

# Electroweak baryogenesis

Mark Trodden\*

*Particle Astrophysics Theory Group, Department of Physics, Case Western Reserve University, Cleveland, Ohio 44106-7079*

Contrary to naive cosmological expectations, all evidence suggests that the universe contains an abundance of matter over antimatter. This article reviews the currently popular scenario in which testable physics, present in the standard model of electroweak interactions and its modest extensions, is responsible for this fundamental cosmological datum. A pedagogical explanation of the motivations and physics behind electroweak baryogenesis is provided, and analytical approaches, numerical studies, up to date developments, and open questions in the field are also discussed. [S0034-6861(99)00105-1]

## CONTENTS

I. Introduction	1463	IX. Concluding Remarks and a Look to the Future	1495
II. Baryon Number Violation in the Electroweak Theory	1465	Acknowledgments	1496
A. Zero-temperature results	1468	References	1497
B. Nonzero-temperature results	1469		
1. Mechanical analogy	1469		
2. Electroweak theory	1469		
III. $C$ and $CP$ Violation	1471		
A. Standard-model $CP$ violation: The Kobayashi-Maskawa matrix	1471		
B. Two-Higgs-doublet model	1472		
C. $CP$ violation from higher-dimension operators	1472		
D. General treatment	1473		
IV. Electroweak Phase Transition	1473		
A. Finite-temperature effective potential	1475		
B. Dimensional reduction	1477		
C. Numerical simulations—lattice approaches	1478		
D. Other approaches	1479		
E. Subcritical fluctuations	1479		
F. Erasure of the baryon asymmetry: Washout	1480		
G. Summary	1481		
V. Local Electroweak Baryogenesis	1481		
A. Local baryogenesis through unwinding	1482		
B. Kicking configurations across the barrier	1484		
C. Making progress	1486		
VI. Nonlocal Baryogenesis	1486		
A. Thin bubble walls	1487		
B. Thick bubble walls	1489		
1. Spontaneous baryogenesis	1489		
2. Classical force baryogenesis	1489		
C. Summary and progress	1490		
VII. A Realistic Model of Electroweak Baryogenesis: The MSSM	1490		
A. Electroweak phase transition in the MSSM	1490		
B. Extra $CP$ violation in the MSSM	1491		
VIII. Topological Defects and the Departure from Thermal Equilibrium	1492		
A. Electroweak symmetry restoration and the baryogenesis volume	1492		
B. Local baryogenesis and diffusion in defects	1493		
C. Specific geometry and examples	1494		
D. Particle physics model	1495		
E. Summary	1495		

## I. INTRODUCTION

Since antiparticles were first predicted (Dirac, 1930, 1931) and observed (Anderson, 1932, 1933) it has been clear that there is a high degree of symmetry between particles and antiparticles. This means, among other things, that a world composed of antimatter would behave in a similar manner to our world. This basic tenet of particle physics, the symmetry between matter and antimatter, is in stark contradiction to the wealth of everyday and cosmological evidence that the universe is composed almost entirely of matter with little or no primordial antimatter.

The evidence that the universe is devoid of antimatter comes from a variety of different observations. On the very small scale, the absence of proton-antiproton annihilations in our everyday actions constitutes strong evidence that our world is composed only of matter and no antimatter. Moving up in scale, the success of satellite launches, lunar landings, and planetary probes indicates that our solar system is made up of the same type of matter that we are and that there is negligible antimatter on that scale. To determine whether antimatter exists in our galaxy, we turn to cosmic rays. Here we see the first detection of antimatter outside particle accelerators. Mixed in with the many protons present in cosmic rays are a few antiprotons, present at a level of around  $10^{-4}$  in comparison with the number of protons ( $\sim 200$  per  $m^2$  per s) (see for example Ahlen *et al.*, 1988). However, this number of antiprotons is consistent with their secondary production through acceleratorlike processes,  $p + p \rightarrow 3p + \bar{p}$ , as the cosmic rays stream towards us. Thus there is no evidence for primordial antimatter in our galaxy. Finally, if matter and antimatter galaxies were to coexist in clusters of galaxies, then we would expect there to be a detectable background of  $\gamma$  radiation from nucleon-antinucleon annihilations within the clusters. This background is not observed, and so we conclude that there is negligible antimatter on the scale of clusters [for a review of the evidence for a baryon asymmetry see Steigman (1976)].

\*Electronic address: trodden@huxley.phys.cwru.edu

A natural question to ask is, what is the largest scale on which we can say that there is no antimatter? This question was addressed by Steigman (1976) and by Stecker (1985), and in particular was the subject of a careful recent analysis by Cohen *et al.* (1998). If large domains of matter and antimatter exist, then annihilations would take place at the interfaces between them. If the typical size of such a domain were small enough, then the energy released by these annihilations would result in a diffuse  $\gamma$ -ray background and a distortion of the cosmic microwave radiation, neither of which is observed. Quantitatively, the result obtained by Steigman and Stecker is that we may safely conclude that the universe consists entirely of matter on all scales up to the Hubble size. It therefore seems that the universe is fundamentally matter-antimatter asymmetric.

The above observations put an experimental upper bound on the amount of antimatter in the universe. However, strict quantitative estimates of the relative abundances of baryonic matter and antimatter may also be obtained from the standard cosmology. Primordial nucleosynthesis [for a review see Copi *et al.* (1995)] is one of the most powerful tools of the standard cosmological model. The theory allows accurate predictions of the cosmological abundances of all the light elements, H,  $^3\text{He}$ ,  $^4\text{He}$ , D, B, and  $^7\text{Li}$ , while requiring only a single input parameter. Define  $n_b$  to be the number density of baryons in the universe. Similarly define  $n_{\bar{b}}$  to be the number density of antibaryons, and the difference between the two to be  $n_B$ . Then, if the entropy density in the universe is given by  $s$ , the single parameter required by nucleosynthesis is the baryon-to-entropy ratio

$$\eta \equiv \frac{n_B}{s} = \frac{n_b - n_{\bar{b}}}{s}, \quad (1)$$

and one may conservatively say that calculations of the primordial light element abundances are correct if

$$1.5 \times 10^{-10} < \eta < 7 \times 10^{-10}. \quad (2)$$

Although the range of  $\eta$  within which all light element abundances agree with observations is quite narrow (see Fig. 1), its existence at all is remarkable and constitutes a strong confirmation of the standard cosmology. For recent progress in nucleosynthesis see Tytler *et al.* (1996) and Hogan (1997).

