" ($k/SH \simeq 1$) during the inflationary phase. The amplitude of the perturbation when the dynamics freezes out depends upon the details of the particle-physics model; we will consider two different models in Sec. III. Second, we must trace the evolution of the perturbation in the kinematic regime from the time when $k/SH \simeq 1$ in the inflationary epoch through the time when the perturbation reenters the horizon ($k/SH \simeq 1$) in the FRW phase. Although the final physical results do not differ from one gauge to another, the identification of the relevant perturbation amplitude and the description of the evolution of the perturbation is gauge dependent when $k/SH \ll 1$.

The amplitude of perturbations at the time when the dynamics freezes out in the de Sitter phase is expected to be only weakly dependent on k. The microphysics that acts when the physical wave number (k/S) is greater than H is essentially independent of k since during the exponential expansion ϕ, $\dot{\phi}$, and H are nearly constant.[9] The only property that distinguishes the history of one scale (characterized by comoving wave number k) from another is the time at which $k/SH \simeq 1$ in the inflationary phase and dynamics freezes out. Different scales reach $k/SH \simeq 1$ at different times but with essentially the same perturbation amplitude. As the Universe expands these scales grow; at reheating the physical sizes of these scales range from that of the entire fluctuation region down to the Hubble radius ($\simeq H^{-1}$), depending upon when $k/SH \simeq 1$. A comoving scale corresponding to the present observable universe had $k/SH \simeq 1$ about 57 e-foldings before reheating and the scale corresponding to the size of a galaxy had $k/SH \simeq 1$ about 48 e-foldings before reheating (assuming reheating to a typical GUT scale $\sim 10^{14}$ GeV). Although the freeze-out time does differ from scale to scale, the time depends only logarithmically on scale since $S \sim \exp(Ht)$.

Given the amplitude of the perturbation when $k/SH \sim 1$ in the inflationary epoch (which we have argued is nearly independent of k), the evolution of the perturbation can be traced through the period when $k/SH \ll 1$ (in the late inflationary epoch and early FRW epoch) via the classical evolution equations. Independent of gauge, the evolution of the perturbation when $k/SH \ll 1$ can be shown to be roughly scale independent; the basic reason is that nothing can alter the amplitude of the real physical perturbation when its scale is large compared to the Hubble radius $O(H^{-1})$ (assuming microphysics is due to local causal effects and there are no nonadiabatic stresses on scales much larger than H^{-1}). For some choices of gauge, e.g., the comoving gauge, it may appear as if the perturbation (whose definition differs from gauge to gauge) grows significantly during reheating when $k/SH \ll 1$; this is a gauge artifact stemming from a choice of hypersurfaces of constant time which undergo great distortions during reheating. The final result of the analysis of the comoving gauge agrees identically with the result found for the uniform Hubble constant gauge in which the perturbation remains constant during reheating, but is produced initially with a larger amplitude.

The nearly scale-independent amplitude of a perturbation when it leaves the horizon ($k/SH \simeq 1$) during the de Sitter phase and the scale-independent evolution while $k/SH \lesssim 1$, result in a spectrum of perturbations which reenter the horizon with a nearly k-independent amplitude. The quantitative determination of the amplitude must be derived directly from the conditions which exist at the time when, in the inflationary epoch, $k/SH \simeq 1$ for a given scale.

Although the quantitative details which will follow are important, albeit rather involved, we believe that the qualitative explanation just presented is the underlying reason why each scale enters the horizon in the FRW phase with nearly the sample amplitude, $\delta\rho/\rho\,|_H$. This spectrum was first considered (in a different context) by Zel'dovich and Harrison.[11]

To summarize, the de Sitter phase plays a crucial role in this result in three ways: (1) it produces a "clean slate" in which any previously existing inhomogeneities disappear on all observable scales; (2) it provides a source—the event horizon—of new perturbations on small scales $O(H^{-1})$ which are then inflated to produce a spectrum of fluctuations on large scales; and (3) it is approximately time-translation invariant so that perturbations for any comoving scale are produced under approximately identical conditions and, as we have argued, evolve in an identical way into the FRW phase. Because of the special properties of the de Sitter phase it is possible to compute the spectrum of fluctuations for a cosmological model from first principles, with essentially no assumptions being made about the initial state of the Universe.

B. Notation and conventions

Although there is an extensive literature on the evolution of density perturbations, we have found it most convenient to employ the approach developed by Bardeen.[19] Some changes in notation and emphasis will be made to adapt Bardeen's approach to our problem, in which the absolute value of the sum of the energy density ρ_0 and the pressure p_0 is much less than ρ_0.

The results will be derived using the comoving gauge in Sec. III C and the uniform Hubble constant frame in Sec. III D; the latter gauge is particularly useful for demonstrating that the results are insensitive to the details of reheating in the inflationary-universe scenario (which is not obvious in other treatments of the problem[10]). In this subsection we will introduce the notation and conventions that will be used throughout.

All perturbations are measured with respect to an isotropic and homogeneous background, with the line element

$$ds^2 = -dt^2 + S^2(t)\delta_{ab}dx^a dx^b \ . \tag{2.1}$$

It would not be difficult to take into account any deviations from spatial flatness in the background, but in the inflationary scenarios we are exploring, this is negligible over the scale of our observable Universe. The proper expansion rate (Hubble constant) H is

$$H = S^{-1}dS/dt \equiv \dot{S}/S = \left[\frac{8\pi}{3}\rho_0/M_{\rm Pl}{}^2\right]^{1/2} \ . \tag{2.2}$$

We shall use units where $\hbar = c = k_B = 1$, so that $G^{-1/2} = M_{\rm Pl} = 1.2\times 10^{19}$ GeV. The relationship between the background pressure p_0 and energy density ρ_0 is described by the parameters

$$w \equiv p_0/\rho_0 \ , \tag{2.3a}$$

$$c_s{}^2 \equiv dp_0/d\rho_0 \ . \tag{2.3b}$$

Note that c_s is a purely formal "speed of sound"; no assumption is made about physics of energy-momentum tensor beyond what is enforced by the symmetry.

During the inflationary phase we can decompose ρ_0 into a contribution from the Higgs scalar field ϕ and a thermal radiation contribution ρ_r,

$$\rho_0 = V(\phi) + \frac{a}{2}\dot{\phi}^2 + \rho_r \ . \tag{2.4}$$

(Note the absence of spatial gradient terms due to the homogeneity of the background in the inflationary scenario.) Contributions to ρ_0 from zero-point fluctuations in other fields can be included in $V(\phi)$, so $V(\phi)$ becomes something like a finite temperature effective potential with a constant temperature the order of the Hawking temperature. Any primordial thermal radiation present at the beginning of inflation is rather quickly red-shifted away and will be ignored in discussing the perturbations. However, dissipation of dynamical oscillations in the Higgs field through quantum particle creation, etc.,[9] will produce eventually a new thermal component and it is this which is being described by ρ_r. The background pressure is

$$p_0 = -V(\phi) + \frac{a}{2}\dot{\phi}^2 + p_r \ . \tag{2.5}$$

Eventually we expect that $p_r = \frac{1}{3}\rho_r$, but during the reheating process we need not assume any particular equation of state.

Through the inflationary epoch the dominant contribution to ρ_0 is

$$\rho_v \simeq V(0) = \text{constant} \ , \tag{2.6}$$

so $H \simeq$ constant, and

$$S \cong S_0 \exp(Ht) \ . \tag{2.7}$$

Note that

$$\rho_0 + p_0 = a\dot{\phi}^2 + \rho_r + p_r \ , \tag{2.8}$$

and not zero as it would be in a spacetime which is exactly de Sitter.

The classical field ϕ is really the vacuum expectation value of the quantum scalar field responsible for SSB breaking, and a is a group-theoretic factor [e.g., for a model in which an adjoint $\underline{24}$ of Higgs breaks SU(5) to SU(3)×SU(2)×U(1), $a = \frac{15}{2}$]. The evolution of ϕ and ρ_r is given by

$$a\ddot{\phi}+3aH\dot{\phi}+V'(\phi)+\delta/\dot{\phi}=0\ , \tag{2.9}$$

where $V'(\phi)=\partial V/\partial\phi$, and

$$\dot{\rho}_r+3H(\rho_r+p_r)-\delta=0\ . \tag{2.10}$$

The δ terms in Eqs. (2.9) and (2.10) account for the creation of particles due to the time variation of ϕ. During the inflationary epoch, the $\ddot{\phi}$ and $\delta/\dot{\phi}$ terms in the evolution equations for ϕ are negligible and to a good approximation

$$3aH\dot{\phi}=-V'(\phi); \tag{2.11}$$

ρ_r is negligible, as noted above

Once reheating is completed the Universe enters a standard radiation-dominated FRW phase, with

$$p_0=\tfrac{1}{3}\rho_0=\tfrac{1}{3}\rho_r\ ,\ \ \rho_0\sim S^{-4},\ H\sim S^{-2}\sim t^{-1}\ . \tag{2.12}$$

A perturbation is characterized by its comoving (coordinate) wave number k or its corresponding comoving wavelength $\lambda=2\pi k^{-1}$, which remain constant as the Universe expands. The physical wave number is k/S and does change as the Universe expands. The spatial dependence of a given k component of the density perturbation is given by a solution of the scalar Helmholtz equation.

