



I. INTRODUCTION: THE HORIZON AND FLATNESS PROBLEMS

The standard model of hot big bang cosmology relies on the assumption of initial conditions which are very puzzling in two ways, which I will explain below. The purpose of this paper is to suggest a modified scenario which avoids both of these puzzles.

By "standard model," I refer to an adiabatically expanding radiation-dominated universe described by a Robertson-Walker metric. Details will be given in Section II.

Before explaining the puzzles, I would first like to clarify my notion of "initial conditions." The standard model has a singularity which is conventionally taken to be at time $t = 0$. As $t \rightarrow 0$, the temperature $T \rightarrow \infty$. Thus, no initial value problem can be defined at $t = 0$. However, when T is of the order of the Planck mass ($M_P \equiv 1/\sqrt{G} = 1.22 \times 10^{19} \text{ GeV}$)¹ or greater, the equations of the standard model are undoubtedly meaningless, since quantum gravitational effects are expected to become essential. Thus, within the scope of our knowledge, it is sensible to begin the hot big bang scenario at some temperature T_0 which is comfortably below M_P ; let us say $T_0 = 10^{17} \text{ GeV}$. At this time one can take the description of the universe as a set of initial conditions, and the equations of motion then describe the subsequent evolution. Of course, the equation of state for matter at these temperatures is not really known, but one can make various hypotheses and pursue the consequences.

In the standard model, the initial universe is taken to be homogeneous and isotropic, and filled with a gas of effectively massless particles in thermal equilibrium at temperature T_0 . The initial value of the Hubble expansion "constant" H is taken to be H_0 , and the model universe is then completely described.

THE INFLATIONARY UNIVERSE:

A POSSIBLE SOLUTION TO THE HORIZON AND FLATNESS PROBLEMS*

Alan H. Guth**

Stanford Linear Accelerator Center
Stanford University, Stanford, California 94305

ABSTRACT

The standard model of hot big bang cosmology requires initial conditions which are problematic in two ways: (1) the early universe is assumed to be highly homogeneous, in spite of the fact that separated regions were causally disconnected (horizon problem); and (2) the initial value of the Hubble constant must be fine tuned to extraordinary accuracy to produce a universe as flat (i.e., near critical mass density) as the one we see today (flatness problem). These problems would disappear if, in its early history, the universe supercooled to temperatures 28 or more orders of magnitude below the critical temperature for some phase transition. A huge expansion factor would then result from a period of exponential growth, and the entropy of the universe would be multiplied by a huge factor when the latent heat is released. Such a scenario is completely natural in the context of grand unified models of elementary particle interactions. In such models, the supercooling is also relevant to the problem of monopole suppression. Unfortunately, the scenario seems to lead to some unacceptable consequences, so modifications must be sought.

(Submitted to Physical Review D)

*Work supported by the Department of Energy under contract number DE-AC03-76SF00515.

**Present address: Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Now I can explain the puzzles. The first is the well-known horizon problem.² The initial universe is assumed to be homogeneous, yet it consists of at least $\sim 10^{83}$ separate regions which are causally disconnected (i.e., these regions have not yet had time to communicate with each other via light signals).⁵ (The precise assumptions which lead to these numbers will be spelled out in Section II.) Thus, one must assume that the forces which created these initial conditions were capable of violating causality.

The second puzzle is the flatness problem. This puzzle seems to be much less celebrated than the first, but it has been stressed by Dicke and Peebles.⁶ I feel that it is of comparable importance to the first. It is known that the energy density ρ of the universe today is near the critical value ρ_{cr} (corresponding to the borderline between an open and closed universe). One can safely assume that⁷

$$.01 < \Omega_p < 10, \quad (1.1)$$

where

$$\Omega \equiv \rho/\rho_{cr} = (8\pi/3)G\rho/H^2, \quad (1.2)$$

and the subscript p denotes the value at the present time. Although these bounds do not appear at first sight to be remarkably stringent, they in fact have powerful implications. The key point is that the condition $\Omega \approx 1$ is unstable. Furthermore, the only time scale which appears in the equations for a radiation-dominated universe is the Planck time, $1/M_p = 5.4 \times 10^{-44}$ sec. A typical closed universe will reach its maximum size on the order of this time scale, while a typical open universe will dwindle to a value of ρ much less than ρ_{cr} . A universe

can survive $\sim 10^{10}$ years only by extreme fine tuning of the initial values of ρ and H , so that ρ is very near ρ_{cr} . For the initial conditions taken at $T_0 = 10^{17}$ GeV, the value of H_0 must be fine tuned to an accuracy of one part in 10^{55} . In the standard model this incredibly precise initial relationship must be assumed without explanation. (For any reader who is not convinced that there is a real problem here, variations of this argument are given in the Appendix.)

The reader should not assume that these incredible numbers are due merely to the rather large value I have taken for T_0 . If I had chosen a modest value like $T_0 = 1$ MeV, I would still have concluded that the "initial" universe consisted of at least $\sim 10^{22}$ causally disconnected regions, and that the initial value of H_0 was fine tuned to one part in 10^{15} . These numbers are much smaller than the previous set, but they are still very impressive.

Of course, any problem involving the initial conditions can always be put off until we understand the physics of $T \gtrsim M_p$. However, it is the purpose of this paper to show that these puzzles might be obviated by a scenario for the behavior of the universe at temperatures well below M_p .

The paper is organized as follows. The assumptions and basic equations of the standard model are summarized in Section II. In Section III, I describe the inflationary universe scenario, showing how it can eliminate the horizon and flatness problems. The scenario is discussed in the context of grand unified models in Section IV, and comments are made concerning magnetic monopole suppression. In Section V I discuss briefly the key undesirable feature of the scenario: the

inhomogeneities produced by the random nucleation of bubbles. Some vague ideas which might alleviate these difficulties are mentioned in Section VI.

II. THE STANDARD MODEL OF THE VERY EARLY UNIVERSE

In this section I will summarize the basic equations of the standard model, and I will spell out the assumptions which lead to the statements made in the introduction.

The universe is assumed to be homogeneous and isotropic, and is therefore described by the Robertson-Walker metric:⁸

$$d\tau^2 = dt^2 - R^2(t) \left\{ \frac{dr^2}{1 - kr^2} + r^2 [d\theta^2 + \sin^2\theta d\phi^2] \right\} \quad (2.1)$$

where $k = +1, -1$, or 0 for a closed, open, or flat universe, respectively.