The standard cosmological model provides a complete and accurate description of the evolution of the universe from extremely early times (a few minutes) to the present day (10–20 billion years) given a host of initial conditions, one of which is the value of  $\eta$ . This standard picture is based on classical, fluid sources for the Einstein equations of General Relativity. While the success of the standard cosmology is encouraging, there remains the question of the initial conditions. One approach is just to consider the initial values of cosmological parameters as given. However, the values required for many parameters are extremely unnatural in the sense that the tiniest deviation from them leads to a universe entirely different from the one that we observe. One well-known example of this is the initial value of the mass density of

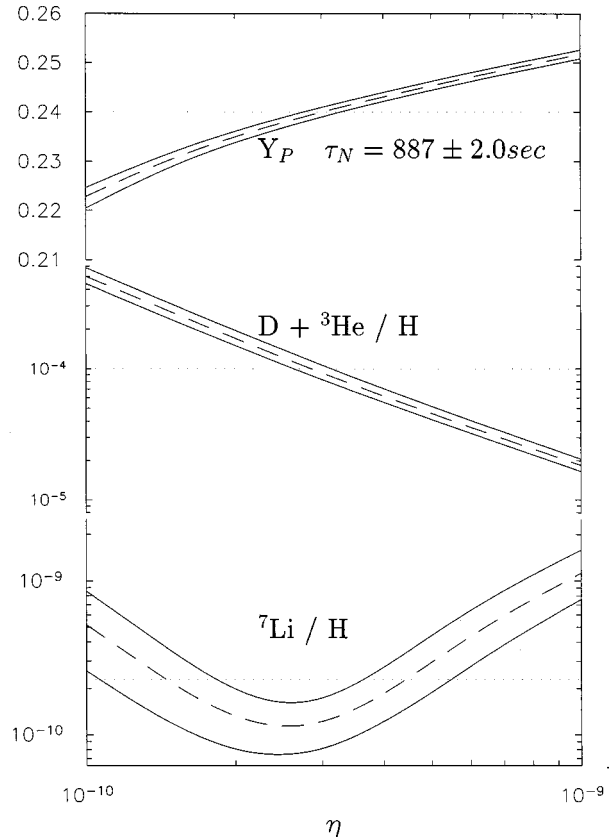


FIG. 1. Zone of concordance for primordial nucleosynthesis.  $\tau_N$  denotes the value of the neutron lifetime used in these calculations. Thanks to Peter Kernan and Lawrence Krauss for providing this figure (Krauss and Kernan, 1995).

the universe, the naturalness of which is at the root of the *flatness problem* of the standard cosmology.

The philosophy of modern cosmology, developed over the last thirty years, is to attempt to explain the required initial conditions on the basis of quantum field theories of elementary particles in the early universe. This approach has allowed us to push our understanding of early-universe cosmology back to much earlier times, conservatively as early as  $10^{-10}$  seconds and perhaps much earlier.

The generation of the observed value of  $\eta$  in this context is referred to as *baryogenesis*. A first step is to outline the necessary properties a particle physics theory must possess. These conditions were first identified by Sakharov (1967) and are now referred to as the three *Sakharov criteria*. They are

- Violation of the baryon number ( $B$ ) symmetry;
- Violation of the discrete symmetries  $C$  (charge conjugation) and  $CP$  (the composition of parity and  $C$ );
- A departure from thermal equilibrium.

The first of these is rather obvious. If no processes ever occur in which  $B$  is violated, then the total number of baryons in the universe must remain constant, and therefore no asymmetry can be generated from symmetric initial conditions. The second Sakharov criterion is required because, if  $C$  and  $CP$  are exact symmetries, then one can prove that the total rate for any process

that produces an excess of baryons is equal to the rate of the complementary process that produces an excess of antibaryons and so no net baryon number can be created. That is to say that the thermal average of  $B$ , which is odd under both  $C$  and  $CP$ , is zero unless those discrete symmetries are violated. Finally, there are many ways to explain the third criterion. One way is to calculate the equilibrium average of  $B$ :

$$\begin{aligned} \langle B \rangle_T &= \text{Tr}(e^{-\beta H} B) = \text{Tr}[(CPT)(CPT)^{-1} e^{-\beta H} B] \\ &= \text{Tr}[e^{-\beta H} (CPT)^{-1} B (CPT)] = -\text{Tr}(e^{-\beta H} B). \end{aligned} \quad (3)$$

In the third step I have used the requirement that the Hamiltonian  $H$  commutes with  $CPT$ , and in the last step used the property of  $B$  that it is odd under  $C$  and even under  $P$  and  $T$  symmetries. Thus  $\langle B \rangle_T = 0$  in equilibrium and there is no generation of net baryon number. This may be loosely described in the following way. In quantum field theories in thermal equilibrium, the number density of any particle species,  $X$  say, depends only on the energy of that species, through

$$n_{eq}(X) = \frac{1}{e^{(E-\mu)/T} \pm 1}, \quad (4)$$

where  $\mu$  is the chemical potential corresponding to baryon number. Since the masses of particle and antiparticle are equal by virtue of the  $CPT$  theorem, and  $\mu = 0$  if baryon number is violated, we have

$$N_{eq}(X) = \int \frac{d^3 p}{(2\pi)^3} n_{eq} = N_{eq}(\bar{X}), \quad (5)$$

and again there is no net asymmetry produced.

The focus of this article is to review one popular scenario for generating the baryon asymmetry of the universe (BAU), as quantified in Eq. (2), within the context of modern cosmology. In general, such scenarios involve calculating  $n_B$  and then dividing by the entropy density,

$$s = \frac{2\pi^2}{45} g_* T^3, \quad (6)$$

where  $g_*$  is the effective number of massless degrees of freedom at temperature  $T$ . While there have been many attempts in the literature to explain the baryon asymmetry of the universe (for a review see Dolgov, 1992), I shall concentrate on those scenarios which involve anomalous electroweak physics, when the universe was at a temperature of  $10^2$  GeV [for earlier reviews see Turok (1993) Cohen *et al.* (1993), and Rubakov and Shaposhnikov (1996)]. The production of baryon asymmetry through these models is referred to as *electroweak baryogenesis*.

In the next section I shall describe baryon number violation in the electroweak theory both at zero and at nonzero temperature. In Sec. III, I shall move on to the subject of  $CP$  violation, explaining how this arises in the standard model and how it is achieved in some popular extensions. Section IV contains an account of the electroweak phase transition, including discussions of both analytic and numerical approaches. Having set up the framework for electroweak baryogenesis, I turn in Sec.

V to the dynamics of the case in which baryon production occurs close to a phase boundary during a phase transition. In Sec. VI, I then extend these ideas to include the effects of particle transport, or diffusion. Section VII contains a description of how baryogenesis is implemented in a popular extension of the standard model, the minimal supersymmetric standard model (MSSM). In Sec. VIII, I explain how, in some extensions of the electroweak theory, baryogenesis may be mediated by topological defects, alleviating the constraints on the order of the phase transition. Finally, in Sec. IX, I summarize the results and comment on open questions and future directions in the field.

It is my hope that this article will serve as both a review of the background and basic material for beginners in the field and a summary of and commentary on the most recent results and directions in the subject. However, the focus of this article, as in any such endeavor, is quite idiosyncratic, and I apologize to any of my colleagues whose work has been omitted or incorrectly detailed. A different focus can be found in other accounts of the subject and, in particular, for a comprehensive modern review of numerical approaches I recommend that of Rubakov and Shaposhnikov (1996).

A note about conventions. Throughout I use a metric with signature  $+2$  and, unless explicitly stated otherwise, I employ units such that  $\hbar = c = k = 1$  so that Newton's constant is related to the Planck mass through  $G = M_{pl}^{-2}$ .