While the variables describing the perturbation can be made mathematically gauge invariant, there is a fundamental physical ambiguity in the interpretation of perturbations whose physical wave number k/S is smaller than the expansion rate H (wavelength larger than the Hubble radius). Perturbations in physical quantities such as the energy density which are nonzero and time dependent in the background are *hypersurface dependent* even though the physical quantity may be a frame-independent scalar in the physical spacetime. The question is which constant-time hypersurface in the physical spacetime is to be identified with the constant-time hypersurface in the background spacetime.

Two choices of hypersurface are particularly convenient for different aspects of the problem at hand. Before there is significant reheating only the scalar field terms contribute to the energy-momentum tensor. The dynamics of the perturbations are greatly simplified by taking *comoving* hypersurfaces on which, by definition, the energy flux (momentum density) T_i^0 vanishes. Since the energy flux of the scalar field is proportional to $\dot{\phi}\phi_{,i}$, in the absence of any other significant contribution to the energy-momentum tensor, the scalar field is homogeneous on comoving hypersurfaces, i.e., the perturbation in the scalar field vanishes. These hypersurfaces must be distorted geometrically, since the geometry of the hypersurfaces is what carries information about the dynamics of the scalar field perturbations.

The natural geometric choice of hypersurfaces is to require that the extrinsic curvature scalar K, the rate of convergence of the hypersurface normals, be the same everywhere on a given constant-t hypersurface. Since the average rate of expansion (Hubble constant) is just $H=-\frac{1}{3}K$, we call these hypersurfaces *uniform Hubble constant* or *uniform expansion* hypersurfaces. In a pure de Sitter space there is no unique foliation in these hypersurfaces; there are an infinite number of uniform expansion hypersurfaces through every point corresponding to all possible directions of a unit timelike vector. However, the background is not exactly de Sitter and in the real background the unique uniform expansion hypersurfaces coincide with the comoving hypersurfaces. In the presence of perturbations the uniform expansion and comoving hypersurfaces differ, so, e.g., the amplitude of the fractional energy-density perturbation on uniform expansion hypersurfaces is not the same as in the comoving gauge.

C. The view from the comoving gauge

We will denote the perturbation amplitudes on comoving hypersurfaces by a subscript c (instead of the m used in Ref. 19). The perturbed metric tensor can be read off of

$$ds^2=-(1+2\alpha_c Q)dt^2 +S^2[(1+2h_c Q)\delta_{ij}+2f_c Q_{,ij}]dx^i dx^j\ , \tag{2.13}$$

where $Q(x_i)$ is a spatial harmonic with coordinate wave number k, and where we choose to propagate the spatial coordinates normal to the hypersurface. The amplitudes α_c, h_c, and

$$\sigma_c\equiv\dot{f}_c \tag{2.14}$$

measure properties of the hypersurface independent of the choice of spatial coordinates. The perturbation in the intrinsic curvature scalar for the hypersurface is

$${}^3R=4\left[\frac{k}{S}\right]^2 h_c(t)Q(x^i) \tag{2.15}$$

and σ_c is the amplitude of the shear of the world lines normal to the comoving hypersurface. The coefficient α_c measures the perturbation in the ratio of proper time to coordinate time intervals along normals to the comoving hypersurface.

The fractional energy-density perturbation is described by ϵ_c, where $(\delta\rho/\rho_0)\equiv Q\epsilon_c$. As long as the scalar field is the only significant contribution to the energy-momentum tensor, the vanishing of $\Delta\phi$ on comoving hypersurfaces means that the only contribution to $\delta\rho$ comes from the kinetic term, due to the difference between proper time and coordinate time, and

$$Q\epsilon_c=\delta\rho/\rho_0=\delta(a\dot{\phi}^2/2)/\rho_0=-(a\dot{\phi}^2/\rho_0)Q\alpha_c\ . \tag{2.16}$$

The only perturbation in the stress tensor is a similar contribution to the isotropic pressure perturbation,

$$\delta p/\rho_0\equiv\pi_c Q=-(a\dot{\phi}^2/\rho_0)\alpha_c Q\ . \tag{2.17}$$

The "nonadiabatic" pressure perturbation

$$\delta p/\rho_0-(\delta\rho/\rho_0)(dp_0/d\rho_0)\equiv\eta Q \tag{2.18}$$

plays an important role in the perturbation equations since to first order it is independent of the hypersurface condition and plays the formal role of a source term in the equation for the evolution of ϵ_c. Physically, η measures the difference between the actual pressure perturbation and that expected from the energy-density perturbation and the background equation of state. [Note that the η defined by Eq. (2.18) is $w=p_0/\rho_0$ times the η defined in Ref. 19.] Equations (2.16)–(2.18) can be combined to give

$$\eta=(1-c_s^2)\epsilon_c \; . \tag{2.19}$$

The equation for the dynamic evolution of the density perturbations can be derived straightforwardly from Einstein's equations and for the case of comoving hypersurfaces can be written as [See Eq. (4.9) of Ref. 19]

$$\ddot{Z}-(\gamma-1)H\dot{Z}-[\gamma+3(1+w)]H^2Z +(k^2c_s^2/S^2)Z=-H^2\eta \; , \tag{2.20}$$

where

$$Z\equiv\left[\frac{HS}{k}\right]^2\epsilon_c \; , \tag{2.21}$$

$$\gamma\equiv-3(1+c_s^2)=H^{-1}d[\ln(\rho_0+p_0)]/dt \; . \tag{2.22}$$

We have assumed the anisotropic stress (shear stress) vanishes. To first order in the perturbations this follows if the energy-momentum tensor is the Higgs-field energy-momentum tensor plus a perfect-fluid energy-momentum tensor, as in the reheating scenario of Albrecht *et al.*[9] As was argued in Sec. II A microphysics (quantum fluctuations, etc.) can affect perturbations on a comoving scale only when $k/SH>1$. This is formally realized in Eq. (2.20) through contributions to the nonadiabatic stress (η) or anisotropic stress (π_T) amplitudes, the source terms which are formally $O(k^2/S^2H^2)$ and $O(k^2\epsilon_c/S^2H^2)$ smaller than the terms on the left-hand side of Eq. (2.20), respectively. When Eq. (2.19) is used to determine η, the evolution equations are purely classical and only describe the evolution of quantum fluctuations after they are generated.

In the comoving gauge the relevant perturbation amplitude is Z, and its evolution is given by Eq. (2.20). The main advantage of Eq. (2.20) is that it is a single differential equation which describes the complete evolution of the perturbation. Assuming efficient reheating, Eq. (2.20) can be integrated from the time the perturbation leaves the horizon ($k/SH\simeq1$) in the de Sitter epoch, through reheating, until the perturbation reenters the horizon in the subsequent FRW phase. (In Sec. II D we will switch to uniform Hubble expansion hypersurfaces which allow us to also consider the case of inefficient reheating.)

From Eq. (2.20) it follows that, with the exception of the few Hubble times before and after reheating, Z remains roughly constant from the time $k/SH\simeq1$ in the de Sitter phase until the perturbation reenters the horizon during the FRW phase which follows. During reheating Z grows by a factor of $O(M_G^4/\dot{\phi}^2)$. This occurs because as ϕ speeds up and begins to rapidly approach σ,

$$c_s^2=-1-(\tfrac{1}{3})H^{-1}d\ln(\rho_0+p_0)/dt$$
$$\simeq-1-2\ddot{\phi}/(3H\dot{\phi})$$

becomes less than -1 and $\gamma\equiv-3(1+c_s^2)$ becomes greater than zero. For $\gamma>0$ the "growing mode" solution for Z increases with time. Near $\phi=\sigma$, $c_s^2\simeq-2\sigma/3H$, and $\gamma\simeq2\sigma/H\gg1$. As ϕ oscillates around $\phi=\sigma$, c_s^2 oscillates between $\pm O(\sigma/H)$, settling down to $\frac{1}{3}$ as the vacuum energy is converted to radiation. In exact de Sitter space $c_s^2=-1$, $\gamma=0$, and $Z=$constant is the growing mode solution to Eq. (2.20). The fact that the inflationary phase is not exactly de Sitter ($c_s^2<-1$, $\gamma>0$), is why Z grows during the inflationary phase—de Sitter space is marginally stable with respect to the growth of perturbations. Geometrically, the growth of Z during reheating can be understood because complex distortions of the comoving hypersurfaces are necessary to keep the energy flux zero as the coherent energy of the ϕ field is converted to radiation. A detailed discussion of the solution of Eq. (2.20) now follows.

