PHYSICAL REVIEW D VOLUME 23, NUMBER 4 15 FEBRUARY 1981

Cosmological consequences of a first-order phase transition in the SU_5 grand unified model

Alan H. Guth

*Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139**
and Stanford Linear Accelerator Center—P.O. Box 4349, Stanford, California 94305

Erick J. Weinberg

*Department of Physics, Columbia University, New York, New York 10027**
and Stanford Linear Accelerator Center—P.O. Box 4349, Stanford, California 94305

(Received 3 November 1980)

The SU_5 grand unified model is considered in the context of big-bang cosmology. If the Higgs potential contains a cubic term there is a first-order transition to the $SU_3 \times SU_2 \times U_1$-symmetric phase. The bubble nucleation rate for this transition is calculated. For typical choices of parameters, the transition proceeds according to one of two scenarios depending on whether or not its rate is ever large relative to the expansion rate of the universe. Both possibilities lead to difficulties: In the former case the transition is rapidly completed, but leads to the production of too many superheavy magnetic monopoles. In the latter case monopole production is suppressed, but there is an extreme supercooling from which the universe never recovers.

I. INTRODUCTION

There has recently been considerable interest in grand unified theories of the strong, weak, and electromagnetic interactions. Characteristic of these is that symmetry breaking occurs at an energy scale of $\sim 10^{14}$ GeV. Clearly, such energies are unattainable in the laboratory. However, the standard model of cosmology suggests that the corresponding temperatures were reached shortly after the big bang. One may therefore hope to gain some insight by studying the interplay between elementary particle interactions at very high energies and the expansion of the early universe. Indeed, such an approach has already led to a possible explanation for the observed baryon number to entropy ratio.[1]

Spontaneously broken symmetries are usually restored at sufficiently high temperatures. Thus we expect the universe to have begun in a state in which the grand unified gauge symmetry was manifest and, as it expanded and cooled, to have undergone a number of phase transitions before reaching the present state of broken symmetry. Of particular interest is the possibility that one or more of these was a first-order phase transition. Such transitions proceed by the formation and growth of bubbles of the new phase and can be relatively slow. When they occur in a rapidly expanding universe there is the potential for a considerable degree of supercooling. It has been suggested that such supercooling may provide a mechanism for suppressing the production of superheavy magnetic monpoles[2,3] and may also lead to a solution of the horizon and flatness problems of cosmology.[4]

In this paper we study one such first-order transition in some detail. We consider the SU_5 model of Georgi and Glashow,[5,6] in which the grand unified symmetry is broken by means of an adjoint-representation Higgs field ϕ. In Sec. II we discuss the various possible patterns of symmetry breaking and determine the range of parameters for which the desired $SU_3 \times SU_2 \times U_1$ is obtained. The effects of high temperature are discussed in Sec. III. We show that for a wide range of parameters the universe does not go directly from the SU_5-symmetric phase to the $SU_3 \times SU_2 \times U_1$-symmetric one, but instead passes through an $SU_4 \times U_1$-symmetric phase. It is the transition out of this intermediate phase on which we concentrate. In Sec. IV we use the zero-temperature methods of Callan and Coleman[6,7] to calculate the rate at which bubbles of the new phase form. In Sec. V we show how these methods must be modified in order to obtain the bubble nucleation rate at high temperature. In Sec. VI the results of the previous two sections are used to determine the degree to which the expanding universe supercools. We find that two quite different types of behavior are possible. In Sec. VII we discuss the implications of each and make some concluding remarks. There is an appendix containing the proof of a formula used in Sec. VI.

II. PHASE STRUCTURE

In this section we discuss the patterns of symmetry breaking which can arise from the adjoint-representation Higgs field. We elaborate the results of Ref. 2, which in turn elaborates the results of Li.[8]

The most general renormalizable potential describing the self-interaction of the adjoint-representation Higgs field[9] is [10]

$$V = -\frac{\mu^2}{2}\mathrm{Tr}\phi^2 + \frac{a}{4}(\mathrm{Tr}\phi^2)^2 + \frac{b}{2}\mathrm{Tr}\phi^4 + \frac{c}{3}\mathrm{Tr}\phi^3 . \tag{2.1}$$

It is convenient to define the dimensionless variables

$$\beta = \frac{\mu^2 b}{c^2} \tag{2.2}$$

and

$$\gamma = \frac{a}{b} + \frac{7}{15} . \tag{2.3}$$

If we also define a rescaled field $\chi = (b/c)\phi$, then the potential may be written as

$$\begin{aligned} V &= \frac{c^4}{b^3} U , \\ U &= -\frac{\beta}{2}\mathrm{Tr}\,\chi^2 + \tfrac{1}{4}(\gamma - \tfrac{7}{15})(\mathrm{Tr}\,\chi^2)^2 + \tfrac{1}{2}\mathrm{Tr}\,\chi^4 \\ &\quad + \tfrac{1}{3}\mathrm{Tr}\,\chi^3 . \end{aligned} \tag{2.4}$$

In order that V have a minimum corresponding to symmetry breakdown to $SU_3 \times SU_2 \times U_1$, we must take $b > 0$. Positivity of the quartic terms then requires $\gamma > 0$.

The stable and metastable vacuums are determined by the global and local minima of the potential; each exists only for a limited region of parameter space. With $b > 0$, the following occur[11]:

I. $\phi = (\lambda_I/\sqrt{30})\,\mathrm{diag}(2, 2, 2, -3, -3)$,

$$\lambda_I = \frac{c}{b}\left(\frac{\beta}{\gamma}\right)^{1/2}\left[\left(1 + \frac{1}{120\beta\gamma}\right)^{1/2} + \frac{1}{(120\beta\gamma)^{1/2}}\right]$$
$$\equiv \frac{c}{b}\left(\frac{\beta}{\gamma}\right)^{1/2} h(\beta\gamma) .$$

The unbroken symmetry is $SU_3 \times SU_2 \times U_1$. This is a local minimum if

$$\beta > \begin{cases} \frac{15}{32}(\gamma - \frac{4}{15}), & \gamma > \frac{2}{15} \\ -\frac{1}{120\gamma}, & \gamma < \frac{2}{15} . \end{cases}$$

II. $\phi = (\lambda_{II}/\sqrt{20})\,\mathrm{diag}(1, 1, 1, 1, -4)$,

$$\lambda_{II} = \frac{c}{2b(\gamma + 5/6)}\left(\frac{3}{\sqrt{20}} + [\tfrac{9}{20} + 4\beta(\gamma + 5/6)]^{1/2}\right) .$$

The unbroken symmetry is $SU_4 \times U_1$. This is a local minimum if

$$-\frac{9}{80(\gamma + 5/6)} < \beta < \tfrac{5}{4}(\gamma + \tfrac{7}{30}) .$$

III. $\phi = 0$.

There is no symmetry breaking. This is a local minimum for $\beta < 0$.