## II. BARYON NUMBER VIOLATION IN THE ELECTROWEAK THEORY

The standard model of unified electromagnetic and weak nuclear interactions (Glashow, 1961; Weinberg, 1967; Salam, 1968) is based on the gauge groups  $SU(2) \times U(1)$ . The model is described by the Lagrangian density

$$\mathcal{L} = (D_\mu \phi)^\dagger D^\mu \phi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^a W^{a\mu\nu} + V(\phi) + \mathcal{L}_f. \quad (7)$$

Here, the field strengths are

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\ W_{\mu\nu}^a &= \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g \epsilon^{abc} A_\mu^b A_\nu^c, \end{aligned}$$

where  $B_\mu$  is the hypercharge gauge field and  $A_\mu^a$  are the weak isospin gauge fields. The covariant derivative is

$$D_\mu = \partial_\mu - \frac{i}{2} g \tau \cdot \mathbf{A}_\mu - \frac{i}{2} g' B_\mu,$$

and the Higgs potential is

$$V(\phi) = \frac{\lambda}{4} (\phi^\dagger \phi - v^2)^2. \quad (8)$$

In the above,  $g$  is the  $SU(2)$  coupling constant,  $g'$  is the  $U(1)$  coupling constant,  $\lambda$  is the Higgs self-coupling, and  $v = 246$  GeV is the vacuum expectation value (VEV) of the Higgs. The potential  $V(\phi)$  is chosen so that the

gauge symmetry is spontaneously broken down to the  $U(1)$  of electromagnetism that is realized by the vacuum today, and  $\mathcal{L}_f$  denotes the fermionic sector of the theory that I will describe in a moment. Note that it is conventional to define

$$\alpha_W \equiv \frac{g^2}{4\pi} \approx \frac{1}{29} \tag{9}$$

and to write the ratio of  $g'$  to  $g$  as

$$\tan \theta_W \equiv \frac{g'}{g}, \tag{10}$$

where the experimentally measured value of the *weak mixing angle*  $\theta_W$  is given by

$$\sin^2 \theta_W \approx 0.23. \tag{11}$$

It will be useful to describe briefly a slightly different formulation of the  $SU(2)$ +Higgs theory, equivalent to considering the purely bosonic part of the standard model and ignoring the  $U(1)$  hypercharge gauge field. The Lagrangian may be written in the form

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2} \text{Tr}(W_{\mu\nu} W^{\mu\nu}) - \frac{1}{2} \text{Tr}(D^\mu \Phi)^\dagger D_\mu \Phi \\ & - \frac{\lambda}{4} [\text{Tr}(\Phi^\dagger \Phi) - v^2]^2. \end{aligned} \tag{12}$$

In this form, the standard Higgs doublet  $\phi = (\phi_1, \phi_2)$  is related to the matrix  $\Phi$  by

$$\Phi(\mathbf{x}, t) = \begin{pmatrix} \phi_2^* & \phi_1 \\ -\phi_1^* & \phi_2 \end{pmatrix}. \tag{13}$$

For reference,  $g=0.65$ , the gauge boson mass is  $m_W = \frac{1}{2} g v$ , and the Higgs boson mass is  $m_H = \sqrt{2\lambda} v$ . Note that

$$\Phi^\dagger \Phi = (\varphi_1^* \varphi_1 + \varphi_2^* \varphi_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{14}$$

so that one can write

$$\Phi = \frac{\sigma}{\sqrt{2}} U, \tag{15}$$

where  $\sigma^2 = 2(\varphi_1^* \varphi_1 + \varphi_2^* \varphi_2) = \text{Tr} \Phi^\dagger \Phi$ , and  $U$  is an  $SU(2)$  valued field that is uniquely defined at any space-time point where  $\sigma$  does not vanish. Without loss of generality, impose the condition that at all times

$$\lim_{|\mathbf{x}| \rightarrow \infty} \sigma(\mathbf{x}, t) = v, \tag{16}$$

$$\lim_{|\mathbf{x}| \rightarrow \infty} U(\mathbf{x}, t) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{17}$$

In  $A_0=0$  gauge, a vacuum configuration is of the form

$$\begin{aligned} \Phi &= \frac{v}{\sqrt{2}} U \\ A_j &= \frac{1}{ig} \partial_j U U^\dagger. \end{aligned} \tag{18}$$

At any time  $t$  when  $\sigma(\mathbf{x}, t) \neq 0$  for all  $\mathbf{x}$  we have that  $U(\mathbf{x}, t)$  is a map from  $\mathbf{R}^3$  into  $SU(2)$  with the points at infinity identified [that is from  $S^3$  into  $SU(2)$ ], and therefore  $U(\mathbf{x}, t)$  can be associated with an integer-valued winding,

$$\begin{aligned} N_H(t) &= w[U] \\ &= \frac{1}{24\pi^2} \int d^3x \epsilon^{ijk} \text{Tr}[U^\dagger \partial_i U U^\dagger \partial_j U U^\dagger \partial_k U], \end{aligned} \tag{19}$$

the Higgs winding number. If  $\Phi(\mathbf{x}, t)$  evolves continuously in  $t$  then  $N_H(t)$  can change only at times when there is a zero of  $\sigma$  at some point in space. At such times,  $N_H$  is not defined; at all other times, it is integer valued. Note that the Higgs winding number of a vacuum configuration (18) is equal to its Chern-Simons number,

$$N_{CS}(t) \equiv \frac{g^2}{32\pi^2} \int d^3x \epsilon^{ijk} \text{Tr} \left( A_i \partial_j A_k + \frac{2}{3} ig A_i A_j A_k \right). \tag{20}$$

However, for a general nonvacuum configuration the Chern-Simons number is not integer valued.

Finally, the fermionic sector of the full  $SU(2) \times U(1)$  theory is described by

$$\mathcal{L}_f = \mathcal{L}_l + \mathcal{L}_q, \tag{21}$$

where the lepton- and quark-sector Lagrangians for a single family are

$$\mathcal{L}_l = -i\bar{\Psi} \gamma^\mu D_\mu \Psi - i\bar{e}_R \gamma^\mu D_\mu e_R + h(\bar{e}_R \phi^\dagger \Psi + \bar{\Psi} \phi e_R), \tag{22}$$

$$\begin{aligned} \mathcal{L}_q = & -i(\bar{u}, \bar{d})_L \gamma^\mu D_\mu \begin{pmatrix} u \\ d \end{pmatrix}_L - i\bar{u}_R \gamma^\mu D_\mu u_R - i\bar{d}_R \gamma^\mu D_\mu d_R \\ & - G_d \left[ (\bar{u}, \bar{d})_L \begin{pmatrix} \phi^+ \\ \phi \end{pmatrix} d_R + \bar{d}_R (\phi^-, \phi^*) \begin{pmatrix} u \\ d \end{pmatrix}_L \right] \\ & - G_u \left[ (\bar{u}, \bar{d})_L \begin{pmatrix} -\phi^* \\ \phi^- \end{pmatrix} u_R + \bar{u}_R (-\phi, \phi^+) \begin{pmatrix} u \\ d \end{pmatrix}_L \right], \end{aligned} \tag{23}$$

where I have written  $\phi = (\phi^+, \phi^0)$ , with  $\phi^- = (\phi^+)^*$ . Here,  $e$  is the electron,  $u$  and  $d$  are the up and down quarks, respectively, and  $\Psi$  represents the left-handed lepton doublet. The indices  $L$  and  $R$  refer to left- and right-handed components, and  $G_d$  and  $G_u$  are Yukawa coupling constants.