When $k/SH\sim1$ in the de Sitter phase for a given perturbation, the only significant contribution to the energy-momentum tensor is that of the Higgs scalar field ϕ; Eqs. (2.19) and (2.20) may be combined to give

$$\ddot{Z}-(\gamma-1)H\dot{Z}-[\gamma+3(1+w)]H^2Z+(k^2/S^2)Z=0 \; . \tag{2.23}$$

Because the potential energy density $V(\phi)$ dominates the kinetic energy density $a\dot{\phi}^2/2$,

$$w=p_0/\rho_0=-1+(a\dot{\phi}^2+\rho_r+p_r)/\rho_0\simeq-1 \; . \tag{2.24}$$

Also, the coefficient γ as defined by Eq. (2.22) will be rather small and vary slowly relative to the background expansion rate, since during most of the long inflation epoch the evolution of the Higgs field ϕ is very slow. For a few Hubble times when $k/SH\sim1$ in the de Sitter phase we can approximate (2.23) by

$$\ddot{Z}+H\dot{Z}+(k^2/S^2)Z=0 \; , \tag{2.25}$$

with H approximately constant and $S\sim\exp(Ht)$. A WKBJ solution of Eq. (2.25) when $k/SH>1$ is

$$Z\simeq Z_0[1-\tfrac{1}{4}(HS/k)^2]^{-1/4} \times\cos\left\{\int^t dt'(k/S)[1-\tfrac{1}{4}(HS/k)^2]^{1/2}\right\} . \tag{2.26}$$

The quantity Z oscillates with nearly constant amplitude and an effective propagation speed equal to the speed of light, even though $c_s^2\cong-1$. As k/SH becomes less than one the $(k^2/S^2)Z$ term in Eq. (2.25) quickly becomes negligible; depending on the phase of the wave at freeze-out, Z will approach a roughly constant value Z_1 somewhere between $-Z_0$ and $+Z_0$, where $|Z_1|\simeq Z_0\simeq\epsilon_c$ when $k/SH\sim1$.

Since γ is not exactly zero and becomes large and positive during the last few Hubble times before reheating, one cannot neglect γ in Eq. (2.23). However, by the time the effects of a nonzero γ become appreciable, $k/SH\ll1$, and to a good approximation the $(k^2/S^2)Z$ term can be neglected;

$$\ddot{Z}-(\gamma-1)H\dot{Z}-\gamma H^2Z\simeq0 \; . \tag{2.27}$$

With $E(t)\equiv\rho_0+p_0$, an exact solution of Eq. (2.27), which has $Z\simeq Z_1\simeq\epsilon_c$ at $t=t_1$ (the time $k/SH\sim1$), is

$$Z(t)=Z_1\int_{t_1}^{t}H\exp[H(t'-t)]E(t')/E(t_1)dt' +Z_1\exp[-H(t-t_1)] \; . \tag{2.28}$$

Note that

$$\dot{Z}=Z_1HE(t)/E(t_1)-HZ \tag{2.29}$$

vanishes at $t=t_1$, and as long as $E(t)$ varies slowly on a Hubble time scale,

$$Z(t)\simeq Z_1\langle E(t)\rangle/E(t_1)\ , \tag{2.30}$$

where $\langle E(t)\rangle$ is an average of $E(t)$ over the last Hubble time or so.

The solution given by Eqs. (2.28) and (2.29) remains valid until reheating, $t=t_*$, when ϕ increases to $\phi\sim\sigma\sim M_G$. The characteristic dynamical time scale for the Higgs field ϕ is then $\sigma^{-1}\simeq M_G{}^{-1}\ll H(t_*)^{-1}\simeq M_{\rm Pl}/M_G{}^2$. The conversion of the coherent energy of the Higgs field to radiation by decay of the Higgs-field oscillations and thermalization of the decay products is expected to take place on this dynamical time scale (see Albrecht *et al.*[9]) In a time $(\Delta t)_{RH}\ll H(t_*)^{-1}$ the energy-momentum tensor transforms from being dominated by the old vacuum energy density $\rho_v=V(0)$, to being dominated by incoherent radiation with $p_r=\frac{1}{3}\rho_r$. During $(\Delta t)_{RH}$ the total energy density does not change appreciably, so at the beginning of the radiation dominated era $\rho_r\cong\rho_v$.

With $(\Delta t)_{RH}\ll H(t_*)^{-1}$ the comoving gauge solution for Z can easily be continued right through reheating. Once the radiation becomes appreciable the nonadiabatic stress amplitude η is no longer given by Eq. (2.19), but it should never be substantially larger. Because $k/SH\,|_{t=t_*}\ll 1$ the η term in Eq. (2.20) can still be neglected. The other assumption made in approximating Eq. (2.20) by Eq. (2.27), that $w\simeq -1$, is not valid during $(\Delta t)_{RH}$, but it is clear from Eq. (2.23) that a change in w of order unity can have an effect on Z only after a time the order of $H(t_*)^{-1}$. With $(\Delta t)_{RH}\ll H(t_*)^{-1}$ our approximate solution embodied in Eqs. (2.28)–(2.30) remains valid through reheating. This gives at the beginning of the radiation era

$$Z(t_*)\cong[\tfrac{4}{3}Z_1\rho_v/E(t_1)]O(H/M_G)\ , \tag{2.31a}$$

after averaging over the previous Hubble time and

$$\dot Z(t_*)\cong\tfrac{4}{3}Z_1H_*\rho_v/E(t_1)[1-O(H/M_G)]\ , \tag{2.31b}$$

where $H_*\equiv H(t_*)$.

If the background remains radiation dominated, $p=\frac{1}{3}\rho$, there is a well-known analytic solution of Eq. (2.20) valid for $t>t_*$. This is

$$Z(t)=a_1(x^{-3}\sin x-x^{-2}\cos x)+a_2(x^{-3}\cos x+x^{-2}\sin x)\ , \tag{2.32}$$

where

$$x=\int^t c_s kS^{-1}dt'=k/(\sqrt{3}SH)=[1+2H_*(t-t_*)]^{1/2}k/(\sqrt{3}S_*H_*)\ . \tag{2.33}$$

The coefficients a_1 and a_2 in Eq. (2.32) are determined by matching to the above initial data at $t=t_*$. With $k/S_*H_*\ll 1$, $x_*\equiv x(t_*)\ll 1$, and

$$Z(t_*)\cong\tfrac{1}{3}a_1+x_*{}^{-3}a_2\ , \tag{2.34a}$$

$$\dot Z(t_*)\cong-\tfrac{1}{15}x_*{}^2H_*a_1-3x_*{}^{-3}H_*a_2\ . \tag{2.34b}$$

Since $|\dot Z(t_*)|\gg|HZ(t_*)|$,

$$a_2\cong-\tfrac{4}{9}x_*{}^3Z_1\rho_v/E(t_1)\ , \tag{2.35a}$$

$$a_1\cong-3x_*{}^{-3}a_2=\tfrac{4}{3}Z_1\rho_v/E(t_1)\ . \tag{2.35b}$$

By the time $S/S(t_*)\gg 1$, the contribution of the "decaying mode" to Z is negligible, and using Eq. (2.33),

$$\epsilon_c\cong 4Z_1[\rho_v/E(t_1)](x^{-1}\sin x-\cos x)\ . \tag{2.36}$$

If the background is still radiation dominated when $k/SH\sim 1$, then the amplitude of the sound wave which the perturbation becomes once it is well within the Hubble radius (reenters the horizon) is $(x\gg 1)$

$$\delta\rho/\rho_0\,|_H=4Z_1\rho_v/E(t_1) \tag{2.37a}$$

$$=4\Delta\dot\phi/\dot\phi(t_1)\ . \tag{2.37b}$$

Expression (2.37b) follows since in the comoving gauge $\phi=$constant, and

$$Z_1=\epsilon_c(t_1)=\Delta(a\dot\phi^2/2)\,|_{t_1}/\rho_0\simeq(a\dot\phi\Delta\dot\phi)\,|_{t_1}/\rho_0\ .$$

We would like to relate $\Delta\dot\phi$ to $\Delta\phi$, the fluctuation in ϕ measured in the uniform Hubble expansion gauge (a frame in which the geometry is "smooth," since the $t=$constant hypersurfaces have not been distorted to conform to $\phi=$constant). This is straightforward to do because when $k/SH\lesssim 1$, $\epsilon_c\simeq\epsilon_u/3$. [$\epsilon_u$ is the fractional energy-density perturbation measured in the uniform Hubble constant gauge; $\epsilon_c\simeq\epsilon_u/3$ follows from Eqs. (2.38) and (2.43) in Sec. III D.] Since $\epsilon_u\simeq V'\Delta\phi/\rho_0\simeq-3aH\dot\phi\Delta\phi/\rho_0$ (recall during inflation $V'\simeq-3aH\dot\phi$), it follows that $\Delta\dot\phi\simeq H\Delta\phi$. Thus, Eq. (2.37b) can be written in terms of $\Delta\phi$ as

$$\delta\rho/\rho_0\,|_H=4H\Delta\phi/\dot\phi(t_1)\ . \tag{2.37c}$$

How these waves eventually evolve into condensations on various scales from galaxies to clusters of galaxies, and produce anisotropy in the microwave background, has been treated extensively elsewhere and will not be reiterated in this paper.