It should be emphasized that any value of k is possible, but by convention r and $R(t)$ are rescaled so that k takes on one of the three discrete values. The evolution of $R(t)$ is governed by the Einstein equations

$$\ddot{R} = -\frac{4\pi}{3} G(\rho + 3p)R \quad , \quad (2.2a)$$

$$H^2 + \frac{k}{R^2} = \frac{8\pi}{3} G\rho \quad , \quad (2.2b)$$

where $H \equiv \dot{R}/R$ is the hubble "constant" (the dot denotes the derivative with respect to t). Conservation of energy is expressed by

$$\frac{d}{dt} (\rho R^3) = -p \frac{d}{dt} (R^3) \quad , \quad (2.3)$$

where p denotes the pressure. In the standard model one also assumes that the expansion is adiabatic, in which case

$$\frac{d}{dt} (sR^3) = 0 \quad , \quad (2.4)$$

where s is the entropy density.

To determine the evolution of the universe, the above equations must be supplemented by an equation of state for matter. It is now standard to describe matter by means of a field theory, and at high temperatures this means that the equation of state is to a good approximation that of an ideal quantum gas of massless particles. Let $N_b(T)$ denote the number of bosonic spin degrees of freedom which are effectively massless at temperature T (e.g., the photon contributes two units to N_b); and let $N_f(T)$ denote the corresponding number for fermions (e.g., electrons and positrons together contribute four units). Provided that T is not near any mass thresholds, the thermodynamic functions are given by

$$\rho = 3p = \frac{\pi^2}{30} \mathcal{N}(T) T^4 \quad (2.5)$$

$$s = \frac{2\pi^2}{45} \mathcal{N}(T) T^3 \quad (2.6)$$

$$n = \frac{\zeta(3)}{\pi^2} \mathcal{N}'(T) T^3 \quad (2.7)$$

where

$$\mathcal{N}(T) = N_b(T) + \frac{7}{8} N_f(T) \quad (2.8)$$

$$\mathcal{N}'(T) = N_b'(T) + \frac{3}{4} N_f'(T) \quad . \quad (2.9)$$

Here n denotes the particle number density, and $\zeta(3) = 1.20206\dots$ is the Riemann zeta function.

The evolution of the universe is then found by rewriting (2.2b) solely in terms of the temperature. Again assuming that T is not near any mass thresholds, one finds

$$\left(\frac{\dot{T}}{T} \right)^2 + \epsilon(T) T^2 = \frac{4\pi^3}{45} G \mathcal{N}(T) T^4 \quad , \quad (2.10)$$

where

$$\epsilon(T) = \frac{k}{R^2 T^2} = k \left[\frac{2\pi^2}{45} \frac{\mathcal{N}(T)}{S} \right]^{2/3}, \quad (2.11)$$

where $S \equiv R^3 s$ denotes the total entropy in a volume specified by the radius of curvature R .

Since S is conserved, its value in the early universe can be determined (or at least bounded) by current observations. Taking $\rho < 10 \rho_{cr}$ today, it follows that today

$$\left| \frac{k}{R^2} \right| < 9H^2. \quad (2.12)$$

From now on I will take $k = \pm 1$; the special case $k = 0$ is still included as the limit $R \rightarrow \infty$. Then today $R > \frac{1}{3} H^{-1} \sim 3 \times 10^9$ years. Taking the present photon temperature T_γ as 2.7°K, one then finds that the photon contribution to S is bounded by

$$S_\gamma > 3 \times 10^{85} \quad (2.13)$$

Assuming that there are three species of massless neutrinos (e , μ , and τ), all of which decouple at a time when the other effectively massless particles are the electrons and photons, then $S_\nu = (21/22)S_\gamma$. Thus,

$$S > 10^{86}, \quad (2.14)$$

and

$$|\epsilon| < 10^{-58} \mathcal{N}^{2/3}. \quad (2.15)$$

But then

$$\left| \frac{\rho - \rho_{cr}}{\rho} \right| = \frac{45}{4\pi^3} \frac{M_P^2}{\mathcal{N} T^2} |\epsilon| < 3 \times 10^{-59} \mathcal{N}^{-1/3} (M_P/T)^2. \quad (2.16)$$

Taking $T = 10^{17}$ GeV and $\mathcal{N} \sim 10^2$ (typical of grand unified models), one finds $|\rho - \rho_{cr}|/\rho < 10^{-55}$. This is the flatness problem.

The ϵT^2 term can now be deleted from (2.10), which is then solved (for temperatures higher than all particle masses) to give

$$T^2 = \frac{M_P}{2\gamma t} \quad (2.17)$$

where $\gamma^2 = (4\pi^3/45)\mathcal{N}$. (For the minimal SU_5 grand unified model, $N_b = 82$, $N_f = 90$, and $\gamma = 21.05$.) Conservation of entropy implies $RT = \text{constant}$, so $R \propto t^{1/2}$. A light pulse beginning at $t = 0$ will have travelled by time t a physical distance

$$\ell(t) = R(t) \int_0^t dt' R^{-1}(t') = 2t, \quad (2.18)$$

and this gives the physical horizon distance. This horizon distance is to be compared with the radius $L(t)$ of the region at time t which will evolve into our presently observed region of the universe. Again using conservation of entropy,

$$L(t) = [s_p/s(t)]^{1/3} L_p, \quad (2.19)$$

where s_p is the present entropy density and $L_p \sim 10^{10}$ years is the radius of the presently observed region of the universe. One is interested in the ratio of volumes, so

$$\begin{aligned} \frac{\ell^3}{L^3} &= \frac{11}{43} \left(\frac{45}{4\pi^3} \right)^{3/2} \mathcal{N}^{-1/2} \left[\frac{M_P}{L_p T_\gamma T} \right]^3 \\ &= 4 \times 10^{-89} \mathcal{N}^{-1/2} (M_P/T)^3. \end{aligned} \quad (2.20)$$

Taking $\mathcal{N} \sim 10^2$ and $T_0 = 10^{17}$ GeV, one finds $\ell_o^3/L_o^3 = 10^{-83}$. This is the horizon problem.

III. THE INFLATIONARY UNIVERSE

In this section I will describe a scenario which is capable of avoiding the horizon and flatness problems.