IV. The same as I, but with λ_I replaced by

$$\lambda_{IV} = -\frac{c}{b}\left(\frac{\beta}{\gamma}\right)^{1/2}\left[\left(1 + \frac{1}{120\beta\gamma}\right)^{1/2} - \frac{1}{(120\beta\gamma)^{1/2}}\right] .$$

This is a local minimum if

$$\beta > \tfrac{15}{2}(\gamma + \tfrac{1}{15}) .$$

The phase diagram in Fig. 1 indicates the global minimum for various values of β and γ. The boundary between regions I and III is given by $\beta^{-1} = -135\gamma$, and that between regions II and III by $\beta^{-1} = -10(\gamma + \frac{5}{6})$. The boundary between regions I and II asymptotically approaches the line $\beta = 0.610\gamma - 0.079$. Note that solution IV is never the global minimum; it will play no further role in our considerations.

We require that the parameters be such as to give solution I as the global minimum of the potential. Twelve of the gauge bosons then acquire a mass

$$M_X = (\tfrac{5}{12})^{1/2} g\lambda . \tag{2.5}$$

Here g is the gauge coupling constant and $\lambda \equiv \lambda_I$. One expects $M_X \sim 10^{14}$ GeV. The surviving Higgs particles separate into an $SU_3 \times SU_2$ singlet, an SU_3 singlet–SU_2 triplet, and an SU_3 octet–SU_2 singlet, with masses

$$\begin{aligned} m_0^2 &= 2b\gamma\left[1 - \frac{1}{1 + (1 + 120\beta\gamma)^{1/2}}\right]\lambda^2 , \\ m_3^2 &= \left[\frac{4}{3} - \frac{5}{\sqrt{30}}\left(\frac{\gamma}{\beta}\right)^{1/2}\frac{1}{h(\beta\gamma)}\right]b\lambda^2 , \\ m_8^2 &= \left[\frac{1}{3} + \frac{5}{\sqrt{30}}\left(\frac{\gamma}{\beta}\right)^{1/2}\frac{1}{h(\beta\gamma)}\right]b\lambda^2 , \end{aligned} \tag{2.6}$$

respectively.[12]

Finally, let us comment on the magnitude of the parameters which appear in the potential. While renormalization-group methods can be used to obtain λ and the gauge coupling constant (at the grand unification mass $\alpha \sim 0.02$), there is little that we can say about the remaining parameters. However, the validity of our approach requires on the one hand that the coupling constants not be so large as to invalidate perturbation theory, and on the other hand that the scalar quartic couplings not be so weak that they are dominated by the α^2

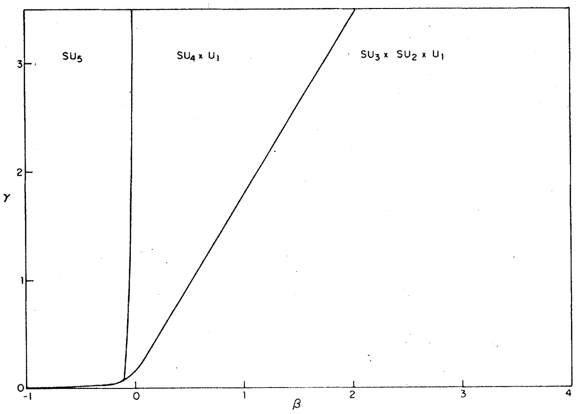

FIG. 1. Phase diagram for the adjoint Higgs system. The triple point is at $\gamma=\frac{1}{15}$, $\beta=-\frac{1}{9}$.

terms arising from vector-boson loop contributions to the effective potential.

III. HIGH-TEMPERATURE BEHAVIOR

In order to study high-temperature behavior, V must be modified to include thermal effects. At temperatures large compared to all masses, the finite-temperature effective potential is of the same form as V, but with β replaced by[13,14]

$$\beta_{\rm eff}(T)=\beta-\frac{b\sigma}{c^2}T^2, \tag{3.1}$$

where[15]

$$\sigma=\tfrac{1}{60}(130a+94b+75g^2). \tag{3.2}$$

Thus, at very high temperatures the universe would be in the unbroken SU_5 phase. As it cooled, it would move horizontally to the right across the phase diagram in Fig. 1, with transitions first to phase II (if $\gamma>\frac{1}{15}$) and then to phase I. In both cases the phase transition would be first order and would proceed by the formation and growth of bubbles of critical size, a relatively slow process. At the same time, the expansion of the universe would cause the temperature to continue to fall. However, this supercooling cannot proceed below the temperature at which the old phase ceases to be metastable; at this temperature the critical bubble size goes to zero and the old phase rapidly disappears.

For the transition out of phase III there are several possibilities. If $\gamma>\frac{4}{15}$, phases I and III are never simultaneously metastable, so the transition is into phase II and is completed by the time that the line $\beta=0$ is crossed. For $\gamma<\frac{4}{15}$ and $\beta>-\frac{9}{80}(\gamma+\frac{5}{6})^{-1}$, there are temperatures at which all three phases are metastable, so the transition will involve competition between formation of regions of phases I and those of phase II. For $\beta<\frac{9}{80}(\gamma+\frac{5}{6})^{-1}$, phase II does not exist, so the transition will be directly to phase I.