On large enough scales ($M\gtrsim 10^{15}$–$10^{16}M_\odot$) the Universe will change to being matter dominated before $k/SH\sim 1$. This modifies the behavior of Z, but in a well-understood way. The fractional-density perturbation is of the order given by Eq. (2.37) when $k/SH\sim 1$ (the factor of 4 is replaced by $\frac{2}{5}$), and thereafter grows linearly in S.

It may appear puzzling, in light of the discussion in Sec. II A, that Z grows significantly during reheating even through $k/SH\ll 1$. This is an example of how the description of the evolution of the perturbation up to the point where $k/SH\simeq 1$ in the FRW regime can be gauge dependent. In the comoving gauge, we trace the evolution of the quantity Z whose value when $k/SH\simeq 1$ in the de Sitter phase is given by the fractional perturbation in the total energy density. The formal quantity Z grows dramatically during reheating so that by the time $k/SH\simeq 1$ during the FRW phase, it is a factor of $O(M_G{}^4/\dot\phi^2)$ larger than when $k/SH=1$ in the de Sitter phase. By contrast in the uniform Hubble constant gauge, as we shall see, we trace a formal quantity whose value when $k/SH\simeq 1$ in the de Sitter phase is given by the fractional perturbation in the nonvacuum part of the energy density only. This quantity does not change during reheating and remains more or less constant up to the time when $k/SH \simeq 1$ in the FRW regime. At that time, it is meaningful to compare the formal quantities in the two gauges and identify them as the physical perturbation amplitude. At this point the amplitudes found in the two gauges must and do agree.

D. The view from the uniform Hubble constant gauge

We will denote the perturbation amplitudes on uniform Hubble constant hypersurfaces by the subscript u. For example, the amplitude of the fractional-energy-density perturbation is denoted by ϵ_u. The metric-perturbation amplitudes relative to uniform expansion hypersurfaces are denoted by α_u, h_u, and f_u, and are defined by the corresponding version of Eq. (2.13). Also, let $\sigma_u = \dot{f}_u$ measure the shear of the uniform-expansion hypersurface normals. Note that h_u is a measure of the distortion of the intrinsic geometry of the uniform-expansion hypersurfaces, just as h_c is for the comoving hypersurfaces via Eq. (2.15).

The perturbation equations written in terms of these new variables are Eqs. (6.24)—(6.26) of Ref. 19. With appropriate changes of notation these are

$$\epsilon_u = \frac{2}{3}\left[\frac{k}{SH}\right]^2 h_u , \tag{2.38}$$

$$[k^2 + 12\pi G(\rho_0 + p_0)S^2]\alpha_u = -k^2(1+3c_s{}^2)h_u - \tfrac{9}{2}H^2S^2\eta , \tag{2.39}$$

$$(\dot{h}_u - H\alpha_u)\dot{} + 3H(\dot{h}_u - H\alpha_u) - \tfrac{1}{3}\left[\frac{k}{S}\right]^2 (h_u + \alpha_u) = H^2\pi_T . \tag{2.40}$$

Here π_T is a measure of the anisotropic stress,

$$S^{-2}T_{ij} - \tfrac{1}{3}\delta_{ij}T_{kk} = \rho_0\pi_T(k^{-2}Q_{,ij} + \tfrac{1}{3}\delta_{ij}Q) , \tag{2.41}$$

which vanishes except possibly during reheating. Remember the amplitude η of the nonadiabatic pressure perturbation is to first-order hypersurface independent, so η in Eq. (2.39) is the same as in Eq. (2.20). We could substitute Eq. (2.39) into Eq. (2.40) to get a single second-order equation for h_u given that η is a known function of time. Actually, η is not a known function of time, but at least before reheating is related to ϵ_c by Eq. (2.19). Still, we will see that in the limit $k/SH \ll 1$, Eq. (2.40) has a very simple approximate solution which holds all the way through reheating.

Equations (2.39)—(2.40) and (2.20) are mathematically equivalent. The exact transformation between the variables is generated by a displacement of the comoving hypersurfaces relative to the uniform-expansion hypersurfaces by an amount Δt, with

$$H\Delta t = \frac{\rho_0}{\rho_0 + p_0}\,\frac{Z + H^{-1}\dot{Z}}{1 + 12\pi G(\rho_0 + p_0)S^2/k^2} . \tag{2.42}$$

This gives

$$h_u = h_c - H\Delta t = \tfrac{3}{2}Z + \frac{9}{2}\left[\frac{HS}{k}\right]^2 \frac{Z + H^{-1}\dot{Z}}{1 + 12\pi G(\rho_0 + p_0)S^2/k^2} . \tag{2.43}$$

An exact solution of Eq. (2.40) before reheating is obtained by substituting

$$H^2S^2\eta = k^2(1 - c_s{}^2)Z \tag{2.44}$$

in Eq. (2.39), and by solving Eq. (2.43) simultaneously with Eq. (2.40). This is obviously more complicated than solving Eq. (2.20) with the same substitution for η and represents an important advantage for the comoving-gauge approach.

The difficulty with the comoving-gauge derivation comes when the scalar field ϕ begins to oscillate about the true vacuum minimum at $\phi = \sigma$ and reheating takes place. Once there is an appreciable thermal contribution to the energy-momentum tensor, the perturbation in ϕ no longer vanishes on comoving hypersurfaces and to find η one must solve the perturbed evolution equation for ϕ. Also, $c_s{}^2$ becomes very large and negative, and may fluctuate wildly during the reheating process. We will see that one can argue the η term is unimportant in either Eqs. (2.20) or (2.39), but the uniform expansion hypersurface approach isolates the growing mode of density perturbation more cleanly and allows a more straightforward handling of the variations in $c_s{}^2$.

In the uniform Hubble constant gauge the evolution during the period when $k/SH \ll 1$ is quite simple because the amplitude

$$\zeta \equiv h_u\{1 + k^2/[12\pi G(\rho_0 + p_0)S^2]\} \tag{2.45}$$

is nearly time independent throughout the period when $k/SH \ll 1$, as can be seen from Eqs. (2.39)—(2.40). From Eqs. (2.38) and (2.2) it follows that

$$\zeta = \tfrac{3}{2}\epsilon_u\left[(HS/k)^2 + \frac{2}{9}\frac{\rho_0}{\rho_0 + p_0}\right] \tag{2.46}$$

when $k/SH \ll 1$. During inflation $\rho_0/(\rho_0 + p_0) \gg 1$ so that $\zeta \sim (\delta\rho)_u/(\rho_0 + p_0)$ when $k/SH \sim 1$. Thus, ζ is a measure of the fractional perturbation in the nonvacuum part of the energy density. The vacuum stress tensor has $\rho_v = -p_v =$ constant and contributes to neither $(\delta\rho)_u$ nor $\rho_0 + p_0$. When the perturbation reenters the horizon $(k/SH \sim 1)$ in the FRW era, $\rho_0/(\rho_0 + p_0) \sim 1$ and Eq. (2.46) implies $\zeta \sim \epsilon_u$. In a sense, there is no change in amplitude of the perturbation while $k/SH \ll 1$.

Recall, in the comoving gauge the amplitude Z is given by $Z = (HS/k)^2\epsilon_c$. When $HS/k \simeq 1$, Z represents the fractional perturbation in the *total* energy density. In contrast to ζ, Z does not remain constant while the scale in question is outside the horizon.

The argument for ζ being constant is as follows. The last term on the left-hand side of Eq. (2.40) is too small to have any appreciable effect on the solution of Eqs. (2.39) and (2.40) while $k/SH \ll 1$, since it is intrinsically of order $(k/SH)^2$ times the first two terms. Assuming the anisotropic stress (π_T) is negligible, Eq. (2.40) says that there are two independent modes, a decaying mode $\dot{h}_u - H\alpha_u \sim \exp(-3Ht)$ which rapidly becomes negligible and a "constant mode" with

$$\dot{h}_u = H\alpha_u = -\frac{(\dot{S}/S)[(1+3c_s{}^2)h_u + \frac{9}{2}(HS/k)^2\eta]}{1 + 12\pi G(\rho_0 + p_0)S^2/k^2} . \tag{2.47}$$

The nonadiabatic stress amplitude η has a negligible effect in Eq. (2.47) if

$$\eta \ll (k/SH)^2 h_u \sim \epsilon_u \; . \tag{2.48}$$

From Eq. (2.43),

$$(k/SH)^2 h_u \simeq Z/[1+\tfrac{9}{2}(HS/k)^2(\rho_0+p_0)/\rho_0] \; ,$$

while at least before reheating $\eta \sim \epsilon_c \sim (k/SH)^2 Z$. As long as $(\rho_0+p_0)/\rho_0 \ll 1$, the inequality (2.48) is satisfied. Once a substantial amount of vacuum energy is converted into radiation during reheating, Eq. (2.48) breaks down, but by then

$$[1+12\pi G(\rho_0+p_0)S^2/k^2] \sim (HS/k)^2 \gg 1$$

and the whole right-hand side of Eq. (2.47) is negligibly small.