From Section II one can see that both problems could disappear if the assumption of adiabaticity were grossly incorrect. Suppose instead that

$$S_p = Z^3 S_0, \quad (3.1)$$

where S_p and S_0 denote the present and initial values of $R^3 s$, and Z is some large factor.

Let us look first at the flatness problem. Given (3.1), the RHS of (2.16) is multiplied by a factor of Z^2 . The "initial" value (at $T_0 = 10^{17}$ GeV) of $|\rho - \rho_{cr}|/\rho$ could be of order unity, and the flatness problem would be obviated, if

$$Z > 3 \times 10^{27}. \quad (3.2)$$

Now consider the horizon problem. The RHS of (2.19) is multiplied by Z^{-1} , which means that the length scale of the early universe, at any given temperature, was smaller by a factor of Z than had been previously thought. If Z is sufficiently large, then the initial region which evolved into our observed region of the universe would have been smaller than the horizon distance at that time. To see how large Z must be, note that the RHS of (2.20) is multiplied by Z^3 . Thus, if

$$Z > 5 \times 10^{27}, \quad (3.3)$$

then the horizon problem disappears. (It should be noted that the horizon will still exist; it will simply be moved out to distances which have not been observed.)

It is not surprising that the RHS's of (3.2) and (3.3) are approximately equal, since they both correspond roughly to S_0 of order unity.

I will now describe a scenario, which I call the inflationary universe, which is capable of such a large entropy production.

Suppose the equation of state for matter (with all chemical potentials set equal to zero), exhibits a first order phase transition at some critical temperature T_c . Then as the universe cools through the temperature T_c , one would expect bubbles of the low temperature phase to nucleate and grow. However, suppose the nucleation rate for this phase transition is rather low. The universe will continue to cool as it expands, and it will then supercool in the high temperature phase. Suppose that this supercooling continues down to some temperature T_s , many orders of magnitude below T_c . When the phase transition finally takes place at temperature T_s , the latent heat is released. However, this latent heat is characteristic of the energy scale T_c , which is huge relative to T_s . The universe is then reheated to some temperature T_r which is comparable to T_c . The entropy density is then increased by a factor of roughly $(T_r/T_s)^3$ (assuming that the number \mathcal{N} of degrees of freedom for the two phases are comparable), while the value of R remains unchanged. Thus,

$$Z \approx T_r/T_s. \quad (3.4)$$

If the universe supercools by 28 or more orders of magnitude below some critical temperature, the horizon and flatness problems disappear.

In order for this scenario to work, it is necessary for the universe to be essentially devoid of any strictly conserved quantities. Let n denote the density of some strictly conserved quantity, and let $r \equiv n/s$

denote the ratio of this conserved quantity to entropy. Then $r_p = Z^{-3} r_0 < 10^{-84} r_0$. Thus, only an absurdly large value for the initial ratio would lead to a measurable value for the present ratio. Thus, if baryon number were exactly conserved, the inflationary model would be untenable. However, in the context of grand unified models, baryon number is not exactly conserved. The net baryon number of the universe is believed to be created by CP-violating interactions at a temperature of $10^{13}-10^{14}$ GeV.⁹ Thus, provided that T_c lies in this range or higher, there is no problem. The baryon production would take place after the reheating. (However, strong constraints are imposed on the entropy which can be generated in any phase transition with $T_c \ll 10^{14}$ GeV: in particular, the Weinberg-Salam phase transition.³⁶)

Let us examine the properties of the supercooling universe in more detail. Note that the energy density $\rho(T)$, given in the standard model by (2.5), must now be modified. As $T \rightarrow 0$, the system is cooling not toward the true vacuum, but rather toward some metastable false vacuum with an energy density ρ_0 which is necessarily higher than that of the true vacuum. Thus, to a good approximation (ignoring mass thresholds)

$$\rho(T) = \frac{\pi^2}{30} \mathcal{N}(T) T^4 + \rho_0 \quad (3.5)$$

Perhaps a few words should be said concerning the zero point of energy. Classical general relativity couples to an energy-momentum tensor of matter, $T_{\mu\nu}$, which is necessarily (covariantly) conserved. When matter is described by a field theory, the form of $T_{\mu\nu}$ is determined by the conservation requirement up to the possible modification

$$T_{\mu\nu} \rightarrow T_{\mu\nu} + \lambda g_{\mu\nu} \quad (3.6)$$

for any constant λ . (λ cannot depend on the values of the fields, nor can it depend on the temperature or the phase.) The freedom to introduce the modification (3.6) is identical to the freedom to introduce a cosmological constant into Einstein's equations. One can always choose to write Einstein's equations without an explicit cosmological term; the cosmological constant Λ is then defined by

$$\langle 0 | T_{\mu\nu} | 0 \rangle = \Lambda g_{\mu\nu} \quad (3.7)$$

where $|0\rangle$ denotes the true vacuum. Λ is identified as the energy density of the vacuum, and in principle there is no reason for it to vanish. Empirically Λ is known to be very small ($|\Lambda| < 10^{-46}$ GeV⁴)¹⁰ so I will take its value to be zero.¹¹ The value of ρ_0 is then necessarily positive, and is determined by the particle theory.¹² It is typically of $\mathcal{O}(T_c^4)$.

Using (3.5), Eq. (2.10) becomes

$$\left(\frac{\dot{T}}{T}\right)^2 = \frac{4\pi^3}{45} G \mathcal{N}(T) T^4 - \varepsilon(T) T^2 + \frac{8\pi}{3} G \rho_0 \quad (3.8)$$

This equation has two types of solutions, depending on the parameters.