We therefore divide the $SU_3\times SU_2\times U_1$ region of the phase diagram into several parts (see Fig. 2) and, according to where the zero-temperature parameters lie, distinguish the following types of behavior.

(a) The III–II transition is completed before the critical temperature for the II–I transition is reached. The latter is slow and may entail considerable supercooling.

(b) There is competition between direct III–I transition and transition via phase II. If $\beta<0$, the transitions are both slow, while if $\beta>0$ there will be rapid transition out of phase III when the line $\beta=0$ is crossed.

(c) There is a direct III–I transition with the possibility of extreme supercooling.

(d) If $\beta>\frac{5}{4}(\gamma+\frac{7}{30})$, neither phase II nor phase III is metastable at $T=0$, so there will be a temperature T' by which the transition to phase I will be completed and below which supercooling cannot persist.

For the remainder of this paper we will concentrate on case (a), rather than on (b) and (c), which require rather fine tuning of parameters, or on (d), which allows only limited supercooling. Furthermore, we will focus our attention on the II–I phase transition.

The above discussion has been based on the high-temperature approximation for the effective potential. However, using Eq. (2.1) to determine the critical temperature gives

$$T_c=\left[\frac{b\gamma}{\sigma}\left(1-\frac{\beta_c}{\beta}\right)\right]^{1/2}\frac{1}{h(\beta\gamma)}\lambda\,. \tag{3.3}$$

[Here $\beta_c(\gamma)$ is the value of β at the I–II phase

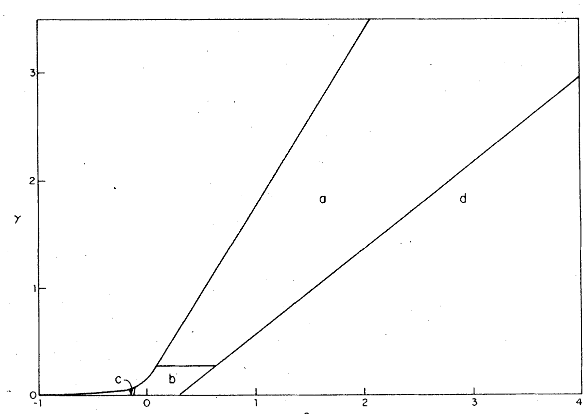

FIG. 2. Parts of the $SU_3\times SU_2\times U_1$ region corresponding to the types of phase-transition behavior discussed in the the text.

boundary.] For typical choices of parameters, this is of the same order of magnitude as M_X, casting some doubt on the validity of the approximation. It is therefore reassuring to note that the critical temperature can also be calculated using the massless ideal quantum gas approximation.[2] For $T \ll M_X$ (but much greater than the Weinberg-Salam mass scale), the free energy density may be written as

$$F = V - \frac{\pi^2}{90}\mathfrak{N}T^4 , \tag{3.4}$$

where $\mathfrak{N}$ is the number of effectively massless degrees of freedom, with fermion degrees of freedom counting $\frac{7}{8}$. Since $\mathfrak{N}_{II} = \mathfrak{N}_I + 8$, this gives

$$T_c = \left[\frac{45}{4\pi^2}(V_{II} - V_I)\right]^{1/4} , \tag{3.5}$$

which differs from Eq. (3.3) by a factor which we write as $(\sigma^2/b\gamma^2)^{1/4}k(\beta,\gamma)$. Except in a narrow strip near the phase boundary, $k(\beta,\gamma)$ lies between $\frac{1}{2}$ and 2,[16] so there is reasonable agreement between the two approximations. Equation (3.5) is displayed graphically in Fig. 3.

IV. BUBBLE NUCLEATION AT ZERO TEMPERATURE

At zero temperature, the transition from false vacuum to true begins with a purely quantum-mechanical process in which a finite region of false vacuum tunnels through the potential barrier to form a bubble of greater than critical size. Callan and Coleman[7] have shown that the rate of this process may be obtained by solving the Euler-Lagrange equations of the theory in four-dimensional Euclidean space with the boundary condition that the fields approach the false vacuum at infinity. The probability per unit time per unit volume of bubble nucleation is then given by

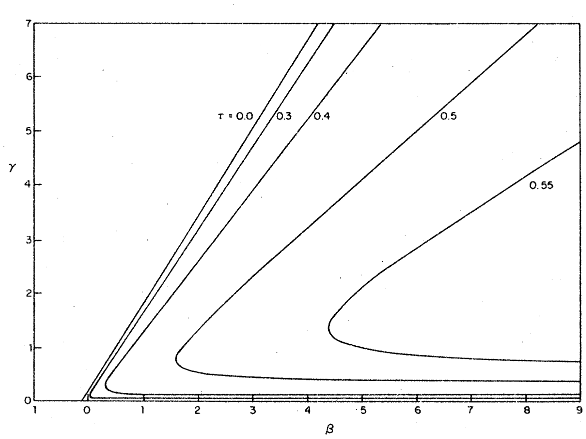

FIG. 3. Critical temperature as a function of β and γ. The diagram shows contours of constant $\tau = b^{-1/4}T_c/\lambda$, with T_c calculated using Eq. (3.5).

$$f_0 = Ae^{-B_0} , \tag{4.1}$$

where B_0 is the Euclidean action corresponding to the tunneling solution with least action. The determination of A requires calculation of radiative corrections; it is expected to be of the form $\eta\mathfrak{M}^4$, where $\mathfrak{M}$ is a characteristic mass of the theory and η is a dimensionless number of order unity. For definiteness, $\mathfrak{M} \equiv T_c$.