A heuristic physical argument for the neglect of η and π_T is based on the fact that on the scales we are interested in, the decay of the false vacuum occurs while the perturbation wavelength is large compared to the Hubble radius. The initial fluctuations in ϕ imply fluctuations in the roll-down time for ϕ to reach $\phi \sim \sigma$, just because ϕ starts out further down the barrier in some regions than others. Spatial gradient terms are negligible and the vacuum decay process should proceed as the same function of local proper time in different regions. This means the energy density ρ and pressure p should be the same functions of proper time as in the background spacetime, except for an offset δt due to different starting conditions. Then

$$\delta\rho = \frac{d\rho_0}{dt}\delta t, \quad \delta p = \frac{dp_0}{dt}\delta t \; ,$$

and by Eq. (2.18), $\eta \sim 0$, $\delta p/\delta\rho = c_s{}^2$. This argument applies all the way through reheating, whether or not it is efficient.

The anisotropic stress amplitude π_T is zero at all times in the simple reheating model of Albrecht *et al.*,[9] since $\phi_{,i}\phi_{,j}$ is second order in the perturbation amplitude and the thermal stress tensor is assumed to be isotropic. The largest π_T could reasonably be is the ratio of some effective particle mean-free path to the perturbation wavelength times ϵ_u, but the effective particle mean-free path cannot be greater than a few times H^{-1}. Thus, we expect $\pi_T \ll \epsilon_u \ll h_u$, while $\pi_T \sim h_u$ is necessary to invalidate Eq. (2.47).

The *exact* solution of Eq. (2.47) once $\eta = 0$ is

$$\zeta = h_u\{1+k^2/[12\pi G(\rho_0+p_0)S^2]\} \equiv \zeta_f = \text{constant} \; , \tag{2.49}$$

since

$$1+3c_s{}^2 = -d\ln[(\rho_0+p_0)S^2]/d\ln S \; .$$

To find ζ_f we match to our approximate solution for Z [Eq. (2.28)] when both it and Eq. (2.49) are valid, i.e., when $k/SH \ll 1$ and $(\rho_0+p_0)/\rho_0 \ll 1$. In this regime Eq. (2.43) gives

$$h_u \cong \tfrac{9}{2}(HS/k)^2(Z+H^{-1}\dot{Z})[1+12\pi G(\rho_0+p_0)S^2/k^2]^{-1} \tag{2.50a}$$

or by Eq. (2.29),

$$\zeta = \zeta_f \cong [\rho_0/(\rho_0+p_0)][Z_1 E(t)/E(t_1)] \; . \tag{2.50b}$$

Since until reheating $\rho_0 \cong \rho_r = \text{constant}$,

$$\zeta \cong Z_1[\rho_0/(\rho_0+p_0)]_{t=t_1} \; . \tag{2.51}$$

After reheating and as long as $k/SH \ll 1$, $k^2/(12\pi G(\rho_0+p_0)S^2) \ll 1$ and Eq. (2.51) becomes

$$h_u \cong Z_1[\rho_0/(\rho_0+p_0)]_{t=t_1} \; . \tag{2.52}$$

No assumption about the background equation of state during or after reheating was involved in deriving Eqs. (2.50) or (2.51). Reheating may be very inefficient and take many expansion times. In order not to affect nucleosynthesis calculations, reheating and rethermalization should be complete at a temperature of at least 10 MeV or so. Of course, baryogenesis must work out properly, but this may not require rethermalization to a high temperature. In any case, one should be back to a conventional radiation-dominated FRW model well before k/SH increases to $O(1)$ on any scale relevant to galaxy formation or anisotropy of the microwave background.

The transition through $k/SH \sim 1$ is most easily handled by going back to Z (or ϵ_c) as a variable. With $p_0 = \frac{1}{3}\rho_0$ and $k/SH \ll 1$, Eq. (2.43) gives

$$\zeta \cong h_u \cong [\tfrac{3}{2}+\rho_0/(\rho_0+p_0)]Z = \tfrac{9}{4}Z \; . \tag{2.53}$$

Following through the matching of solutions as before, we find with greater generality that the amplitude of the sound wave at $k/SH \gg 1$ is given by Eq. (2.37).

If the Universe is matter dominated by the time $k/SH \sim 1$, the value of Z at $k/SH \ll 1$ from Eqs. (2.43) and (2.36) is

$$Z = Z_2 = \tfrac{2}{5}\zeta_f = \tfrac{2}{5}Z_1[\rho_0/(\rho_0+p_0)]_{t=t_1} \; . \tag{2.54}$$

The relevant solution of Eq. (2.20) with $p_0 = 0$ is just Z remaining constant even after $k/SH > 1$, and if S_2 is the scale factor when $k/SH = 1$,

$$\epsilon_c = \tfrac{2}{5}\zeta_f(S/S_2) \; . \tag{2.55}$$

The amplitude of the perturbations as they reenter the horizon (Hubble radius) in the FRW epoch is governed by the parameter ζ_f as defined by Eq. (2.50), either through Eqs. (2.37a) or (2.55). What are plausible values for ζ_f based on the microphysics of the Higgs field? From Eqs. (2.16) and (2.8),

$$\zeta_f = -\alpha_c\,|_{t=t_1} \; , \tag{2.56}$$

but a physically more natural and intuitive way of looking at the microphysics of the perturbations is from a frame of reference in which the geometrical perturbations are small when $k/SH \sim 1$. The uniform-expansion hypersurfaces provide such a frame of reference, since at $t = t_1$ and with $(\rho_0+p_0)/\rho_0 \ll 1$ Eqs. (2.39)–(2.43) imply $h_u \sim \alpha_u \sim Z_1 \ll \zeta_f$.

On uniform-expansion hypersurfaces the perturbation in the scalar-field expectation value $(\delta\phi)_u$ is found from

$$0 = (\delta\phi)_c = (\delta\phi)_u + \dot{\phi}\Delta t \; , \tag{2.57}$$

where the displacement Δt of comoving relative to uniform-expansion hypersurfaces is given by Eq. (2.42). Since Z is slowly varying at $t \sim t_1$,

$$\Delta t/Q \cong H^{-1}[\rho_0/(\rho_0+\rho_0)]Z \cong H^{-1}\zeta_f \ . \qquad (2.58a)$$

The amplitude corresponding to $(\delta\phi)_u$ is

$$(\Delta\phi)_u \equiv (\delta\phi)_u/Q \cong -H^{-1}\zeta_f\dot{\phi} \ , \qquad (2.58b)$$

so

$$\zeta_f = -H(\Delta\phi)_u/\dot{\phi}\,|_{t=t_1} \ , \qquad (2.58c)$$

which confirms Eq. (2.37b). Any gauge in which the metric perturbation amplitudes are small compared with ζ_f at $t=t_1$ will see the initial perturbation as a scalar-field fluctuation, with $\Delta\phi \cong (\Delta\phi)_u$, in an appropriate de Sitter background. How rapidly the metric perturbations grow and whether they become comparable with ζ_f before or during reheating depends on the gauge. In the uniform-expansion gauge the spatial metric perturbations grow as S^2 once $k/SH<1$, while in a synchronous gauge[10] the metric perturbations stay small until reheating, when they suddenly increase to order ζ_f. All the usual gauges describe the perturbation as being predominantly a "curvature perturbation" with roughly constant amplitude after reheating up until the time $k/SH \sim 1$ again.

III. RESULTS IN TWO SPECIFIC MODELS OF NEW INFLATION

In the previous section we determined the spectrum of the density perturbations which result from quantum fluctuations in the scalar field responsible for SSB in the generic model of new inflation. The basic result is that when the physical length scale of a density perturbation equals the Hubble radius during the FRW epoch that follows inflation $\delta\rho/\rho\,|_H = (4 \text{ or } \frac{2}{5})H\Delta\phi/\dot{\phi}(t_1)$, where t_1 is the time when that scale expanded beyond the Hubble radius during the de Sitter epoch and 4 ($\frac{2}{5}$) applies if the Universe is radiation (matter) dominated when the scale reenters the horizon. As we discussed in Sec. II B, the spectrum should be nearly scale independent in all models of new inflation; however, the amplitude depends upon $\Delta\phi$ and $\dot{\phi}(t_1)$ which, of course, depend upon the shape of the scalar potential. In this section we will consider two very different models: (1) an SU(5) GUT with Coleman-Weinberg SSB; (2) a supersymmetric GUT with a potential of the O'Raifeartaigh-Witten type.