If $\varepsilon > \varepsilon_0$, where

$$\varepsilon_0 = \frac{8\pi^2 \sqrt{30}}{45} G \sqrt{\mathcal{N} \rho_0} \quad (3.9)$$

then the expansion of the universe is halted at a temperature T_{\min} given by

$$T_{\min}^4 = \frac{30\rho_0}{\pi^2 \mathcal{N}} \left[\frac{\varepsilon}{\varepsilon_0} + \sqrt{\left(\frac{\varepsilon}{\varepsilon_0}\right)^2 - 1} \right]^2 \quad (3.10)$$

and then the universe contracts again. Note that T_{\min} is of $\mathcal{O}(T_c)$, so this is not the desired scenario. The case of interest is $\varepsilon < \varepsilon_0$,

in which case the expansion of the universe is unchecked. (Note that $\epsilon_0 \sim \sqrt{\Lambda} T_c^2/M_P^2$ is presumably a very small number. Thus $0 < \epsilon < \epsilon_0$ (a closed universe) seems unlikely, but $\epsilon < 0$ (an open universe) is quite plausible.) Once the temperature is low enough for the ρ_0 term to dominate over the other two terms on the RHS of (3.8), one has

$$T(t) \approx \text{const} \times e^{-\chi t}, \quad (3.11)$$

where

$$\chi^2 = \frac{8\pi}{3} G\rho_0. \quad (3.12)$$

Since $RT = \text{const}$, one has¹³

$$R(t) = \text{const} \times e^{\chi t}. \quad (3.13)$$

The universe is expanding exponentially, in a false vacuum state of energy density ρ_0 . The Hubble constant is given by $H = \dot{R}/R = \chi$. (More precisely, H approaches χ monotonically from above. This behavior differs markedly from the standard model, in which H falls as t^{-1} .)

The false vacuum state is Lorentz-invariant, so $T_{\mu\nu} = \rho_0 g_{\mu\nu}$. It follows that $p = -\rho_0$: the pressure is negative. This negative pressure allows for the conservation of energy, Eq. (2.3). From the second order Einstein equation (2.2a), it can be seen that the negative pressure is also the driving force behind the exponential expansion.

The Lorentz invariance of the false vacuum has one other consequence: the metric described by (3.13) (with $k = 0$) does not single out a comoving frame. The metric is invariant under an $O(4,1)$ group of transformations, in contrast to the usual Robertson-Walker invariance of $O(4)$.¹⁴ It is known as the de Sitter metric, and it is discussed in the standard literature.¹⁵

Now consider the process of bubble formation in a Robertson-Walker universe. The bubbles form randomly, so there is a certain nucleation rate $\lambda(t)$, which is the probability per (physical) volume per time that a bubble will form in any region which is still in the high temperature phase. I will idealize the situation slightly, and assume that the bubbles start at a point and expand at the speed of light. Furthermore, I neglect k in the metric, so $d\tau^2 = dt^2 - R^2(t)d\vec{x}^2$.

I want to calculate $p(t)$, the probability that any given point remains in the high temperature phase at time t . Note that the distribution of bubbles is totally uncorrelated, except for the exclusion principle that bubbles do not form inside of bubbles. This exclusion principle causes no problem, because one can imagine fictitious bubbles which form inside the real bubbles with the same nucleation rate $\lambda(t)$. With all bubbles expanding at the speed of light, the fictitious bubbles will be forever inside the real bubbles, and will have no effect on $p(t)$. The distribution of all bubbles, real and fictitious, is then totally uncorrelated.

$p(t)$ is the probability that there are no bubbles which engulf a given point in space. But the number of bubbles which engulf a given point is a Poisson distributed variable, so $p(t) = \exp[-\bar{N}(t)]$, where $\bar{N}(t)$ is the expectation value of the number of bubbles engulfing the point. Thus¹⁶

$$p(t) = \exp \left\{ - \int_0^t dt_1 \lambda(t_1) R^3(t_1) V(t, t_1) \right\}, \quad (3.14)$$

where

$$V(t, t_1) = \frac{4\pi}{3} \left[\int_{t_1}^t \frac{dt_2}{R(t_2)} \right]^3 \quad (3.15)$$

is the coordinate volume at time t of a bubble which formed at time t_1 .

I will now assume that the nucleation rate is sufficiently slow so that no significant nucleation takes place until $T \ll T_c$, when exponential growth has set in. I will further assume that by this time $\lambda(t)$ is given approximately by the zero temperature nucleation rate, λ_0 . One then has

$$p(t) = \exp \left\{ -\frac{t}{\tau} + \mathcal{O}(1) \right\}, \quad (3.16)$$

where

$$\tau = \frac{3\chi^3}{4\pi\lambda_0}, \quad (3.17)$$

and $\mathcal{O}(1)$ refers to terms which approach a constant as $\chi t \rightarrow \infty$. During one of these time constants, the universe will expand by a factor

$$Z_\tau = \exp(\chi\tau) = \exp \left\{ \frac{3\chi^4}{4\pi\lambda_0} \right\}. \quad (3.18)$$

If the phase transition is associated with the expectation value of a Higgs field, then λ_0 can be calculated using the method of Coleman and Callan.¹⁷ The key point is that nucleation is a tunnelling process, so that λ_0 is typically very small. The Coleman-Callan method gives an answer of the form

$$\lambda_0 = A\rho_0 \exp(-B) \quad (3.19)$$

where B is a barrier penetration term, and A is a dimensionless coefficient of order unity. Since Z_τ is then an exponential of an exponential,

one can very easily^{18,19,36} obtain values as large as $\log_{10} Z \approx 28$, or even $\log_{10} Z \approx 10^{10}$.

Thus, if the universe reaches a state of exponential growth, it is quite plausible for it to expand and supercool by a huge number of orders of magnitude before a significant fraction of the universe undergoes the phase transition.

So far I have assumed that the early universe can be described from the beginning by a Robertson-Walker metric. If this assumption were really necessary, then it would be senseless to talk about "solving" the horizon problem; perfect homogeneity was assumed at the outset. Thus, I must now argue that the assumption can probably be dropped.

Suppose instead that the initial metric, and the distribution of particles, was rather chaotic. One would then expect that statistical effects would tend to thermalize the particle distribution on a local scale.²⁰ It has also been shown (in idealized circumstances) that anisotropies in the metric are damped out on the time scale of $\sim 10^3$ Planck times.²¹ The damping of inhomogeneities in the metric has also been studied,²² and it is reasonable to expect such damping to occur. Thus, assuming that at least some regions of the universe started at temperatures high compared to T_c , one would expect that, by the time the temperature in one of these regions falls to T_c , it will be locally homogeneous, isotropic, and in thermal equilibrium. By locally, I am talking about a length scale ξ which is of course less than the horizon distance. It will then be possible to describe this local region of the universe by a Robertson-Walker metric, which will be accurate at distance scales small compared to ξ . When the temperature of such a

region falls below T_c , the inflationary scenario will take place. The end result will be a huge region of space which is homogeneous, isotropic, and of nearly critical mass density. If Z is sufficiently large, this region can be bigger than (or much bigger than) our observed region of the universe.