We now wish to determine the exponent B_0 for the transition from phase II to phase I. It is a formidable task to find the general solution to the field equations, so we will seek only solutions with a high degree of symmetry. It is plausible that these are the solutions of lowest action, but we have not been able to prove this. To begin, we consider only configurations which are O(4)-symmetric and which involve only trivial gauge fields; i.e., for which there exists a gauge in which the Yang-Mills field vanishes everywhere. This reduces the field equations to

$$\ddot{\phi}_a + \frac{3}{s}\dot{\phi}_a = \frac{\partial V}{\partial \phi_a} , \tag{4.2}$$

where dots denote differentiation with respect to $s = (x^2 + t^2)^{1/2}$. The boundary conditions are that $\phi(\infty)$ correspond to the $SU_4 \times U_1$-symmetric false vacuum [without any loss of generality we require $\phi(\infty) \sim \text{diag}(1, 1, 1, 1, -4)$] and, in order that ϕ be nonsingular at the origin, $\dot{\phi}(0) = 0$. Finally, we distinguish tunneling solutions by the requirement that $\phi(0)$ be on the same side of the potential barrier as the $SU_3 \times SU_2 \times U_1$-symmetric true vacuum.

Even the restrictions we have made thus far are not sufficient to make the problem tractable; we therefore make the further assumption that $\phi(s)$ may be taken to be diagonal not only at $s = \infty$, but for all s. Two types of solutions must then be distinguished: those which tunnel from $\phi \sim \text{diag}(1, 1, 1, 1, -4)$ toward the vacuum state $\phi \sim \text{diag}(2, 2, 2, -3, -3)$ and those which tunnel toward $\phi \sim \text{diag}(-3, -3, 2, 2, 2)$. In both cases the equations of motion allow us to impose the further requirement that $\phi(s)$ have at most three distinct eigenvalues for each s; this gives an $SU_3 \times U_1 \times U_1$-symmetric solution in the first case and an $SU_2 \times SU_2 \times U_1$-symmetric solution in the second. With this last symmetry requirement, the number of independent components of ϕ is reduced to two, and it becomes feasible to find the appropriate solution to Eq. (4.2) by having a computer search for a $\phi(0)$ such that $\phi(s)$ approaches the correct asymptotic value as s tends towards infinity.

The value thus obtained for B depends on the three dimensionless variables β, γ, and b. The dependence on the last of these is particularly simple. By rescaling coordinates according to

$$x^\mu \to y^\mu = \frac{c}{\sqrt{b}} x^\mu \,, \tag{4.3}$$

we can write the action as

$$B = \frac{1}{b} \int d^4y \left[\tfrac{1}{2} \,\mathrm{Tr}\left(\frac{\partial \chi}{\partial y_\mu}\right)^2 + U(\chi) \right], \tag{4.4}$$

where χ and U are the rescaled field and potential defined in Sec. II. Since the latter depends only on β and γ, we may write

$$B = \frac{1}{b} I_4(\beta, \gamma)\,. \tag{4.5}$$

The computer calculations show that the $SU_3 \times U_1 \times U_1$-symmetric solution has a lower action than the $SU_2 \times SU_2 \times U_1$-symmetric one,[17] and gives the values of $I_4(\beta,\gamma)$ shown in Fig. 4.

V. BUBBLE NUCLEATION AT HIGH TEMPERATURE

In the previous section, the method of Callan and Coleman was used to determine the zero-temperature rate of nucleation of bubbles of true vacuum in an $SU_4 \times U_1$-symmetric false vacuum. In this section, we discuss nucleation at high temperatures; i.e., at temperatures of the order of T_c. At these temperatures critical bubbles can be formed not only by quantum-mechanical barrier penetration, but also as a result of thermal fluctuations. Also, a correct formulation of the problem should predict no nucleation by either mechanism for $T > T_c$.

At $T = 0$, the bubble nucleation rate is obtained from the imaginary part of the energy density; at finite temperature one must use the imaginary part of the free energy.[18] In the path-integral formalism used by Callan and Coleman, this is done by restricting the integration to configurations with periodicity $1/T$ in imaginary time. Making this change, and proceeding as before, we would find that for sufficiently large T the corrections due to the prefactor A would not be small relative to B, thus invalidating the perturbative scheme on which the calculation is based. The failure of perturbation theory at high temperature has been investigated by S. Weinberg.[13] He showed that the validity of the perturbation expansion may be restored (for T large compared to all masses) by a redefinition of the mass terms in the Lagrangian, while compensating for the change by the addition of appropriate counterterms. In the case at hand, this corresponds to using an effective potential in which β is replace by the β_{eff} defined in Eq. (3.1).

As with the zero-temperature calculation, we expect the solution with least action to have a high degree of symmetry; however, because of the periodicity requirement, we can only impose O(3)—rather than O(4)—symmetry. For large T, we expect the solution to be independent of t, giving an exponent of the form $B(T) = E(T)/T$, where $E(T)$ is the energy, calculated using the effective potential, of a bubble of critical size.[19,20] On the other hand, for temperatures sufficiently low that T^{-1} is large compared to the critical bubble radius, we expect solutions with approximate O(4) symmetry and with $B \approx B_0$ to dominate.

Turning now to the explicit calculation of $E(T)$, we are faced with two difficulties. The first, which was also present in the $T = 0$ calculation, is that even with the assumption of spherical symmetry, the field equations involve too many independent fields. We deal with this by making the same simplifying ansatz for the fields as before. The second is that we must use not V, but rather a finite-temperature effective potential in the field equations. At high temperatures this only requires that β be replaced by the $\beta_{eff}(T)$ defined by Eq. (3.1). This prescription is also valid for