When extended to three families, this contains the six types of quarks  $U^i = (u, c, t)$ ,  $D^i = (d, s, b)$ , the electron, muon, and tau lepton, and their associated neutrinos. Note that gauge-invariant fermion mass terms are generated through couplings to the Higgs doublet  $\Phi$ .

At the classical level, the number of each fermionic species is separately conserved in all processes. This is reflected in the fact that for each species there exists a global  $U(1)$  current, which is exactly classically conserved. In the standard model there are a total of 12 different species and so there are 12 separate conserved global currents. In particular, for baryons, one may write a vectorlike current

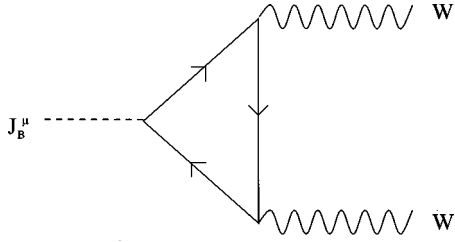


FIG. 2. The triangle diagram contributing to the anomaly in the baryon and lepton number currents.

$$j_B^\mu = \frac{1}{2} \bar{Q} \gamma^\mu Q, \quad (24)$$

where  $Q$  represents quarks, and there is an implied sum over the color and flavor indices. Now, due to quantum effects, any axial current  $\bar{\psi} \gamma^\mu \gamma^5 \psi$  of a gauge-coupled Dirac fermion  $\psi$  is anomalous (Adler, 1969; Bell and Jackiw, 1969). This is relevant to baryon number since the electroweak fermions couple chirally to the gauge fields. If one writes the baryon current as

$$j_B^\mu = \frac{1}{4} [\bar{Q} \gamma^\mu (1 - \gamma^5) Q + \bar{Q} \gamma^\mu (1 + \gamma^5) Q], \quad (25)$$

only the axial part of this vector current is important when one calculates the divergence. This effect can be seen by calculating triangle graphs (see Fig. 2) and leads to the following expressions for the divergences of the baryon number and lepton number currents:

$$\partial_\mu j_B^\mu = \partial_\mu j_L^\mu = n_f \left( \frac{g^2}{32\pi^2} W_{\mu\nu}^a \tilde{W}^{a\mu\nu} - \frac{g'^2}{32\pi^2} F_{\mu\nu} \tilde{F}^{\mu\nu} \right), \quad (26)$$

where  $n_f$  is the number of families,

$$\tilde{W}^{\mu\nu} = \frac{1}{2} \epsilon^{\mu\nu\alpha\beta} W_{\alpha\beta} \quad (27)$$

is the dual of the  $SU(2)$  field strength tensor, and an analogous expression holds for  $\tilde{F}$ .

The anomaly is important because of the multiple vacuum structure of the theory (see Fig. 3), as I shall describe in the following subsections. Equation (26) may be written as

$$\partial_\mu j_B^\mu = \partial_\mu j_L^\mu = n_f \left( \frac{g^2}{32\pi^2} \partial_\mu K^\mu - \frac{g'^2}{32\pi^2} \partial_\mu k^\mu \right), \quad (28)$$

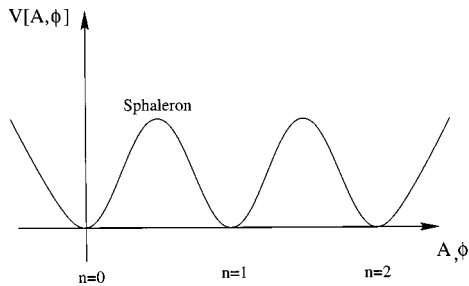


FIG. 3. The multiple vacua of the electroweak theory depicted along some direction in configuration space. The sphaleron is also shown.

where the *gauge-noninvariant* currents are defined by

$$K^\mu = \epsilon^{\mu\nu\alpha\beta} \left( W_{\nu\alpha}^a A_\beta^a - \frac{1}{3} g \epsilon_{abc} A_\nu^a A_\alpha^b A_\beta^c \right), \quad (29)$$

$$k^\mu = \epsilon^{\mu\nu\alpha\beta} F_{\nu\alpha} B_\beta.$$

For simplicity, we shall consider space to be a 3-sphere and consider the change in baryon number from time  $t=0$  to some arbitrary final time  $t=t_f$ . For transitions between vacua, the average values of the field strengths are zero at the beginning and the end of the evolution. Since the final term in Eq. (28) is strictly proportional to the field strength of the  $U(1)$  gauge fields, we may ignore this term. Then, using Eq. (20), we may write the change in baryon number as

$$\Delta B = \Delta N_{CS} \equiv n_f [N_{CS}(t_f) - N_{CS}(0)]. \quad (30)$$

Although the Chern-Simons number is not gauge invariant, the change  $\Delta N_{CS}$  is. Thus, since  $N_{CS}$  is integral (as we have mentioned), changes in Chern-Simons number result in changes in baryon number that are integral multiples of the number of families  $n_f$ .

To understand this structure more completely, since the  $U(1)$  gauge fields are unimportant here, I shall return to the  $SU(2)$  theory. Consider fermion production in the background of the evolving Higgs and gauge fields and ignore the back reaction of the fermions on the bosonic fields. Consider the dynamics of nonzero energy configurations with nonzero Higgs winding. A simple example (Ambjørn *et al.*, 1989) is

$$\Phi(\mathbf{x}) = \frac{v}{\sqrt{2}} U_{[1]}(\mathbf{x}), \quad (31)$$

$$A_\mu(\mathbf{x}) = 0,$$

where  $U_{[1]}(\mathbf{x})$  is a winding-number-one map, say,

$$U_{[1]}(\mathbf{x}) = \exp(i \eta(r) \boldsymbol{\tau} \cdot \hat{\mathbf{x}}), \quad (32)$$

with  $\eta(0) = -\pi$  and  $\eta(\infty) = 0$ . The configuration (31) has no potential energy but does carry gradient energy because the covariant derivatives  $D_t \Phi$  do not vanish. This configuration has  $N_H = 1$ , where  $N_H$  is defined in Eq. (19). If the configuration (31) were released from rest it would radiate away its energy and relax towards a vacuum configuration. There are two very different ways for this to occur (Turok and Zdrozny, 1990.) If the characteristic size of  $U_{[1]}$  is large compared to  $m_W^{-1}$ , then the gauge field will evolve until it lines up with the Higgs field, making the covariant derivatives zero, and at late times  $N_H$  will still be one. If the characteristic size is small the configuration will shrink, the Higgs field  $\sigma$  will go through a zero, and at late times  $N_H$  will be zero. These dynamics are the subject of Sec. V.