A. Coleman-Weinberg SU(5) GUT

We will first evaluate the amplitude of the density perturbation spectrum for the simplest model of new inflation, the SU(5) model with Coleman-Weinberg SSB. The behavior of the fluctuations of ϕ in an inflationary universe based on Coleman-Weinberg SSB has been considered by several authors but probably the most complete analysis has been provided in a recent paper by Linde.[17] For Coleman-Weinberg SSB with the effective mass of the scalar field (including the effects of the coupling of the scalar field to the curvature) set to zero, Linde finds that the quantum zero-point fluctuations at a time Δt after the beginning of the de Sitter phase are given by

$$\langle a\phi^2\rangle = \frac{H^3}{4\pi^2}\Delta t \ , \qquad (3.1)$$

where Δt is the "clock" time elapsed since the spinodal domains appeared, at a physical temperature $T \lesssim H$. The original thermal contribution to $\langle a\phi^2\rangle$ can be ignored compared with the zero-point fluctuations after several Hubble times. The physics of Eq. (3.1) is that fluctuations in ϕ are generated on each comoving scale, represented by the comoving wave number k, when $k/SH \sim 1$ in the de Sitter phase, starting when the temperature falls below the Hawking temperature say, at $t \simeq t_0$. This is when the conformal noninvariance of the minimally coupled Higgs-field equation first makes itself felt. Once generated, and expanded so that $k/SH \lesssim 1$, the fluctuations on a given scale are frozen in place. As smaller and smaller scales (larger and larger k) pass through $k/SH \sim 1$, the total contribution to $\langle a\phi^2\rangle$ steadily increases. The contribution to $\langle a\phi^2\rangle$ from one e-folding of k is produced in one Hubble time H^{-1}. The amplitude of the fluctuation $\Delta\phi$ associated with a given scale k is naturally taken from Eq. (3.1) to be

$$\Delta\phi = (H^2/4a\pi^2)^{1/2} = a^{-1/2}(H/2\pi) \ ; \qquad (3.2)$$

note that $H/2\pi$ is the Hawking temperature.

The cumulative effect of fluctuations on scales larger than the scale in which one is interested contributes to the background value of ϕ. In this context, the evolution of the background is not just the evolution of the uniform classical field governed by

$$a\ddot{\phi} + 3aH\dot{\phi} + \frac{\partial V}{\partial\phi} = 0 \ , \qquad (3.3)$$

but should contain a source term corresponding to the growth of ϕ due to quantum fluctuations. During the de Sitter phase after $t \sim t_0$ the $a\ddot{\phi}$ term in Eq. (3.3) is small compared with the other two terms, and the classical evolution of ϕ is

$$\dot{\phi} \simeq -\frac{1}{3aH}\frac{\partial V}{\partial\phi} \ . \qquad (3.4)$$

On the other hand, evolution by quantum fluctuations according to Eq. (3.1), whose derivation assumes $\partial V/\partial\phi \sim 0$, gives

$$\dot{\phi} \sim \frac{H^3}{8\pi^2}(a\phi)^{-1} \ . \qquad (3.5)$$

As ϕ increases, the solution including fluctuations should go from Eqs. (3.5) to (3.4). A modification to Eq. (3.3) whose solution has this property is

$$3aH\dot{\phi} + \frac{\partial V}{\partial\phi} = \frac{3}{8\pi^2}H^4\phi^{-1} \ . \qquad (3.6)$$

This makes sense as a background equation for ϕ only as long as $\ddot{\phi} \ll H\dot{\phi}$. Within about one Hubble time of the end of the de Sitter phase the $\ddot{\phi}$ term is important, but by this time the de Sitter fluctuation term is no longer significant. At this point one can go back to Eq. (3.3) with the appropriate terms added to model the reheating process [see Eq. (2.9) and Albrecht *et al.*[9]].

In the conventional SU(5) GUT model one can choose the parameters of the Higgs field so that in the de Sitter background the zero-temperature effective potential has the Coleman-Weinberg form

$$V(\phi) = V_0 - \tfrac{1}{4}\lambda(\phi)\phi^4 \ , \qquad (3.7)$$

where

$$\lambda(\phi)=8Bg^4\left[\ln\frac{g\sigma}{\max(H,g\,|\,\phi\,|\,)}+\frac{1}{4}\right], \tag{3.8}$$

$$V_0=\tfrac{1}{2}B(g\sigma)^4 . \tag{3.9}$$

Here g is the gauge coupling constant. The true vacuum is at $\phi=\sigma$, where $V(\phi)=0$, $\partial V/\partial\phi=0$. The coefficient $B=5625/(1024\pi^2)$ in a SU(5) model, which we will use for our calculations. We have ignored in Eqs. (3.8) and (3.9) the fact that the relevant true vacuum should be in a flat-spacetime background rather than a de Sitter background. The corresponding corrections to V_0 are negligible. Whether the choice of parameters which gives Eq. (3.7) is in any sense "natural" will not be argued here.[18] A mass term in Eq. (3.7) small enough to leave the inflationary scenario intact would not modify any of our subsequent results in a qualitatively important way. As long as $|\,\phi\,| \ll\sigma$ and $H \ll g\sigma$, λ is nearly constant, and

$$\frac{\partial V}{\partial\phi}\simeq-\lambda\phi^3 . \tag{3.10}$$

During the de Sitter phase both of these conditions are easily satisfied.

A classical treatment of the evolution of perturbations makes sense only when the evolution of the background is classical, i.e., when [see Eq. (3.6)]

$$\frac{3}{8\pi^2}H^4\phi^{-1}<\lambda\phi^3 . \tag{3.11}$$

The solution of Eq. (3.6) when the classical evolution is dominant is

$$(\phi/H)^2=(3a/2\lambda)[H(t_*-t)]^{-1} , \tag{3.12}$$

where as usual $t\simeq t_*$ corresponds to the time of reheating. Thus, inequality (3.11) is satisfied as long as

$$H(t_*-t)<a(6\pi^2/\lambda)^{1/2} . \tag{3.13}$$

Assuming efficient reheating [to a temperature $\simeq O(g\sigma)$], the time t_1 at which the scale corresponding to the present Hubble radius ($\simeq 10^{28}$ cm) leaves the horizon during the de Sitter phase $[k_U/S(t_1)H\simeq 1]$ satisfies

$$\Delta\beta\equiv H(t_*-t_1)\cong 57+\ln(g\sigma/10^{14}\text{ GeV}) . \tag{3.14}$$

In general, for the comoving scale k,

$$\Delta\beta=H(t_*-t_1)$$
$$\cong 57+\ln(k_U/k)+\ln(g\sigma/10^{14}\text{ GeV}) . \tag{3.15}$$

For the scale corresponding to a large galaxy ($\lambda\simeq 1$ Mpc$\simeq 3\times 10^{24}$ cm)

$$\Delta\beta\simeq 48+\ln(g\sigma/10^{14}\text{ GeV}) .$$

The exponential dependence of S on t during the de Sitter phase means that at $t=t_*$, k/SH is $\exp(-\Delta\beta)$. From Eq. (3.13) we see that for $\lambda<1$ and $a=\frac{15}{2}$, the classical evolution equation, Eq. (3.12), is appropriate for all observable scales. In the standard SU(5) GUT model $\lambda\cong 4$. With this value of λ, the amount of inflation is not sufficient to make a current Hubble volume scale satisfy $k/SH \gg 1$ at the time of freeze-out of thermal fluctuations, which is a fundamental requirement for the new inflation to work. This is the conclusion drawn in Ref. 16.

As we shall soon see, the requirement that the density perturbations produced from the quantum fluctuations in ϕ be consistent with the isotropy of the microwave background constrains λ to be $\ll 1$. So, for the sake of argument, let us assume λ is small and that ϕ evolves classically during the whole time that perturbations on observable scales are generated. From Eqs. (2.37), (3.2), (3.12), (3.14), and (3.15) we find that

$$\left.\frac{\delta\rho}{\rho}\right|_H\simeq(4\text{ or }\tfrac{2}{5})(a^{-1/2}H^2/2\pi)/[H^2(3a/8\lambda)^{1/2}(\Delta\beta)^{-3/2}] \tag{3.16}$$

$$\simeq(4\text{ or }\tfrac{2}{5})110(\lambda^{1/2}/a)[1+\ln(k_U/k)/57+\ln(g\sigma/10^{14}\text{ GeV})/57]^{3/2} . \tag{3.17}$$

As advertised, the predicted amplitude of the curvature perturbations depends on the scale only logarithmically.

To get $\delta\rho/\rho\simeq 3\times 10^{-4}$, say, on a current Hubble volume scale, we must have from Eq. (3.16),

$$\lambda\simeq 5\times 10^{-11}a^2\simeq 3\times 10^{-9} , \tag{3.18}$$

assuming that $g\sigma\sim 2.5\times 10^{14}$ GeV, $a=\frac{15}{2}$. The only way to get λ this small is to reduce the coupling g by a factor of $\sim 3\times 10^{-3}$. (The coupling constant g depends only logarithmically on the unification scale.) In other words, taking $\lambda\simeq 4$, the value obtained from $\alpha_{\text{GUT}}\equiv g(\sigma)^2/4\pi\simeq\frac{1}{45}$ through the renormalization-group equations (see Ref. 4), we find that $\delta\rho/\rho\,|_H\simeq 10$ on the current Hubble scale.