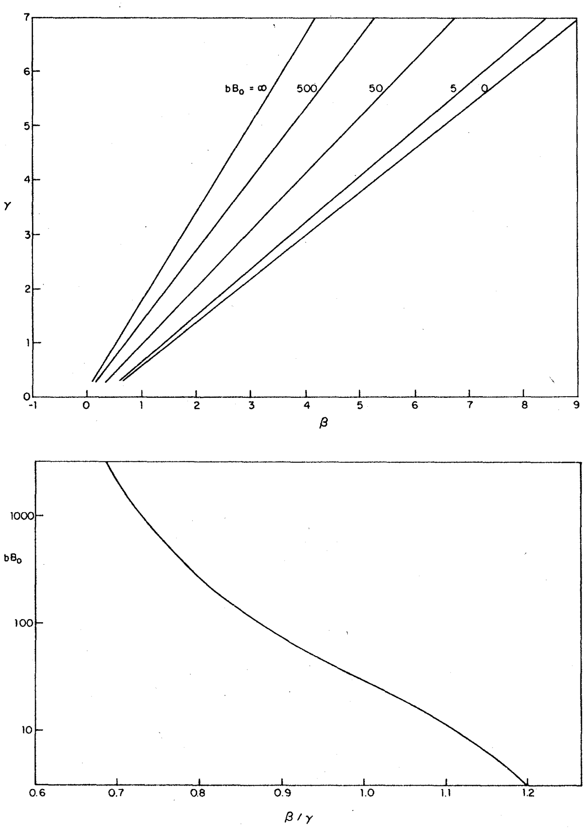

FIG. 4. Behavior of $I_4 = bB_0$ in part (a) of the $SU_3 \times SU_2 \times U_1$ region of the phase diagram. (a) Contours of constant I_4. (b) Variation of I_4 as β/γ is varied with γ held fixed. The data shown correspond to the limit $\gamma \to \infty$. In this limit I_4 vanishes at $\beta/\gamma = 1.25$ and $I_4 = \infty$ at $\beta/\gamma = 0.6103$.

temperatures small compared to T_c, in which case thermal corrections are negligible and $\beta_{\rm eff} \approx \beta$. However, there is an intermediate range (which for certain choices of parameters may include T_c) in which this approximation is of doubtful validity.[21] Nevertheless, let us proceed for the moment using it; we will return later to the question of the reliability of the results.

By using the same coordinate rescaling as in the previous section, one obtains

$$E(T)=\frac{c}{b^{3/2}}I_3(\beta_{\rm eff},\gamma)\,, \tag{5.1}$$

where I_3 is the three-dimensional analog of the integral in Eq. (4.4). Figure 5 illustrates the behavior of I_3. Note first of all the divergence as T_c is approached from below; thus, as required, there is no nucleation for $T>T_c$. Note also the rapid decrease as $\beta_{\rm eff}$ increases. We expect this decrease of the critical bubble energy to persist under an exact treatment, although the detailed temperature dependence should not be trusted beyond the range of validity of the high-temperature approximation. Since the thermal effects decrease rather rapidly with temperature [see, e.g., Eq. (3.4)], E will in general be well approximated by its zero-temperature value for $T \lesssim T_c/4$. Thus $B(T)$ should reach a minimum (corresponding to a maximum for the nucleation rate) at a temperature $T^* \gtrsim T_c/4$.[22] Its value at this minimum then lies in the range

$$\frac{E(0)}{\frac{1}{4}T_c} > B(T^*) > \frac{E(0)}{T_c}\,. \tag{5.2}$$

The behavior of $b^{3/4}\,E(0)/T_c$ is shown in Fig. 6. By comparing this figure with Fig. 4(b) and using the relation (5.2), we see that $B(T^*)$ is roughly an order of magnitude smaller than B_0. The consequent sharp decrease in nucleation rate as the temperature falls from T^* can have a significant effect on the development of the phase transition, a subject to which we now turn.

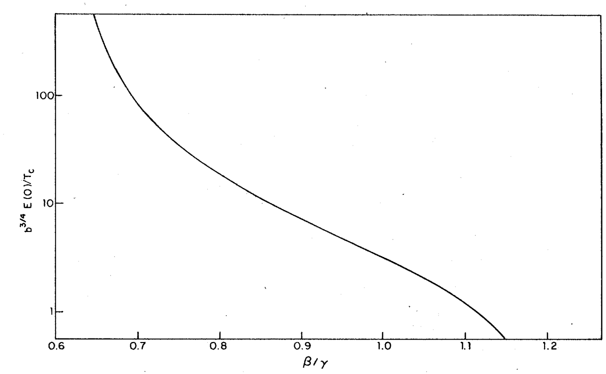

FIG. 6. Variation of $E(0)/T_c$ as β/γ is varied with γ held fixed. The data shown correspond to the limit $\gamma\to\infty$.

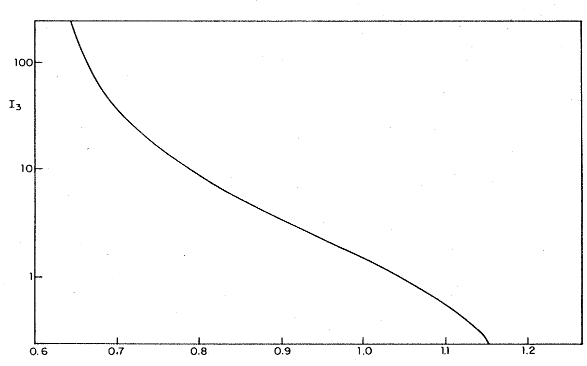

FIG. 5. Variation of I_3 as $\beta_{\rm eff}/\gamma$ is varied with γ held fixed. The data shown correspond to the limit $\gamma\to\infty$.

VI. DEVELOPMENT OF THE PHASE TRANSITION

The progress of the phase transition depends on the rate of expansion of the universe. A homogeneous isotropic universe can be described by a Robertson-Walker metric (in comoving coordinates

$$d\tau^2 = d\tau^2 - R^2(t)\left(\frac{dr^2}{1-kr^2}+r^2 d\,\Omega^2\right)\,. \tag{6.1}$$

Here $k=1$, 0, or -1 according to whether the universe is closed, flat, or open. Its expansion is governed by the equation

$$\left(\frac{\dot R}{R}\right)^2=\frac{8\pi}{3{M_P}^2}\rho-\frac{k}{R^2}\,, \tag{6.2}$$

where $M_P=1.2\times10^{19}$ GeV is the Planck mass. The energy density may be written as

$$\rho=\rho_0+\left(\frac{\pi^2}{30}\right)\mathfrak{N}T^4 \tag{6.3}$$

with $\mathfrak{N}$ being the number of effectively massless degrees of freedom, with fermion degrees of freedom counting $\frac{7}{8}$. The vacuum energy density ρ_0 is equivalent to a cosmological constant; to agree with observation it must be taken to be very close to zero in the present phase of the universe, so in the $SU_4\times U_1$-symmetric phase,

$$\rho_0 \approx V_{\rm II} - V_{\rm I}\,. \tag{6.4}$$

At sufficiently small temperatures, the energy density is dominated by ρ_0, leading to an R which grows exponentially with time. Taking $\mathfrak{N}\approx 10^2$ and using Eq. (3.5), we see that this occurs when $T \lesssim 0.4T_c$.