Note that  $N_H$  is not invariant under large gauge transformations. However, the change  $\Delta N_H$  in Higgs winding is gauge invariant, and the two distinct relaxation processes are distinguished by whether  $\Delta N_H$  is zero or nonzero. To be definite, I shall always choose the gauge such that the prototypical initial configuration is of the form (31), which has  $N_H = 1$ .

Now, if the fields relax to the vacuum by changing the Higgs winding then there is no anomalous fermion number production. However, if there is no net change in Higgs winding during the evolution (for example,  $\sigma$  never vanishes) then there is anomalous fermion number production.

To understand these claims consider two sequences of configurations beginning with the wound-up configuration [Eq. (31)] and ending at the classical vacuum [Eq. (18)]. The first sequence ends at the vacuum (18) with  $U=1$ , while the second ends up at  $U=U_{[1]}$ . Note that these sequences cannot be solutions to the classical equations of motion since the initial configurations carry energy whereas the final ones do not. Throughout both sequences the boundary conditions (16) and (17) are maintained. For the first sequence,  $\sigma$  must vanish at some intermediate configuration since the Higgs winding changes. For the second sequence, the change in Higgs winding is zero and  $\sigma$  need not vanish.

Now introduce an  $SU(2)_L$  weak fermionic doublet  $\psi$ . The fermion is given mass through the usual gauge-invariant coupling to the Higgs field  $\Phi$ , and for simplicity we assume that both the up and down components of the doublet have the same mass  $m$ . The fermion field is quantized in the background of the bosonic fields given by the above interpolation.

The anomaly equation (26) reduces here to

$$\partial_\mu J^\mu = \frac{g^2}{32\pi^2} \text{Tr}(W\tilde{W}), \quad (33)$$

which, when integrated, implies that the change in the fermion number from the beginning to the end of a sequence is given by

$$\int d^3x J^0 \Big|_{\text{final}} - \int d^3x J^0 \Big|_{\text{initial}} = -w[U], \quad (34)$$

where  $U$  is that of the final configuration (18). For the first sequence  $w$  is one, whereas for the second it is zero. Thus fermion number is violated in processes for which the configuration (31) unwinds via gauge unwinding, but is not violated when such a configuration unwinds via a Higgs unwinding.

### A. Zero-temperature results

We have seen that the vacuum of the electroweak theory is degenerate, labeled by  $N_H$  and  $N_{CS}$ . The field theories constructed around these vacua are entirely equivalent and may be obtained from one another through large gauge transformations. However, transitions between these vacua result in the anomalous production of fermions via the anomaly equation (26). Such transitions violate baryon number and so are classically forbidden, since baryon number is an exact global symmetry of the theory. In fact, as I shall describe, at zero temperature, baryon-number-violating processes occur through quantum tunneling between the classical vacua of the theory.

In the infinite-dimensional gauge and Higgs-field configuration space, adjacent vacua of the electroweak

theory are separated by a ridge of configurations with energies larger than that of the vacuum (see Fig. 3). The lowest-energy point on this ridge is a saddle-point solution (Dashen *et al.*, 1974; Christ, 1980) to the equations of motion with a single negative eigenvalue, and is referred to as the *sphaleron* (Manton, 1983; Klinkhamer and Manton, 1984). The sphaleron plays a central role in the calculation of the rate of baryon-number-violating processes.

The calculation of the tunneling rate between degenerate vacua in quantum field theories is well established (for a review see Coleman, 1979). One first constructs the Euclideanized action  $S_E$ , obtained from the Minkowski-space action by performing a Wick rotation,

$$t \rightarrow -it \equiv \tau. \quad (35)$$

In the case of pure gauge theories, one then finds the solution to the Euclidean equations of motion that interpolates between the two vacuum states and minimizes the Euclidean action. Such a solution to the full four-dimensional Euclidean system is known as an *instanton* and may be seen as a time series of three-dimensional configurations. The transition rate between the degenerate vacua is then proportional to

$$\exp(-S_E)|_{\text{instanton}}. \quad (36)$$

When the theory in question has a Higgs field, as in the electroweak theory, the calculation of the tunneling amplitude is a little more complicated. In this case, there are no nonzero size configurations that minimize the Euclidean action, and a slightly different approach is used. The procedure is to fix the size of the Euclidean configurations in the functional integral by explicitly introducing a constraint (Affleck, 1981a). The functional integral is then evaluated using these *constrained instantons* and, finally, an integral over the instanton size is performed.

As an estimate of the zero-temperature  $B$ -violating rate in the electroweak theory, I shall follow the approach of 't Hooft (1976). Consider the pure gauge  $SU(2)$  theory (note no Higgs field). The relevant configurations are finite-action solutions to the Euclidean equations of motion. Further, as I described earlier, the configurations of interest must possess nonzero gauge winding in order for baryon number violation to take place. If we consider the case in which the instanton interpolates between adjacent vacua of the theory, then the topological charge of such a solution is  $\Delta N_{CS}=1$ . Now, the quantity

$$\int d^4x (W_{\mu\nu}^a - \tilde{W}_{\mu\nu}^a)^2 \quad (37)$$

is positive semidefinite for any gauge-field configuration. This allows us to construct a bound on the Euclidean action of configurations in the following way. Expanding Eq. (37) gives

$$\int d^4x [\text{Tr}(W_{\mu\nu} W^{\mu\nu}) + \text{Tr}(\tilde{W}_{\mu\nu} \tilde{W}^{\mu\nu}) - 2 \text{Tr}(W_{\mu\nu} \tilde{W}^{\mu\nu})] \geq 0. \quad (38)$$

Now, in terms of the Euclidean action and the Chern-Simons number, this may be written as

$$4S_E - 2\left(\frac{16\pi^2}{g^2}\right)N_{CS} \geq 0, \tag{39}$$

which finally yields

$$S_E \geq \frac{8\pi^2}{g^2}N_{CS}. \tag{40}$$

The  $N_{CS}=1$  instanton saturates this bound, which means that the rate per unit volume of baryon-number-violating processes at zero temperature is approximately

$$\Gamma(T=0) \sim \exp(-2S_E) \sim 10^{-170}. \tag{41}$$

This is so small that if the universe were always close to zero temperature, not one event would have occurred within the present Hubble volume ever in the history of the universe.

### B. Nonzero-temperature results

We have seen that the rate of baryon-number-violating processes is negligible at zero temperature. This is to be expected, since such events occur through quantum tunneling under a high potential barrier ( $\sim 10$  TeV). If this were the case at all temperatures, it is safe to say that electroweak baryon number violation would have no observable consequences. In this section we shall see that, when one includes the effects of nonzero temperature, classical transitions between electroweak vacua become possible due to thermal activation.