IV. GRAND UNIFIED MODELS AND MAGNETIC MONOPOLE PRODUCTION

In this section I will discuss the inflationary model in the context of grand unified models of elementary particle interactions.^{23,24}

A grand unified model begins with a simple gauge group G which is a valid symmetry at the highest energies. As the energy is lowered, the theory undergoes a hierarchy of spontaneous symmetry breaking into successive subgroups: $G \rightarrow H_n \rightarrow \dots \rightarrow H_0$, where $H_1 = SU_3 \times SU_2 \times U_1$ (QCD \times Weinberg-Salam) and $H_0 = SU_3 \times U_1^{EM}$. In the Georgi-Glashow model,²³ which is the simplest model of this type, $G = SU_5$ and $n = 1$. The symmetry breaking of $SU_5 \rightarrow SU_3 \times SU_2 \times U_1$ occurs at an energy scale $M_X \sim 10^{14}$ GeV.

At high temperatures, it was suggested by Kirzhnits and Linde²⁵ that the Higgs fields of any spontaneously broken gauge theory would lose their expectation values, resulting in a high temperature phase in which the full gauge symmetry is restored. A formalism for treating such problems was developed²⁶ by Weinberg and by Dolan and Jackiw. In the range of parameters for which the tree potential is valid, the phase structure of the SU_5 model was analyzed by Tye and me.^{16,27} We found that the SU_5 symmetry is restored at $T > \sim 10^{14}$ GeV, and that for most values of the parameters there is an intermediate temperature phase with gauge symmetry $SU_4 \times U_1$, which disappears at $T \sim 10^{13}$ GeV. Thus,

grand unified models tend to provide phase transitions which could lead to an inflationary scenario of the universe.

Grand unified models have another feature with important cosmological consequences: they contain very heavy magnetic monopoles in their particle spectrum. These monopoles are of the type discovered by 't Hooft and Polyakov,²⁸ and will be present in any model satisfying the above description.²⁹ These monopoles typically have masses of order $M_X/\alpha \sim 10^{16}$ GeV, where $\alpha = g^2/4\pi$ is the grand unified fine structure constant. Since the monopoles are really topologically stable knots in the Higgs field expectation value, they do not exist in the high temperature phase of the theory. They therefore come into existence during the course of a phase transition, and the dynamics of the phase transition is then intimately related to the monopole production rate.

The problem of monopole production and their subsequent annihilation, in the context of a second order or weakly first order phase transition, was analyzed by Zeldovich and Khlopov³⁰ and by Preskill.³¹ In Preskill's analysis, which was more specifically geared toward grand unified models, it was found that relic monopoles would exceed present bounds by roughly 14 orders of magnitude. Since it seems difficult to modify the estimated annihilation rate, one must find a scenario which suppresses the production of these monopoles.

Kibble³² has pointed out that monopoles are produced in the course of the phase transition by the process of bubble coalescence. The orientation of the Higgs field inside one bubble will have no correlation with that of another bubble not in contact. When the bubbles coalesce to fill the space, it will be impossible for the uncorrelated Higgs fields to

align uniformly. One expects to find topological knots, and these knots are the monopoles. The number of monopoles so produced is then comparable to the number of bubbles, to within a few orders of magnitude.

Kibble's production mechanism can be used to set a "horizon bound" on monopole production which is valid if the phase transition does not significantly disturb the evolution of the universe.³³ At the time of bubble coalescence t_{coal} , the size ℓ of the bubbles cannot exceed the horizon distance at that time. So

$$\ell < 2t_{\text{coal}} = \frac{M_P}{\gamma_{T_{\text{coal}}}^2} \quad (4.1)$$

By Kibble's argument, the density n_M of monopoles then obeys

$$n_M \gtrsim \ell^{-3} > \frac{\gamma^3 T_{\text{coal}}^6}{M_P^3} \quad (4.2)$$

By considering the contribution to the mass density of the present universe which could come from 10^{16} GeV monopoles, Preskill³¹ concludes that

$$n_M/n_\gamma < 10^{-24} \quad , \quad (4.3)$$

where n_γ is the density of photons. This ratio changes very little from the time of the phase transition, so with (2.7) one concludes

$$T_{\text{coal}} < \left[\frac{10^{-24} \pi^2}{2\zeta(3)} \right]^{1/3} \gamma^{-1} M_P \approx 10^{10} \text{ GeV} \quad . \quad (4.4)$$

If $T_c \sim 10^{14}$ GeV, this bound implies that the universe must supercool by at least about four orders of magnitude before the phase transition is completed.

The problem of monopole production in a strongly first order phase transition with supercooling was treated in more detail by Tye and me.^{16,34}

We showed how to explicitly calculate the bubble density in terms of the nucleation rate, and we considered the effects of the latent heat released in the phase transition. Our conclusion was that (4.4) should be replaced by

$$T_{\text{coal}} < 2 \times 10^{11} \text{ GeV} \quad , \quad (4.5)$$

where T_{coal} refers to the temperature just before the release of the latent heat.

Tye and I omitted the crucial effects of the mass density ρ_0 of the false vacuum. However, our work has one clear implication: if the nucleation rate is sufficiently large to avoid exponential growth, then far too many monopoles would be produced. Thus, the monopole problem seems to also force one into the inflationary scenario.³⁵

In the simplest SU_5 model, the nucleation rates have been calculated (approximately) by E. Weinberg and me.¹⁹ The model contains unknown parameters, so no definitive answer is possible. We do find, however, that there is a sizable range of parameters which lead to the inflationary scenario.³⁶

V. PROBLEMS OF THE INFLATIONARY SCENARIO³⁷

As I mentioned earlier, the inflationary scenario seems to lead to some unacceptable consequences. It is hoped that some variation can be found which avoids these undesirable features, but maintains the desirable ones. The problems of the model will be discussed in more detail elsewhere,³⁷ but for completeness I will give a brief description here.

The central problem is the difficulty in finding a smooth ending to the period of exponential expansion. Let us assume that $\lambda(t)$ approaches a constant as $t \rightarrow \infty$ and $T \rightarrow 0$. To achieve the desired expansion factor $Z > 10^{28}$, one needs $\lambda_0/\chi^4 < 10^{-2}$ (see (3.18)), which means that the nucleation rate is slow compared to the expansion rate of the universe. (Explicit calculations show that λ_0/χ^4 is typically much smaller than this value.^{18,19,36}) The randomness of the bubble formation process then leads to gross inhomogeneities.