Since a universe containing bubbles of one phase expanding within another is hardly homogeneous, it might seem that these formulas are inapplicable after the onset of bubble nucleation. However, the bubbles expand at a speed which rapidly approaches that of light.[7] Therefore the region outside the bubbles cannot be affected by their presence and can be described by a Robertson-Walker metric obeying Eqs. (6.1)–(6.4).

In the Appendix, we show that the fraction of space remaining in the old phase at time t is[2]

$$p(t)=\exp\left[-\int_{t_0}^{t} dt_1 f(t_1)R^3(t_1)V(t_1,t)\right], \qquad (6.5)$$

where

$$V(t_1,t)=\frac{4\pi}{3}\left[\int_{t_1}^{t} dt_2\frac{1}{R(t_2)}\right]^3 \qquad (6.6)$$

is the coordinate volume at time t of a bubble formed at time t_1, and $f(t)$ is the rate of bubble nucleation per unit time per unit physical volume. It is convenient to reexpress this in terms of temperature. We assume adiabatic expansion and take $\mathfrak{N}$ as a constant, so $RT=\text{constant}$. Neglecting the second term on the right-hand side of Eq. (6.2), one obtains

$$\dot{T}=-\chi T g(T),$$

where

$$\chi=\left(\frac{8\pi}{3}\frac{\rho_0}{M_P{}^2}\right)^{1/2} \qquad (6.7)$$

and

$$g(T)=\left(1+\frac{\pi^2}{30}\mathfrak{N}\frac{T^4}{\rho_0}\right)^{1/2}.$$

Note that $g(T)\approx 1$ for small T. Using the nucleation rates obtained in Secs. IV and V gives

$$p(T)=\exp\left\{-d\int_T^{T_c} dT_1\frac{e^{-B(T_1)}}{g(T_1)T_1{}^4}\left[\int_T^{T_1}\frac{dT_2}{g(T_2)}\right]^3\right\} \qquad (6.8)$$

with

$$d=\frac{4\pi}{3}\eta\left(\frac{T_c}{\chi}\right)^4 =0.078\eta\left(\frac{M_P}{T_c}\right)^4. \qquad (6.9)$$

[Eq. (3.5) has been used in obtaining the second equality.] Thus for $T_c\approx 10^{14}$ GeV, $d\approx 10^{19}$. We have seen that, as T falls from T_c, $B(T)$ decreases to a minimum at T^* and then rises, leveling off at B_0. The potentially dominant contributions to the T_1 integral in Eq. (6.8) come from the regions $T_1\approx T^*$ and $T_1\approx 0$. Approximating the integral by the sum of these gives (for $T<T^*$)

$$p(T)\approx\exp\left\{-d\left[\frac{\sqrt{2\pi}\,\xi e^{-B(T^*)}}{[B''(T^*)]^{1/2}T^*}\left(\frac{T^*-T}{T^*}\right)^3 + e^{-B_0}\ln\left(\frac{T_c}{T}\right)\right]\right\}, \qquad (6.10)$$

where ξ is a correction factor of order unity. [More precisely, $\xi=1+O(1/T^{*2}B''(T^*))+O(T^{*4}/\rho_0)$.]

Two rather different types of behavior are possible. If $B(T^*)<\ln d$ [i.e., if the nucleation rate $f(T^*)>\chi^4$], then $p(T)$ decreases rapidly toward 0 and supercooling ceases by $T\approx T^*$. (The situation in which the high-temperature phase ceases to be metastable at some temperature $T>0$ may be considered to be a special case of this.) On the other hand, if $B(T^*)>\ln d$, then the bubbles due to thermal fluctuations at high temperatures are not sufficient to complete the phase transition. Since the second term in the exponent of Eq. (6.10), which may be attributed to the effects of quantum-mechanical tunneling, grows only logarithmically, supercooling will continue to exceedingly small temperatures. While the uncertainties in the calculation of high-temperature effects make it difficult to fix precisely the boundary between these two regimes, our numerical results indicate that both correspond to appreciable regions of parameter space; neither requires unnaturally fine tuning of parameters.

VII. CONCLUDING REMARKS

We have seen that in the rapidly expanding early universe, a first-order phase transition can develop in one of two quite distinct manners, depending on whether or not the bubble nucleation rate ever becomes large relative to the fourth power of the expansion rate of the universe; let us now consider some of the implications of each. We exclude from our discussion the possibility that the parameters have been chosen to have exceptional values corresponding to a T^* many orders of magnitude smaller than T_c; in particular, we assume that radiative corrections to the potential are negligible.

In the former case the universe supercools and then, as the coalescence of the bubbles releases the latent heat stored in the bubble walls, reheats. Since the supercooling is rather moderate, we expect the consequent increase in entropy to have no dramatic effect on the cosmic expansion. However, for the particular case we have considered, that of a grand unified gauge theory, the process of bubble coalescence can lead to other difficulties. Since the orientations of the Higgs field in different bubbles are uncorrelated, there is a certain probability that the collision of several bubbles can lead to a Higgs configuration with nontrivial topology and thus to a superheavy magnetic monopole.[23,2,24] This mechanism alone should produce a density of monopoles which is equal to within several orders of magnitude to the density of bubbles. The latter is[2]

$$n_b = R^{-3}(t)\int_{t_0}^{t} dt' f(t') R^3(t') p(t')$$
$$= T^3 \int_{T}^{T_c} dT' \frac{\eta T_c^4 e^{-B(T')}}{\chi g(T') T'^4} p(T') . \qquad (7.1)$$