One might object that our linear perturbation analysis should not be trusted if the result is $\delta\rho/\rho\,|_H\simeq 10\gg 1$ for the appropriate GUT value of λ. Quantitatively, this objection is valid; complicated nonlinear effects may alter the actual amplitude. However, to obtain the desired result $\delta\rho/\rho\,|_H\sim 10^{-4}$, beginning from a value of the perturbation greater than unity, the nonlinear effects would have to cause the amplitude of the perturbation to decrease through a regime where its magnitude is of order 10^{-1} or 10^{-2}. At this point, a linear perturbation analysis should be valid and then, we conclude, no further decrease is possible. In other words, it is difficult to imagine that the nonlinear effects can reduce $\delta\rho/\rho\,|_H$ below 10^{-1} or 10^{-2}, and it is probably overly optimistic to expect even this decrease. While our patching together of the quantum regime and the classical regime in analyzing the evolution of the background value of ϕ is also very naïve, any model which undergoes slightly more inflation than is necessary to explain the large-scale homogeneity of the Universe should be in the classical regime [cf. Eq. (3.11)], by the time that fluctuations are being created on observationally relevant scales. Both these points deserve further investigation, but we are reasonably confident that our qualitative conclusions will stand.

The root of the large fluctuations in the conventional Coleman-Weinberg GUT is the fact that $\delta\rho/\rho\,|_H \sim H^2/\dot{\phi}(t_1)$ and $\dot{\phi} \lesssim H^2$ during the slow roll-down period. This itself stems from a very basic difficulty with this model: The inflationary epoch occurs when the classical value of the slowly evolving Higgs field is small compared to the de Sitter zero-point fluctuations in ϕ. One obvious solution is to have inflation occur when $\phi \gg H$ and $\dot{\phi} \gg H^2$ (this requires a potential which is flat for $\phi \gg H$). Arbitrarily setting λ to be very small has this effect, because it flattens the potential [cf. Eq. (3.10)] and makes the inflation epoch very long, so that by the time presently observable scales have $k/SH \sim 1$ the Higgs field has grown, by a combination of quantum fluctuations and classical roll down, to a value large compared with H and $\dot{\phi}$ to a value large compared with H^2. [For consistency, setting λ according to Eq. (3.18) also requires resetting σ to a value $\sim 10^{19}$ GeV.] We will next consider a potential which is only flat for $\phi \gg H$ and thus quite naturally avoids the pitfall of $\dot{\phi}(t_1) \lesssim H^2$ and $\delta\rho/\rho\,|_H \gg 1$.

B. Reverse hierarchy supersymmetric models—newer inflation

The disappointing result of Sec. III A makes it rather interesting to consider a class of supersymmetric models which have been recently shown to lead to inflation without any of the fine tuning of the mass parameter required in the SU(5) model with Coleman-Weinberg SSB.[18] The class of models employs the Witten reverse hierarchy scheme[20] based on models with O'Raifeartaigh-type symmetry breaking.[21] For concreteness, we shall consider the three-scale supersymmetric GUT model of Dimopoulos and Raby.[22] In this model there is one fundamental scale $M_I \sim 10^{12}$ GeV which sets the scale of supersymmetry breaking; then, there are two other scales which are generated radiatively: $M_G \sim M_{\rm Pl}$ the grand-unification scale, and $M_W \sim M_I^2/M_G$, equal to the weak scale. The scale which determines the properties of the phase transition which leads to inflation is the intermediate scale M_I. Let the field whose vacuum expectation value leads to the SSB of the GUT be ϕ. For values of ϕ much greater than M_I, but much less than M_G, the scalar potential is given by

$$V(\phi) \simeq c_1 M_I^4 - c_2 M_I^4 \ln(|\phi|/M_{\rm Pl}) \,. \tag{3.19}$$

This peculiar logarithmic dependence on ϕ occurs without any fine tuning of parameters. The behavior of $V(\phi)$ for small $\phi(\ll M_I)$ does depend upon the choice of parameters, but has no effect on the inflation which depends on only the logarithmic region of the potential.[18] The shape of the potential for ϕ near $\sigma \simeq O(M_{\rm Pl})$ (the SSB minimum) has not yet been completely determined. It is not clear whether the potential has the necessary curvature near $\phi = \sigma$, and whether ϕ couples with sufficient strength to the other fundamental fields to ensure good reheating.[18] If this particular model fails to reheat, there may exist similar models in which the reheating is not a problem. We shall not discuss this issue further here.

The salient feature of this potential is that it becomes very flat for large values of ϕ, $V'(\phi) = -c_2 M_I^4/\phi$. In fact, for $c_1/c_2 \gtrsim 10^2$ (in Ref. 18 it is argued that this is a natural choice of parameters), the potential is sufficiently flat so that when $\phi \gtrsim (3c_2/8\pi c_1 a)^{1/2} M_{\rm Pl}$ the growth of ϕ is very slow and inflation takes place. During this epoch the $\ddot{\phi}$ term in the classical evolution equation for ϕ is negligible compared with the "friction term" and once again, to a good approximation $3aH\dot{\phi} = -V' = c_2 M_I^4/\phi$ [in analogy to Eq. (3.10)]. This equation is easily integrated and we find that

$$\phi(t) \simeq (3c_2/8\pi c_1 a)^{1/2} (1 + \tfrac{2}{3} Ht)^{1/2} M_{\rm Pl} \,, \tag{3.20a}$$

$$\dot{\phi}(t) \simeq (c_2/24\pi c_1 a)^{1/2} (1 + \tfrac{2}{3} Ht)^{-1/2} H M_{\rm Pl} \,, \tag{3.20b}$$

where

$$H^2 \simeq 8\pi c_1 M_I^4 / 3M_{\rm Pl}^2 \,, \tag{3.20c}$$

and $t=0$ marks the beginning of the "slow-growth epoch". When

$$Ht_* \equiv \beta_* = 4\pi c_1 a/c_2, \quad \phi \simeq O(M_{\rm Pl}) \simeq \sigma$$

and presumably the potential steepens, ending the inflationary phase and it is hoped, beginning reheating. In terms of $\Delta\beta = H(t_* - t_1)$ which for this model is given by

$$\Delta\beta = 51 + \ln(k_U/k) + \ln(M_I/10^{12}\ {\rm GeV}) \,, \tag{3.21}$$

$$\dot{\phi}(t_1) \simeq (c_2/16\pi c_1 a)^{1/2} (\tfrac{3}{2} + \beta_* - \Delta\beta)^{-1/2} H M_{\rm Pl} \,. \tag{3.22}$$

As long as β_* is much greater than $\Delta\beta$, i.e., the physical size of the fluctuation region is much larger than the mass scale of interest, $\dot{\phi}(t_1) \simeq \dot{\phi}(t_*) = (c_2/8\pi c_1 a) H M_{\rm Pl}$. Thus, from Eqs. (2.37) and (3.2) it follows that

$$\delta\rho/\rho\,|_H = (4 \text{ or } \tfrac{2}{5}) 8 \left[\frac{2\pi}{3}\right]^{1/2} \frac{c_1^{3/2} a^{1/2}}{c_2} \left[\frac{M_I}{M_{\rm Pl}}\right]^2 \tag{3.23}$$

with $ac_1^{1/2} \sim 10$, $M_I \sim 10^{12}$ GeV one needs $(c_1/c_2) \sim 10^6$ to achieve $\delta\rho/\rho\,|_H \simeq 10^{-4}$. In these models c_1 and c_2 are not precisely determined; however, c_1/c_2 is naturally $\gtrsim 10^4$ (Ref. 18). Because $\phi \gg H$ and $\dot{\phi} \gg H^2$ throughout the inflationary epoch, the quickening of the evolution of the scalar field due to scalar-field zero-point fluctuations is also avoided.

IV. DISCUSSION AND SUMMARY

The hot big-bang model provides a reliable framework for describing the evolution of the Universe from about 0.01 sec after the big bang until the present. It quite naturally accounts for the large abundance of ^{4}He and the relatively large abundance of D, the cosmic microwave background, and the expansion of the Universe. However, there are a handful of puzzles which the model sheds no light upon. They include the large-scale homogeneity, the small-scale inhomogeneity, the isotropy, the near-critical expansion rate ($\Omega \sim 1$), and the baryon number of the Universe. The wedding of cosmology with particle physics has resulted in a rather attractive scenario for understanding the origin of the baryon asymmetry;[23] however, at the same time it has introduced additional problems: superheavy magnetic monopoles[24] (and other topological structures), and the puzzle of why the vacuum energy (which acts like a cosmological constant) is so small today.