To understand the effects of this randomness, the reader should bear in mind the following facts:

- (i) All of the latent heat released as a bubble expands is transferred initially to the walls of the bubble.¹⁷ This energy can be thermalized only when the bubble walls undergo many collisions.
- (ii) The de Sitter metric does not single out a comoving frame. The $O(4,1)$ invariance of the de Sitter metric is maintained even after the formation of one bubble. The memory of the original Robertson-Walker comoving frame is maintained by the probability distribution of bubbles, but the local comoving frame can be reestablished only after enough bubbles have collided.
- (iii) The size of the largest bubbles will exceed that of the smallest bubbles by roughly a factor of Z ; the range of bubble sizes is immense. The surface energy density grows with the size of the bubble, so the energy in the walls of the largest bubbles can be thermalized only by colliding with other large bubbles.
- (iv) As time goes on, an arbitrarily large fraction of the space will be in the new phase (see (3.16)). However, one can ask a more

subtle question about the region of space which is in the new phase: is the region composed of finite separated clusters, or do these clusters join together to form an infinite region? The latter possibility is called "percolation." It can be shown³⁸ that the system percolates for large values of λ_0/χ^4 , but that for sufficiently small values it does not. The critical value of λ_0/χ^4 has not been determined, but presumably an inflationary universe would have a value of λ_0/χ^4 below critical. Thus, no matter how long one waits, the region of space in the new phase will consist of finite clusters, each totally surrounded by a region in the old phase.

- (v) Each cluster will contain only a few of the largest bubbles. Thus, the collisions discussed in (iii) cannot occur.

The above statements do not quite prove that the scenario is impossible, but these consequences are at best very unattractive. Thus, it seems that the scenario will become viable only if some modification can be found which avoids these inhomogeneities. Some vague possibilities will be mentioned in the next section.

Note that the above arguments seem to rule out the possibility that the universe was ever trapped in a false vacuum state, unless $\lambda_0/\chi^4 \gtrsim 1$. Such a large value of λ_0/χ^4 does not seem likely, but it is possible.¹⁹

VI. CONCLUSION

I have tried to convince the reader that the standard model of the very early universe requires the assumption of initial conditions which are very implausible for two reasons:

- (i) The horizon problem: causally disconnected regions are assumed to be nearly identical; in particular, they are simultaneously at the same temperature.

(ii) The flatness problem: for a fixed initial temperature, the initial value of the Hubble "constant" must be fine tuned to extraordinary accuracy to produce a universe which is as flat as the one we observe.

Both of these problems would disappear if the universe supercooled by 28 or more orders of magnitude below the critical temperature for some phase transition. (Under such circumstances, the universe would be growing exponentially in time.) However, the random formation of bubbles of the new phase seems to lead to a much too inhomogeneous universe.

The inhomogeneity problem would be solved if one could avoid the assumption that the nucleation rate $\lambda(t)$ approaches a small constant λ_0 as the temperature $T \rightarrow 0$. If instead the nucleation rate rose sharply at some T_1 , then bubbles of an approximately uniform size would suddenly fill space as T fell to T_1 . Of course, the full advantage of the inflationary scenario is achieved only if $T_1 \lesssim 10^{-28} T_c$.

Recently E. Witten³⁹ has suggested that the above chain of events may in fact occur if the parameters of the SU_5 Higgs field potential are chosen to obey the Coleman-Weinberg condition⁴⁰ (i.e., that $\partial^2 V / \partial \phi^2 = 0$ at $\phi = 0$) Witten⁴¹ has studied this possibility in detail for the case of the Weinberg-Salam phase transition. Here he finds that thermal tunneling is totally ineffective, but instead the phase transition is driven when the temperature of the QCD chiral symmetry breaking phase transition is reached. For the SU_5 case, one can hope that a much larger amount of supercooling will be found; however, it is difficult to see how 28 orders of magnitude could arise.

Another physical effect which has so far been left out of the analysis is the production of particles due to the changing gravitational

metric.⁴² This effect may become important in an exponentially expanding universe at low temperatures.

In conclusion, the inflationary scenario seems like a natural and simple way to eliminate both the horizon and the flatness problems. I am publishing this paper in the hope that it will highlight the existence of these problems and encourage others to find some way to avoid the undesirable features of the inflationary scenario.

Acknowledgments

I would like to express my thanks for the advice and encouragement I received from Sidney Coleman and Leonard Susskind, and for the invaluable help I received from my collaborators, Henry Tye and Erick Weinberg. I would also like to acknowledge very useful conversations with Michael Aizenman, Beilok Hu, Harry Kesten, Paul Langacker, Gordon Lasher, So-Young Pi, John Preskill, and Edward Witten. This work was supported by the Department of Energy under contract DE-AC03-76SF00515.

APPENDIX: REMARKS ON THE FLATNESS PROBLEM

This appendix is added in the hope that some skeptics can be convinced that the flatness problem is real.

Some physicists would rebut the argument given in Section I by insisting that the equations might make sense all the way back to $t = 0$. Then if one fixes the value of H corresponding to some arbitrary temperature T_a , one always finds that when the equations are extrapolated backward in time, $\Omega \rightarrow 1$ as $t \rightarrow 0$. Thus, they would argue, it is natural for Ω to be very nearly equal to 1 at early times. For

physicists who take this point of view, the flatness problem must be restated in other terms. Since H_0 and T_0 have no significance, the model universe must be specified by its conserved quantities. In fact, the model universe is completely specified by the dimensionless constant $\epsilon \equiv k/R^2 T^2$, where k and R are parameters of the Robertson-Walker metric, Eq. (2.1). For our universe, one must take $|\epsilon| < 3 \times 10^{-57}$. The problem then is to explain why $|\epsilon|$ should have such a startlingly small value.

Some physicists also take the point of view that $\epsilon \equiv 0$ is plausible enough, so to them there is no problem. To these physicists I point out that the universe is certainly not described exactly by a Robertson-Walker metric. Thus it is difficult to imagine any physical principle which would require a parameter of that metric to be exactly equal to zero.