The integral is cut off by the sharp decrease in $p(T)$ when $B(T) \approx \ln d$, so

$$\frac{n_b}{T^3} \sim \left(\frac{T_c}{M_P}\right)^3 . \qquad (7.2)$$

Taking $T_c \sim 10^{14}$ GeV gives $n_b/T^3 \sim 10^{-15}$, and thus a monopole density far in excess of the limit of 10^{-24} given by Preskill.[3] Consequently, a fast first-order transition of this type is acceptable only if there is a mechanism by which the excess monopoles can be eliminated; an example of such a mechanism has been given by Langacker and Pi.[25]

In the latter case the bubbles formed by thermal fluctuations cannot complete the phase transition and extreme supercooling sets in. In this case there is no difficulty with magnetic monopoles; a calculation similar to that above shows that after reheating the monopole density would be far below the experimental bounds. Instead, there is a potentially more serious difficulty which will be discussed in more detail elsewhere.[26] Because the energy density at low temperatures is dominated by the vacuum contribution ρ_0, the universe grows exponentially with time. With such rapid expansion, bubbles move apart from each other so quickly that collisions between them are very infrequent; since the latent heat is released only through bubble collision and coalescence, there is no general recovery from the supercooling if the expansion is too fast. Specifically, if de^{-B_0} is less than a critical value k, the new phase does not percolate; i.e., the bubbles of the new phase do not coalesce to form an infinite region spreading through space.[27] It can be shown that k is greater than 10^{-9}; since B_0 is typically an order of magnitude larger than $B(T^*)$, which we are assuming to be greater than $\ln d \approx 19 \ln 10$, there is no percolation. In the absence of percolation the presently observed universe must be developed from a finite cluster of bubbles; it seems doubtful that the observed isotropy and homogeneity can be obtained from such a cluster. If they cannot, all choices of parameters leading to such slow first-order transitions must be excluded.

We conclude with the following remarks.

(1) We have calculated the rate of spontaneous nucleation of bubbles; there is also a possibility that bubble nucleation may be induced by the presence of suitable nucleation sites. In particular, magnetic monopoles present in the $SU_4 \times U_1$-symmetric phase might play such a role[28]: At their centers the scalar field corresponds to an unbroken $SU_3 \times SU_2 \times U_1$ symmetry, so for suitable choices of parameters these monopoles might become unstable against expansion of their centers once the temperature had fallen below some $T_N < T_c$. While it is rather difficult to determine precisely the range of parameters for which this is a possibility, it is fairly easy to see how many monopoles would be needed to have an appreciable effect.

Let us assume that there is present a density $D(T)$ of objects, each of which will cause a bubble of true vacuum to form once the temperature has fallen to T_N. Further, let $\bar{p}(t)$ be the fraction of space which would be unconverted at time t if these sites were the only source of bubble nucleation. Equations (6.5)–(6.7) then give

$$\bar{p}(t=\infty) = \exp\left\{-\frac{4\pi}{3}\frac{D(T_N)}{\chi^3}\left[\frac{1}{T_N}\int_0^{T_N}\frac{dT}{g(T)}\right]^3\right\} . \qquad (7.3)$$

If $\bar{p}(\infty)$ is close to unity, induced nucleation will be negligible relative to spontaneous nucleation. This will be the case if $D(T_N)/\chi^3 \lesssim 1$ or, from Eqs. (3.5), (6.4), and (6.7), if

$$\frac{D(T_N)}{T_N^3} \lesssim \left(\frac{T_c}{T_N}\right)^3 \left(\frac{T_c}{M_P}\right)^3 . \qquad (7.4)$$

Specializing now to the case of monopoles as nucleation sites, we note that, since thermal effects decrease rapidly with temperature, T_N is unlikely to be much below T_c; thus for $T_c \sim 10^{14}$ GeV, the above inequality requires $D/T^3 \lesssim 10^{-14}$. By the arguments which led to Eq. (7.2), this is roughly the density of $SU_4 \times U_1$ monopoles which one would expect to be produced by bubble coalescence during the $SU_5 - SU_4 \times U_1$ phase transition (which also occurs in the 10^{14} GeV range). Thus, for those values of the parameters which allow monopoles to act as nucleation sites, their effect may be significant, if not dominant. It should be stressed however that this would not qualitatively affect our previous results: At most, it changes the boundary in parameter space separating the two regimes.

(2) For the sake of definiteness we have considered the SU_5 model. However, we do not expect that the results would have been qualitatively very different had we chosen a different grand unified model (provided, of course, the potential were such as to give a first-order transition). Indeed, it is only in the choice of the energy scale for symmetry breaking, and thus of T_c, that the fact that the model was a grand unified one played a signifi-

cant role. There is perhaps a distinction to be made between phase transitions which are first-order because of the nature of the tree-level potential and those which arise as a result of radiative corrections.[29] Our SU_5 results suggest that in the former case there can be an appreciable range of parameters corresponding to each of the types of behavior we have found. In the latter case, scaling arguments similar to those used in obtaining Eq. (4.5) suggest much slower nucleation rates, with the result that only for a very narrow choice of parameters does the existence of a metastable false vacuum at $T=0$ not lead to extreme supercooling and nonpercolation.

ACKNOWLEDGMENTS

We would like to thank Sidney Coleman and Gordon Lasher for helpful discussions and Benjamin Svetitsky for assistance with the computer calculations. We are grateful for the hospitality of the Theory Group at SLAC where this work was begun. One of us (E.J.W.) was supported in part by an Alfred P. Sloan Fellowship. This work was supported in part by the U.S. Department of Energy.

APPENDIX

In this appendix we derive Eq. (6.5) for $p(t)$, the fraction of space remaining in the old phase at time t. We begin by considering a space containing randomly place spheres (including overlapping and nested spheres) and asking for the probability that a given point is not contained in any sphere. Let $n(V)dV$ be the density of spheres with volume between V and $V+dV$, and let $g(V_1, V_2)$ be the probability that a given point is not contained in any sphere of volume between V_1 and V_2. Then

$$g(V_1, V_2+dV_2)=g(V_1, V_2)g(V_2, V_2+dV_2)$$
$$=g(V_1, V_2)[1-n(V_2)V_2dV_2], \quad \text{(A1)}$$

so

$$\frac{dg(V_1, V_2)}{dV_2}=-n(V_2)V_2g(V_1, V_2) \quad \text{(A2)}$$

and

$$g(V_1, V_2)=\exp\left[-\int_{V_1}^{V_2} dV\,n(V)V\right]. \quad \text{(A3)}$$

In particular, the probabilty of not being in any sphere is

$$g(0, \infty)=e^{-\mathcal{V}}, \quad \text{(A4)}$$

where

$$\mathcal{V}=\int_0^\infty dV\,n(V)V \quad \text{(A5)}$$

is the total volume in spheres (with appropriate multiple counting of overlaps) per unit volume of space.