The fundamental assumption underlying all inflationary scenarios is that the energy density associated with our asymmetric vacuum is very nearly zero and that earlier on when the Universe was in a more symmetric state the vacuum energy density was much larger, large enough so that it had significant effect on the dynamics of the expansion. When the fluctuations in the scalar field are ignored (i.e., ϕ is assumed to be uniform within a given fluctuation region), the simplest model of new inflation, the Coleman-Weinberg SU(5) model, resolves all of the puzzles with the exception of the origin of the small-density inhomogeneities necessary to produce the small-scale structures in the Universe. However, when the scalar-field fluctuations are taken into account in the simplest model, the consequences are disastrous. Firstly, it appears that the fluctuations prevent the evolution of ϕ from proceeding slowly enough for inflation to occur.[16,17] Secondly, although the scalar-field fluctuations lead to a scale-free spectrum of density inhomogeneities in the subsequent FRW phase, their magnitude is ~ 10, resulting in a highly irregular universe.

The basic idea of new inflation still seems very promising since it solves the difficulty of remaining trapped in the metastable symmetric vacuum forever by having the exponential growth occur during a period of slow ($\Delta t \gg H^{-1}$), but inevitable, progress toward the asymmetric vacuum. The problem with the simplest model is that the period of slow growth occurs when $\phi \lesssim H$ and $\dot{\phi} \lesssim H^2$. (In addition, the quantum fluctuations in ϕ, $\Delta\phi \simeq H$ are comparable to the classical value of ϕ.) Since the amplitude of the resulting density perturbation is $(\delta\rho/\rho)_H \simeq (4 \text{ or } \frac{2}{5}) H\Delta\phi/\dot{\phi}(t_1)$, the outcome is a very inhomogeneous Universe. One prescription for an "ultimate" model of particle physics that will yield the right kind of inflation is manifest: When the relevant (galactic to current comoving Hubble volume) scales cross the horizon during the de Sitter phase, $\dot{\phi}(t_1)$ must be $\sim 10^4 \times H^2$. That is, the scalar potential should be flat for $\phi \gg H$, so that inflation occurs for large values of ϕ rather than for small values.

As we have mentioned, the O'Raifeartaigh-Witten-type supersymmetric potentials have this feature. The geometric hierarchy model of Dimopoulos and Raby[22] is a potentially viable candidate for an ultimate model, although the question of adequate reheating remains to be answered.[18] Clearly, the constraint on the density perturbations that we have derived serves as a useful guide for the building of future particle-physics theories.

ACKNOWLEDGMENTS

Much of this work was done at the 1982 Nuffield Workshop on the Very Early Universe at Cambridge. We benefited greatly from the many discussions (and arguments) with other attendees at the workshop. We especially thank the organizers, S. W. Hawking and G. B. Gibbons. We particularly thank J. Barrow, A. Guth, S. W. Hawking, A. A. Starobinskii, and E. T. Vishniac for useful discussions. This work was supported in part by the DOE through contracts DE-ACO2-80ER 10773 (at Chicago) and DOE-EY-76-C-02-3071 (at Penn.), and by the NSF through grant PHY-77-27084 (at Santa Barbara). One of us (P.J.S.) was supported in part by a DOE Junior Investigator Grant.

---

[1]A. Guth, Phys. Rev. D 23, 347 (1981).

[2]A. Guth and E. Weinberg, Nucl. Phys. B212, 321 (1983); S. W. Hawking, I. G. Moss, and I. M. Stewart, Phys. Rev. D 26, 2681 (1982).

[3]A. D. Linde, Phys. Lett. 108B, 389 (1982).

[4]A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. 48, 1220 (1982).

[5]The effects of curved space-time on the phase transition in these models are discussed in S. W. Hawking and I. G. Moss, Phys. Lett. 110B, 35 (1982); see also Refs. 16 and 17.

[6]S. Coleman and E. Weinberg, Phys. Rev. D 7, 1888 (1973).

[7]P. Steinhardt, in *The Birth of the Universe,* edited by J. Andouze and J. Trân Thanh Vân (Éditions Frontières, Dreux, France, 1982).

[8]The domain-wall problem associated with the axion, see P. Sikivie, Phys. Rev. Lett. 48, 1156 (1982), is remedied by inflation; however, there may be additional problems which involve the coherent field energy of axion oscillations, see J. Preskill, M. Wise, and F. Wilczek, Phys. Lett. 120B, 127 (1983); L. Abbott and P. Sikivie, *ibid.* 120B, 133 (1983); M. Dine and W. Fischler, *ibid.* 120B, 137 (1983).

[9]A. Albrecht, P. J. Steinhardt, M. S. Turner, and F. Wilczek, Phys. Rev. Lett. 48, 1437 (1982); A. D. Dolgov and A. D. Linde, Phys. Lett. 116B, 329 (1982); L. Abbott, E. Farhi, and M. Wise, *ibid.* 117B, 29 (1982).

[10]The generation of density perturbations due to scalar field fluctuations has been analyzed in a variety of different ways by a number of authors who all reach essentially the same conclusions as we do; A. Guth and S.-Y. Pi, Phys. Rev. Lett. 49, 1110 (1982); S. W. Hawking, Phys. Lett. 115B, 295 (1982); A. A. Starobinskii, *ibid.* 117B, 175 (1982). Also, see earlier related work by V. F. Mukhanov and G. V. Chibisov, Pis'ma Zh. Eksp. Teor. Fiz. 33, 549 (1981) [JETP Lett. 33, 532 (1981)] and Lebedev Report No. 198, 1982 (unpublished); Lukash, and Vishniac and Gott have analyzed the spontaneous production of scale-free density perturbations due to zero-point fluctuations of quantized sound (pressure) waves in a de Sitter Universe; V. N. Lukash, Acad. of Sci. of the USSR Sp. Res. Inst. Report No. 559, 1980 (unpublished); E. T. Vishniac and J. R. Gott, III, 1982 (unpublished).

[11]Ya. B. Zel'dovich, Mon. Not. R. Astron. Soc. 106, 1P (1972); E. R. Harrison, Phys. Rev. D 1, 2726 (1970). See also A. G. Doroshkevich, R. A. Sunyaev, and Ya. B. Zel'dovich, in *Confrontation of Cosmological Theories with Observation Data,* edited by M. S. Longair (Reidel, Dordrecht, 1974), p. 213.

[12]R. K. Sachs and A. M. Wolfe, Astrophys. J. 147, 73 (1967).

[13]See, e.g., P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University, Princeton, 1980); A. G. Doroshkevich, Ya. B. Zel'dovich, and R. A. Sunyaev, Astron. Zh. 55, 913 (1978) [Sov. Astron. 22, 523 (1978)]; J. Silk and M. L. Wilson, Phys. Scr. 21, 708 (1980); Astrophys. J. 243, 14 (1981).

[14]K. Olive, D. N. Schramm, G. Steigman, M. Turner, and J. Yang, Astrophys. J. 246, 557 (1981); J. Yang, M. Turner, G. Steigman, D. Schramm, and K. Olive, University of Chicago report, 1983 (unpublished).

[15]J. Bond and A. Szalay, in *Neutrino '81*, proceedings of the 1981 International Conference on Neutrino Physics and Astrophysics, Maui, Hawaii, edited by R. J. Cence, E. Ma, and A. Roberts (University of Hawaii, Honolulu, 1981).

[16]A. Vilenkin and L. Ford, Phys. Rev. D 26, 1231 (1982).

[17]A. D. Linde, Phys. Lett. 116B, 335 (1982).

[18]A. Albrecht, S. Dimopoulos, W. Fischler, E. Kolb, S. Raby, and P. Steinhardt, Nucl. Phys. B (to be published). See also P. Steinhardt, in *The Very Early Universe*, edited by G. Gibbons, S. Hawking, and S. Siklos (Cambridge University Press, Cambridge, 1983).

[19]J. M. Bardeen, Phys. Rev. D 22, 1882 (1980).

[20]E. Witten, Phys. Lett. 105B, 267 (1981).

[21]L. O'Raifeartaigh, Nucl. Phys. B96, 331 (1975).

[22]S. Dimopoulos and S. Raby, Nucl. Phys. B (to be published).

[23]A. D. Sakharov, Pis'ma Zh. Eksp. Teor. Fiz. 5, 32 (1967) [JETP Lett. 5, 24 (1967)]; M. Yoshimura, Phys. Rev. Lett. 41, 281 (1978); A. Ignatiev, N. Krasnikov, V. Kuzmin, and A. Tavkhelidze, Phys. Lett. 76B, 436 (1978); D. Toussaint, S. B. Treiman, F. Wilczek, and A. Zee, Phys. Rev. D 19, 1036 (1979); S. Weinberg, Phys. Rev. Lett. 42, 850 (1979); J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, Phys. Lett. 80B, 360 (1978); S. Dimopoulos and L. Susskind, Phys. Rev. D 18, 4500 (1978).

[24]J. P. Preskill, Phys. Rev. Lett. 43, 1365 (1979); Ya. B. Zel'dovich and M. Yu. Khlopov, Phys. Lett. 79B, 239 (1979).