In the end, I must admit that questions of plausibility are not logically determinable, and depend somewhat on intuition. Thus I am sure that some physicists will remain unconvinced that there really is a flatness problem. However, I am also sure that many physicists agree with me that the flatness of the universe is a peculiar situation which at some point will admit a physical explanation.

REFERENCES

1. I use units for which $\hbar = c = k$ (Boltzmann constant) = 1. Then
 $1 \text{ m} = 5.068 \times 10^{15} \text{ GeV}^{-1}$, $1 \text{ kg} = 5.610 \times 10^{26} \text{ GeV}$,
 $1 \text{ sec} = 1.519 \times 10^{24} \text{ GeV}^{-1}$, and $1^\circ\text{K} = 8.617 \times 10^{-14} \text{ GeV}$.
2. W. Rindler, Mon. Not. Roy. Ast. Soc. 116, 663 (1956). See also Ref. (3), pp. 489-90, 525-6; and Ref. (4), pp. 740 and 815.

3. S. Weinberg, Gravitation and Cosmology, J. Wiley and Sons (1972).
4. C. W. Misner, K. S. Thorne and J. A. Wheeler, Gravitation, W. H. Freeman and Company (1973).
5. In order to calculate the horizon distance, one must of course follow the light trajectories back to $t = 0$. This violates my contention that the equations are to be trusted only for $T \lesssim T_0$. Thus, the horizon problem could be obviated if the full quantum gravitational theory had a radically different behavior from the naive extrapolation. Indeed, a solution of this sort has been proposed by A. Zee, Phys. Rev. Lett. 44, 703 (1980). However, it is the point of this paper to show that the horizon problem can also be obviated by mechanisms which are more within our grasp, occurring at temperatures below T_0 .
6. R. H. Dicke and P.J.E. Peebles, General Relativity: An Einstein Centenary Survey, S. W. Hawking and W. Israel, eds., Cambridge University Press (1979).
7. See Ref. (3), pp. 475-81; and Ref. (4), pp. 796-7.
8. For example, see Ref. (3), Chapter 14.
9. M. Yoshimura, Phys. Rev. Lett. 41, 281 (1978), ibid. 42, 746 (1979), Phys. Lett. 88B, 294 (1979); S. Dimopoulos and L. Susskind, Phys. Rev. D 18, 4500 (1978), Phys. Lett. 81B, 416 (1979); A. Yu Ignatiev, N. V. Krashikov, V. A. Kuzmin, and A. N. 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); D. V. Nanopoulos and S. Weinberg, Phys. Rev. D 20, 2484 (1979); J. Ellis, M. K. Gaillard and D. V. Nanopoulos,

- Phys. Lett. 80B, 360 (1979), ibid. 82B, 464 (1979); M. Honda and M. Yoshimura, Prog. Theor. Phys. 62, 1704 (1979); D. Toussaint and F. Wilczek, Phys. Lett. 81B, 238 (1979); S. Barr, G. Segre, and H. A. Weldon, Phys. Rev. D 20, 2494 (1979); A. D. Sakharov, Sov. Phys. JETP 49, 594 (1979); A. Yu Ignatiev, N. V. Krashikov, V. A. Kuzmin and M. E. Shaposhnikov, Phys. Lett. 87B, 114 (1979); E. W. Kolb and S. Wolfram, Phys. Lett. 91B, 217 (1980) and CAL TECH Preprint 79-0956; J. N. Fry, K. A. Olive and M. S. Turner, Chicago University Preprint EFI-80-07; S. B. Treiman and F. Wilczek, Princeton Preprint 80-0255; G. Senjanović and F. W. Stecker, University of Maryland Preprint 80-197.
10. See Ref. (4), p. 411.
 11. The reason Λ is so small is of course one of the deep mysteries of physics. The value of Λ is not determined by the particle theory alone, but must be fixed by whatever theory couples particles to quantum gravity. This appears to be a separate problem from the ones discussed in this paper, and I merely use the empirical fact that $\Lambda \approx 0$.
 12. S. A. Bludman and M. A. Ruderman, Phys. Rev. Lett. 38, 255 (1977).
 13. The effects of a false vacuum energy density on the evolution of the early universe have also been considered by E. W. Kolb and S. Wolfram, "Spontaneous Symmetry Breaking and the Expansion Rate of the Early Universe," CAL TECH Preprint 79-0984, and by S. A. Bludman, "Elementary Particle Symmetry Restoration Can Prevent a Cosmological Singularity," University of Pennsylvania Preprint UPR-0143T (1979).

14. More precisely, the usual invariance is $O(4)$ if $k = 1$, $O(3,1)$ if $k = -1$, and the group of rotations and translations in three dimensions if $k = 0$.
15. See for example, Ref. 3, pp. 385-92.
16. A. H. Guth and S.-H. H. Tye, Phys. Rev. Lett. 44, 631, 963 (1980).
17. S. Coleman, Phys. Rev. D 15, 2929 (1977); C. G. Callan and S. Coleman, ibid. D 16, 1762 (1977); see also S. Coleman, "The Uses of Instantons," lectures delivered at the 1977 International School of Subnuclear Physics, Ettore Majorana.
18. A. H. Guth and E. J. Weinberg, "A Cosmological Lower Bound on the Higgs Boson Mass," SLAC-PUB-2525 (1980).
19. E. J. Weinberg and I are preparing a manuscript on the possible cosmological implications of the phase transitions in the SU_5 grand unified model.
20. J. Ellis and G. Steigman, Phys. Lett. 89B, 186 (1980); J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, ibid. 90B, 253 (1980).
21. B. L. Hu and L. Parker, Phys. Rev. D 17, 933 (1978).
22. See Ref. (4), Chapter 30.
23. The simplest grand unified model is the $SU(5)$ model of H. Georgi and S. L. Glashow, Phys. Rev. Lett. 32, 438 (1974). See also H. Georgi, H. R. Quinn and S. Weinberg, Phys. Rev. Lett. 33, 451 (1974); and A. J. Buras, J. Ellis, M. K. Gaillard and D. V. Nanopoulos, Nucl. Phys. B135, 66 (1978).
24. Other grand unified models include the $SO(10)$ model: H. Georgi, in Particles and Fields, H. J. Lubatti and P. M. Mochett, eds., American Physical Society (1975); H. Fritzsch and P. Minkowski,