In the phase transition the distribution of bubbles is not completely random because of the physical restriction that bubbles do not nucleate within other bubbles. Let us therefore relax this condition and include an appropriate number of fictitious bubbles formed within bubbles; since these fictitious bubbles are entirely contained within real ones, this will cause no error in our determination of $p(t)$. The total number of bubbles (real and fictitious) formed per unit time per unit coordinate volume is $f(t)R^3(t)$, while the coordinate volume at time t of a bubble formed at t' is

$$V(t', t)=\frac{4\pi}{3}\left[\int_{t'}^{t} dt''\frac{1}{R(t'')}\right]^3 \quad \text{(A6)}$$

(the bubbles are formed with a negligible initial radius and expand with a speed rapidly approaching that of light), so

$$\mathcal{V}=\int_{t_0}^{t} dt' f(t')R^3(t')V(t', t)\,. \quad \text{(A7)}$$

Substitution into Eq. (A4) then gives Eq. (6.5).

---

*Present address.

[1]M. Yoshimura, Phys. Rev. Lett. 41, 281 (1978); 42, 746(E) (1979); S. Dimopoulos and L. Susskind, Phys. Rev. D 18, 4500 (1978); D. Toussaint, S. Treiman, F. Wilczek, and A. Zee, *ibid.* 19, 1036 (1979); S. Weinberg, Phys. Rev. Lett. 42, 850 (1979); J. Ellis, M. Gaillard, and D. Nanopoulos, Phys. Lett. 80B, 360 (1979); 82B, 464 (1979). A more complete list of references is contained in Ref. 4.

[2]A. Guth and S.-H. Tye, Phys. Rev. Lett. 44, 631 (1980); 44, 963 (1980).

[3]J. Preskill, Phys. Rev. Lett. 43, 1365 (1979).

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

[5]H. Georgi and S. Glashow, Phys. Rev. Lett. 32, 438 (1974).

[6]A. Buras, J. Ellis, M. Gaillard, and D. Nanopoulos, Nucl. Phys. B135, 66 (1978).

[7]S. Coleman, Phys. Rev. D 15, 2929 (1977); C. Callan and S. Coleman, *ibid.* 16, 1762 (1977).

[8]L.-F. Li, Phys. Rev. D 9, 1723 (1974). For an analysis similar to that given below see A. Kennedy, G. Lazarides, and Q. Shafi, CERN Report No. TH 2944 (unpublished).

[9]We denote this by a Hermitian traceless 5×5 matrix $\phi=\Sigma\phi_a\lambda^a/\sqrt{2}$, where the λ^a are normalized so that $\mathrm{Tr}(\lambda^a\lambda^b)=2\delta^{ab}$.

[10]We follow the notation of Ref. 6. The full potential would also contain terms involving the fundamental-

representation Higgs field, which is expected to acquire a nonzero expectation value only at temperatures of order 10^2 GeV.

[11]Although there are stationary points with ϕ having three distinct eigenvalues, these are never local extrema. This result has been extended to all SU(N) by H. Ruegg, Phys. Rev. D 22, 2040 (1980).

[12]For a discussion of the mixing between these particles and the fundamental-representation scalars, see Ref. 6.

[13]S. Weinberg, Phys. Rev. D 9, 3357 (1974).

[14]L. Dolan and R. Jackiw, Phys. Rev. D 9, 3320 (1974).

[15]There will also be contributions to σ from the couplings between the adjoint and the fundamental Higgs multiplets.

[16]In the region of the β-γ plane which we are considering, $k>2$ only for $(\beta-\beta_c)/\beta_c<0.016$; at the phase boundary k diverges.

[17]Some insight into this result may be gained from noting that as one approaches the line in the β-γ plane at which the $SU_4 \times U_1$-symmetric solution ceases to be metastable, the potential barrier disappears for the $SU_3 \times U_1 \times U_1$-symmetric configurations, but not for the $SU_2 \times SU_2 \times U_1$-symmetric ones.

[18]I. Affleck, Harvard University Report No. HUTP-80/A062 (unpublished).

[19]It is important that the solution be unstable, i.e., it should correspond to a bubble of critical size. The imaginary part of the free energy then results from the unstable mode in the small-oscillation problem.

[20]A. Linde, Phys. Lett. 70B, 306 (1977); 92B, 119 (1980).

[21]The methods of Ref. 14 cannot be used to improve the approximation because they give a contribution to the effective potential from scalar particle loops which becomes complex at certain temperatures in this range.

[22]There is a very narrow strip along the line separating regions (a) and (d) of Fig. 2 for which these statements will not apply; on the boundary between the regions, $T^*=0$.

[23]T. W. B. Kibble, J. Phys. A 9, 1387 (1976).

[24]M. Einhorn, D. Stein, and D. Toussaint, Phys. Rev. D 21, 3295 (1980).

[25]P. Langacker and S.-Y. Pi, Phys. Rev. Lett. 45, 1 (1980).

[26]A. Guth and E. Weinberg (unpublished).

[27]For a review of percolation, see J. W. Essam, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. Green (Academic, London, 1972) Vol. 2; V. Shante and S. Kirkpatrick, Adv. Phys. 20, 325 (1971).

[28]For a discussion of this effect in a simpler model, see P. Steinhardt, Harvard University Report No. HUTP-80/A088 (unpublished).

[29]See, for example, A. Linde, Ref. 20; E. Witten, Harvard University Report No. HUTP 80/A040 (unpublished); A. Guth and E. Weinberg, Phys. Rev. Lett. 45, 1131 (1980).