#### 1. Mechanical analogy

Let us begin with a warmup example. Consider an ideal pendulum of mass  $m$  and length  $l$ , confined to rotate in the plane. The Lagrangian is

$$L = \frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - \cos\theta). \tag{42}$$

The system possesses a periodic vacuum structure labeled by integer  $n$ ,

$$\theta_n = 2n\pi. \tag{43}$$

The Schrödinger equation describing the quantum mechanics of this system is

$$\frac{\hbar\omega}{2}\left(-\alpha\frac{d^2}{d\chi^2} + \frac{1}{\alpha}\sin^2\chi\right)\psi_n(\chi) = E_n\psi_n(\chi), \tag{44}$$

where  $\chi \equiv \theta/2$ ,  $\omega \equiv g/l$ , and  $\alpha \equiv \hbar\omega/(4mgl) \ll 1$ . Since the potential is periodic in the angle  $\theta$ , the wave functions will be periodic and therefore the multiple vacuum structure of the system is manifest. However, if one uses perturbation theory, Taylor expanding the potential about a chosen minimum ( $\theta=0$ ) and keeping only the first term, then the Schrödinger equation becomes that for a simple harmonic oscillator and all information about the periodic vacua is lost. This situation is analogous to most familiar calculations in the electroweak theory, in which perturbation theory is usually a safe

tool to use. Such an approximation scheme is only valid when the energy of the pendulum is much less than the height of the barrier preventing transitions between vacua (see next subsection concerning the sphaleron). In that limit, quantum tunneling between vacua is exponentially suppressed as expected. This may be seen by employing a different approximation scheme which preserves periodicity—the semiclassical (WKB) approximation.

Now consider raising the temperature of the system. The pendulum is coupled to a thermal bath and is thermally excited to states of higher and higher energy as the temperature is raised. As the temperature becomes comparable with the barrier height, it becomes possible for thermal transitions over the barrier to occur. Writing the energy of the barrier between vacua as  $E_{bar}$ , we can show the rate of these transitions to be

$$\Gamma_{pendulum}(T) \propto \exp\left(-\frac{E_{bar}}{T}\right), \tag{45}$$

so that at  $T \sim E_{bar}$ , the pendulum makes transitions between vacua, crossing the point  $\theta = \pi$  randomly, at an unsuppressed rate. The important point here is that the rate of thermally activated transitions between vacua is governed by the barrier height, or more accurately, by the maximum of the free energy of the configurations that interpolate between the vacua.

The general features of this simple mechanical system are very important for a non-Abelian gauge theory such as the electroweak theory.

#### 2. Electroweak theory

The calculation of the thermally activated transition rate for the infinite-dimensional field theory is, of course, much more complicated than that for the one-dimensional example above. The field theory approach to problems of metastability was first outlined by Affleck (1981b; see also Linde, 1981), using techniques developed by Langer (1967) for use in condensed-matter problems. The general framework for evaluating the thermal rate of anomalous processes in the electroweak theory is due to Kuzmin *et al.* (1985).

For any sequence of configurations interpolating between vacua, there is one configuration that maximizes the free energy. Over all such sequences, there is one maximum energy configuration with the smallest free energy (i.e., a saddle point) and it is this configuration, in analogy with the pendulum example, that governs the rate of anomalous transitions due to thermal activation.

In the electroweak theory the relevant saddle-point configuration is the sphaleron. In the full Glashow-Salam-Weinberg theory, the sphaleron solution cannot be obtained analytically. In fact, the original calculation of Klinkhamer and Manton (1984) was performed in the  $SU(2)+\text{Higgs}$  theory, and identified a sphaleron with energy in the range

$$8 \text{ TeV} < E_{sph} < 14 \text{ TeV}, \tag{46}$$

depending on the Higgs mass. The corresponding object in the full electroweak theory can then be obtained from the  $SU(2)$  sphaleron by a perturbative analysis in which the small parameter is the weak mixing angle  $\theta_W$ .

To calculate the thermal baryon-number-violation rate, the following steps are performed. One first computes the rate to cross the barrier over the sphaleron beginning from a given state. This process is essentially a one-degree-of-freedom process, since all field directions perpendicular to the negative mode corresponding to the sphaleron are ignored. The second step is to sum over all such paths, weighting each by the appropriate Boltzmann factor. The calculation for the electroweak theory, properly taking into account the translational and rotational zero modes of the sphaleron, was first carried out by Arnold and McLerran (1987). A final step is to take account of the infinite number of directions transverse to the sphaleron by performing a calculation of the small fluctuation determinant around the sphaleron. This final step is carried out within the Gaussian approximation and was originally performed by Carson and McLerran (1990; see also Carson *et al.*, 1990), with more recent analyses by Baacke and Junker (1993, 1994a, 1994b). If  $M_W(T)$  is the thermal  $W$  boson mass calculated from the finite-temperature effective potential (see next section), the approximations that go into this calculation are valid only in the regime

$$M_W(T) \ll T \ll \frac{M_W(T)}{\alpha_W}. \quad (47)$$

The rate per unit volume of baryon-number-violating events in this range is calculated to be

$$\Gamma(T) = \mu \left( \frac{M_W}{\alpha_W T} \right)^3 M_W^4 \exp\left(-\frac{E_{sph}(T)}{T}\right), \quad (48)$$

where  $\mu$  is a dimensionless constant. Here, the temperature-dependent ‘‘sphaleron’’ energy is defined through the finite-temperature effective potential by

$$E_{sph}(T) \equiv \frac{M_W(T)}{\alpha_W} \mathcal{E}, \quad (49)$$

with the dimensionless parameter  $\mathcal{E}$  lying in the range

$$3.1 < \mathcal{E} < 5.4, \quad (50)$$

depending on the Higgs mass.

Recent approaches to calculating the rate of baryon-number-violating events in the broken phase have been primarily numerical. Several results using real-time techniques (Tang and Smit, 1996; Moore and Turok, 1997a) were found to contain lattice artifacts arising from an inappropriate definition of  $N_{CS}$ . These techniques were later shown to be too insensitive to the true rate when improved definitions of  $N_{CS}$  were used (Ambjørn and Krasnitz, 1997; Moore and Turok, 1997b). The best calculation to date of the broken-phase sphaleron rate is undoubtedly that by Moore (1998a). This work yields a fully nonperturbative evaluation of the broken-phase rate by using a combination of multicanonical weighting and real-time techniques.

Although the Boltzmann suppression in Eq. (48) appears large, it is to be expected that, when the electroweak symmetry is restored (Kirzhnits, 1972; Kirzhnits and Linde, 1972) at a temperature of around 100 GeV, there will no longer be an exponential suppression factor. Although calculation of the baryon-number-violating rate in the high-temperature phase is extremely difficult, a simple estimate is possible. In Yang-Mills Higgs theory at nonzero temperature the infrared modes (those with wave number  $k \ll T$ ) are well described classically in the weak-coupling limit, while the ultraviolet modes ( $k \sim T$ ) are not. Now, the only important scale in the symmetric phase is the magnetic screening length given by

$$\xi = (\alpha_W T)^{-1}. \quad (51)$$

Assuming that any time scale must also behave in this way, then on dimensional grounds, we expect the rate per unit volume of baryon-number-violating events, an infrared spacetime rate, to be

$$\Gamma(T) = \kappa (\alpha_W T)^4, \quad (52)$$

with  $\kappa$  another dimensionless constant, assuming that the infrared and ultraviolet modes decouple from each other.