- Ann. Phys. 93, 193 (1975); H. Georgi and D. V. Nanopoulos, Phys. Lett. 82B, 392 (1979) and Nucl. Phys. B155, 52 (1979); the E(6) model: F. Gürsey, P. Ramond, and P. Sikivie, Phys. Lett. B60, 177 (1975); F. Gürsey and M. Serdaroglu, Nuovo Cim. Lett. 21, 28 (1978); and the E(7) model: F. Gürsey and P. Sikivie, Phys. Rev. Lett. 36, 775 (1976), and Phys. Rev. D 16, 816 (1977); P. Ramond, Nucl. Phys. B110, 214 (1976); for some general properties of grand unified models, see M. Gell-Mann, P. Ramond and R. Slansky, Rev. Mod. Phys. 50, 721 (1978). For a review, see P. Langacker, "Grand Unified Theories and Proton Decay," SLAC-PUB-2544 (1980).
25. D. A. Kirzhnits and A. D. Linde, Phys. Lett. 42B, 471 (1972).
26. S. Weinberg, Phys. Rev. D 9, 3357 (1974); L. Dolan and R. Jackiw, *ibid.* D 9, 3320 (1974); see also D. A. Kirzhnits and A. D. Linde, Ann. of Phys. 101, 195 (1976); A. D. Linde, Rep. Prog. Phys. 42, 389 (1979). ϵ -expansion techniques are employed by P. Ginsparg, "First and Second Order Phase Transitions in Gauge Theories at Finite Temperature," CEN Saclay Preprint DPh-T/80/27, submitted to Nucl. Phys. B.
27. In the case that the Higgs quartic couplings are comparable to g^4 or smaller (g = gauge coupling), the phase structure has been studied by M. Daniel and C. E. Vayonakis, "The Phase Transition in the SU(5) Model at High Temperatures," CERN Preprint TH.2860 (1980); and by P. Suranyi, "Phase Diagram of Broken Gauge Theories," University of Cincinnati Preprint 80-0506.
28. G. 't Hooft, Nucl. Phys. B79, 276 (1974); A. M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. 20, 430 (1974) [JETP Lett. 20, 194 (1974).]

- For a review, see P. Goddard and D. I. Olive, Rep. Prog. Phys. 41, 1357 (1978).
29. If $\Pi_1(G)$ and $\Pi_2(G)$ are both trivial, then $\Pi_2(G/H_0) = \Pi_1(H_0)$. In our case $\Pi_1(H_0)$ is the group of integers. For a general review of topology written for physicists, see N. D. Mermin, Rev. Mod. Phys. 51, 591 (1979).
30. Y. B. Zeldovich and M. Y. Khlopov, Phys. Lett. 79B, 239 (1978).
31. J. P. Preskill, Phys. Rev. Lett. 43, 1365 (1979).
32. T.W.B. Kibble, J. Phys. A: Math. Gen. 9, 1387 (1976).
33. This argument was first shown to me by John Preskill. It is also described by Einhorn *et al.*, Ref. 34, except that they make no distinction between T_{coal} and T_c .
34. The problem of monopole production was also examined by M. B. Einhorn, D. L. Stein and D. Toussaint, Phys. Rev. D 21, 3295 (1980), who focused on second order transitions. The structure of SU(5) monopoles has been studied by C. P. Dokos and T. N. Tomaras, Phys. Rev. D 21, 2940 (1980); and by M. Daniel, G. Lazarides and Q. Shafi, "SU(5) Monopoles, Magnetic Symmetry and Confinement," CERN Preprint TH.2800 (1979). The problem of suppression of the cosmological production of monopoles is discussed by G. Lazarides and Q. Shafi, "The Fate of Primordial Magnetic Monopoles," CERN Preprint TH.2821 (1980), and G. Lazarides, M. Magg and Q. Shafi, "Phase Transitions and Magnetic Monopoles in SO(10)," CERN Preprint TH.2856 (1980); the suppression discussed here relies on a novel confinement mechanism, and also on the same kind of supercooling as in Ref. (16). See also J. N. Fry and D. N. Schramm, Phys. Rev. Lett. 44, 1361 (1980).

35. An alternative solution to the monopole problem has been proposed by S.-Y. Pi and P. Langacker, Phys. Rev. Lett. 45, 1 (1980). By modifying the Higgs structure, they have constructed a model in which the high temperature SU_5 phase undergoes a phase transition to an SU_3 phase at $T \sim 10^{14}$ GeV. Another phase transition occurs at $T \sim 10^3$ GeV, and below this temperature the symmetry is $SU_3 \times U_1^{EM}$. Monopoles cannot exist until $T < 10^3$ GeV, but their production is negligible at these low temperatures. The suppression of monopoles due to the breaking of U_1^{EM} symmetry at high temperatures was also suggested by S.-H. H. Tye, talk given at the "1980 Guangzhou (Canton) Conference on Theoretical Particle Physics," China (1980).
36. The Weinberg-Salam phase transition has also been investigated by a number of authors: E. Witten, Ref. 41; M. A. Sher, "Effects of Phase Transitions on the Evolution of the Early Universe," U. C. Santa Cruz Preprint 80/132; P. J. Steinhardt, "The Weinberg-Salam Model in Early Cosmology," Harvard Preprint (1980); and A. H. Guth and E. J. Weinberg, Ref. (18).
37. This section represents the work of E. J. Weinberg, H. Kesten and myself. Weinberg and I are preparing a manuscript on this subject.
38. The proof of this statement was outlined by H. Kesten (Dept. of Mathematics, Cornell University), with details completed by me.
39. E. Witten, private communication.
40. S. Coleman and E. J. Weinberg, Phys. Rev. D 7, 1888 (1973); see also J. Ellis, M. K. Gaillard, D. Nanopoulos and C. Sachrajda, Phys. Lett 83B, 339 (1979), and J. Ellis, M. K. Gaillard, A. Peterman, and C. Sachrajda, Nucl. Phys. B164, 253 (1980).

41. E. Witten, "Cosmological Consequences of a Light Higgs Boson," Harvard Preprint HUTP-80/A040.
42. L. Parker, in Asymptotic Structure of Spacetime, F. Esposito and L. Witten, eds. (Plenum, N.Y.), 1977; V. N. Lukash, I. D. Novikov, A. A. Starobinsky and Ya. B. Zeldovich, Nuovo Cimento 35B, 293 (1976).