The rate of baryon-number-violating processes can be related to the sphaleron rate, which is the diffusion constant for the Chern-Simons number [Eq. (20)] and is defined by

$$\lim_{V \rightarrow \infty} \lim_{t \rightarrow \infty} \left[ \frac{\langle [N_{CS}(t) - N_{CS}(0)]^2 \rangle}{Vt} \right] \quad (53)$$

[cf. Eq. (30)] by a fluctuation-dissipation theorem [Khlebnikov and Shaposhnikov (1988); for a good description of this, see Rubakov and Shaposhnikov (1996)]. In almost all numerical calculations of the baryon-number-violation rate, this relationship is used and what is actually evaluated is the diffusion constant. The first attempts to estimate  $\kappa$  numerically in this way (Ambjørn *et al.*, 1990, 1991) yielded  $\kappa \sim 0.1-1$ , but the approach suffered from limited statistics and large-volume systematic errors. Nevertheless, more recent numerical attempts (Ambjørn and Krasnitz, 1995; Moore, 1996; Tang and Smit, 1996; Moore and Turok, 1997a) found approximately the same result. However, as I mentioned above, these approaches employ a poor definition of the Chern-Simons number, which compromises their reliability.

The simple scaling argument leading to Eq. (52) was criticized by Arnold *et al.* (1997), who argued that damping effects in the plasma suppress the rate by an extra power of  $\alpha_W$ . The essential reason for this modification of Eq. (52) is that the decoupling between infrared and ultraviolet modes does not hold completely when dynamics are taken into account. Since the transition rate involves physics at soft energies  $g^2 T$  that are small compared to the typical hard energies  $\sim T$  of the thermal excitations in the plasma, the simplest way of analyzing the problem is to consider an effective theory for the soft modes. Thus one integrates out the hard modes and



keeps the dominant contributions, the so-called hard thermal loops. The resulting typical frequency  $\omega_c$  of a gauge-field configuration with spatial extent  $(g^2 T)^{-1}$  immersed in the plasma determines the change of baryon number per unit time and unit volume. This frequency has been estimated to be

$$\omega_c \sim g^4 T, \quad (54)$$

when the damping effects of the hard modes (Arnold *et al.*, 1997; Arnold, 1997) are taken into account.

In recent months these issues seem to have been resolved. The analysis of Moore and Turok (1997b) demonstrated that, when a reliable definition of  $N_{CS}$  is used, there is a lattice spacing dependence in the symmetric phase rate that is consistent with the claims of Arnold *et al.* (1997) and similar results were also obtained by Ambjørn and Krasnitz (1997). Huet and Son (1997) have constructed a nonlocal infrared effective theory that includes the effects of hard thermal loops, which they expect to be responsible for the extra power of  $\alpha_W$  in the rate of baryon number violation. Further, a rigorous field-theoretic derivation of this theory has been derived (Son, 1997). Using ideas developed by Hu and Mueller (1997), Moore *et al.* (1998) have also considered the effects of hard thermal loops [see also Iancu (1997)]. In that work the authors find

$$\Gamma(T \gg T_c) = \kappa' \alpha_W (\alpha_W T)^4, \quad (55)$$

with

$$\kappa' = 29 \pm 6 \quad (56)$$

for the minimal standard model value of the Debye mass. Note that, although this estimate takes into account physics that did not enter the original estimate, this expression is numerically close to Eq. (52).

Finally, the effective dynamics of soft non-Abelian gauge fields at finite temperature have recently been addressed by Bödeker (1998), who finds a further logarithmic correction

$$\Gamma_{sp} \sim \alpha_W^5 T^4 \ln(1/\alpha_W). \quad (57)$$

The physics leading to these corrections has been discussed at length by Moore (1998b), who describes in detail both the intuitive arguments for such a term and the lattice Langevin calculations required to provide an accurate numerical determination of its magnitude.

Now that I have discussed baryon number violation, I shall turn to the second Sakharov criterion and its realizations in the standard model.

### III. C AND CP VIOLATION

As I mentioned earlier, it is necessary that both the  $C$  and  $CP$  symmetries be violated for baryogenesis scenarios to succeed. One cause of the initial excitement over electroweak baryogenesis was the observation that the Glashow-Salam-Weinberg model naturally satisfies these requirements.

Recall that fermions in the theory are chirally coupled to the gauge fields as in Eq. (21). This means that, for

example, only the left-handed electron is  $SU(2)$  gauge coupled. In terms of the discrete symmetries of the theory, these chiral couplings result in the electroweak theory's being maximally  $C$  violating. This is a general feature that remains true in extensions of the theory and makes the model ideal for baryogenesis. However, the issue of  $CP$  violation is more complex.

#### A. Standard-model $CP$ violation: Kobayashi-Maskawa matrix

$CP$  is known not to be an exact symmetry of the weak interactions. This is seen experimentally in the neutral-kaon system through  $K_0, \bar{K}_0$  mixing (Christenson *et al.*, 1964). At present there is no accepted theoretical explanation of this. However, it is true that  $CP$  violation is a natural feature of the standard electroweak model. When Eq. (21) is expanded to include  $N$  generations of quarks and leptons, there exists a charged current which, in the weak interaction basis, may be written as

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} \bar{U}_L \gamma^\mu D_L W_\mu + (\text{H.c.}), \quad (58)$$

where  $U_L = (u, c, t, \dots)_L$  and  $D_L = (d, s, b, \dots)_L$ . Now, the quark mass matrices may be diagonalized by unitary matrices  $V_L^U, V_R^U, V_L^D, V_R^D$  via

$$\text{diag}(m_u, m_c, m_t, \dots) = V_L^U M^U V_R^U, \quad (59)$$

$$\text{diag}(m_d, m_s, m_b, \dots) = V_L^D M^D V_R^D. \quad (60)$$

Thus, in the basis of quark mass eigenstates, Eq. (58) may be rewritten as

$$\mathcal{L}_W = \frac{g}{\sqrt{2}} \bar{U}'_L K \gamma^\mu D'_L W_\mu + (\text{H.c.}), \quad (61)$$

where  $U'_L \equiv V_L^U U_L$  and  $D'_L \equiv V_L^D D_L$ . The matrix  $K$ , defined by

$$K \equiv V_L^U (V_L^D)^\dagger, \quad (62)$$

is referred to as the Kobayashi-Maskawa (KM) quark mass mixing matrix (Kobayashi and Maskawa, 1973). For  $N$  generations,  $K$  contains  $(N-1)(N-2)/2$  independent phases, and a nonzero value for any of these phases signals  $CP$  violation. Therefore  $CP$  violation exists for  $N \geq 3$ , and when  $N=3$ , as in the standard model, there is precisely one such phase  $\delta$ . While this is encouraging for baryogenesis, it turns out that this particular source of  $CP$  violation is not strong enough. The relevant effects are parametrized by a dimensionless constant which is no larger than  $10^{-20}$ . This appears to be much too small to account for the observed baryon asymmetry of the universe and, thus far, attempts to utilize this source of  $CP$  violation for electroweak baryogenesis have been unsuccessful.

When one includes the strong interactions described by quantum chromodynamics (QCD), a second potential source of  $CP$  violation arises due to instanton effects. However, precision measurements of the dipole moment of the neutron constrain the associated dimen-

sionless parameter  $\theta_{QCD}$  to be less than  $10^{-9}$ , which again is too small for baryogenesis. In light of these facts, it is usual to extend the standard model in some minimal fashion that increases the amount of  $CP$  violation in the theory while not leading to results that conflict with current experimental data.

## B. Two-Higgs-doublet model

One particular way of enhancing  $CP$  violation (McLerran, 1989) is to expand the Higgs sector of the theory to include a second Higgs doublet. In a two-Higgs model, the structure described in Sec. II is doubled so that we have scalars,  $\Phi_1$  and  $\Phi_2$ , and the scalar potential is replaced by the most general renormalizable two-Higgs potential (Gunion *et al.*, 1989),

$$\begin{aligned} V(\Phi_1, \Phi_2) = & \lambda_1(\Phi_1^\dagger \Phi_1 - v_1^2)^2 + \lambda_2(\Phi_2^\dagger \Phi_2 - v_2^2)^2 \\ & + \lambda_3[(\Phi_1^\dagger \Phi_1 - v_1^2) + (\Phi_2^\dagger \Phi_2 - v_2^2)]^2 \\ & + \lambda_4[(\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) - (\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1)] \\ & + \lambda_5[\text{Re}(\Phi_1^\dagger \Phi_2) - v_1 v_2 \cos \xi]^2 \\ & + \lambda_6[\text{Im}(\Phi_1^\dagger \Phi_2) - v_1 v_2 \sin \xi]^2. \end{aligned} \quad (63)$$

Here  $v_1$  and  $v_2$  are the respective vacuum expectation values of the two doublets, the  $\lambda_i$  are coupling constants, and  $\xi$  is a phase. To make the  $CP$  violation explicit, let us write the Higgs fields in unitary gauge as

$$\Phi_1 = (0, \varphi_1)^T, \quad \Phi_2 = (0, \varphi_2 e^{i\theta})^T, \quad (64)$$

where  $\varphi_1$ ,  $\varphi_2$ ,  $\theta$  are real, and  $\theta$  is the  $CP$ -odd phase. The important terms for  $CP$  violation in the potential are the final two, with coefficients  $\lambda_5$  and  $\lambda_6$ , since it is these terms that determine the dynamics of the  $CP$ -odd field  $\theta$ .

There are several mechanisms by which  $\theta$  may contribute to the free-energy density of the theory. To be specific, let us concentrate on the one-loop contribution (Turok and Zadrozny, 1990) and for simplicity assume that  $\theta$  is spatially homogeneous. The relevant term is

$$\mathcal{F}_B = -\frac{14}{3\pi^2 n_f} \zeta(3) \left(\frac{m}{T}\right)^2 \dot{\theta} n_B, \quad (65)$$

where  $m$  is the finite-temperature mass of the particle species dominating the contribution to the anomaly and  $\zeta$  is the Riemann function.

The coefficient of  $n_B$  in the above equation can be viewed as a sort of chemical potential,  $\mu_B$ , for baryon number,

$$\mu_B = \frac{14}{3\pi^2 n_f} \zeta(3) \left(\frac{m}{T}\right)^2 \dot{\theta}. \quad (66)$$

The changes in  $\theta$  are dependent on changes in the magnitude of the Higgs fields. In particular, if a point in space makes a transition from a false electroweak vacuum to a true one, then  $\Delta\theta > 0$ , and sphaleron processes result in the preferential production of baryons over antibaryons. For the opposite situation,  $\Delta\theta < 0$ , and sphaleron processes generate an excess of antibaryons.

The total change in the phase  $\theta$  (from before the phase transition to  $T=0$ ) may be estimated to be

$$\Delta\theta \sim \xi - \arctan\left(\frac{\lambda_6}{\lambda_5} \tan \xi\right), \quad (67)$$

and, as we shall see, it is this quantity that enters into estimates of the baryon asymmetry of the universe. This, then, is how the dynamics of the two-Higgs model bias the production of baryons through sphaleron processes in electroweak baryogenesis models. The use of the two-Higgs model is motivated in part by supersymmetry (SUSY), which demands at least two Higgs scalars. However, in the minimal supersymmetric standard model (MSSM), supersymmetry demands that  $\lambda_5 \equiv \lambda_6 = 0$  and so the two-Higgs potential is  $CP$  invariant. In such models  $CP$  violation arises through soft-SUSY breaking, which generates nonzero values for  $\lambda_5$  and  $\lambda_6$ , as we shall see later.

## C. $CP$ violation from higher-dimension operators

The second popular method of extending the standard model is to view the model as an effective-field theory, valid at energies below some mass scale  $M$ . In addition to the standard terms in the Lagrangian, one then expects extra, nonrenormalizable operators, some of which will be  $CP$  odd (Shaposhnikov, 1988; Dine *et al.*, 1991; Dine, Leigh, *et al.*, 1992; Zhang *et al.*, 1994; Lue *et al.*, 1997). A particular dimension-six example is

$$\mathcal{O} = \frac{b}{M^2} \text{Tr}(\Phi^\dagger \Phi) \text{Tr}(F_{\mu\nu} \tilde{F}^{\mu\nu}), \quad (68)$$

with  $b$  a dimensionless constant.  $\mathcal{O}$  is the lowest-dimension  $CP$ -odd operator that can be constructed from minimal standard-model Higgs and gauge fields. Such a term, with  $M = v$ , can be induced in the effective action by  $CP$  violation in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, but in that case the coefficient  $b$  is thought to be tiny.

The operator  $\mathcal{O}$  induces electric dipole moments for the electron and the neutron (Zhang and Young, 1994), and the strongest experimental constraint on the size of such an operator comes from the fact that such dipole moments have not been observed. Working to lowest order (one loop), Lue *et al.* (1997) find

$$\frac{d_e}{e} = \frac{m_e \sin^2(\theta_W)}{8\pi^2} \frac{b}{M^2} \ln\left(\frac{M^2 + m_H^2}{m_H^2}\right). \quad (69)$$

Note that  $M^2$  arises in the logarithm without  $b$  because  $M$ , the scale above which the effective theory is not valid, is the ultraviolet cutoff for the divergent loop integral. Using the experimental limit (Commins *et al.*, 1994)

$$\frac{d_e}{e} < 4 \times 10^{-27} \text{ cm} \quad (70)$$

yields the bound

$$\frac{b}{M^2} \ln \left( \frac{M^2 + m_H^2}{m_H^2} \right) < \frac{1}{(3 \text{ TeV})^2}. \quad (71)$$

The corresponding experimental limit (Smith *et al.*, 1990; Altarev *et al.*, 1992) on the neutron electric dipole moment  $d_n$  is weaker than that on  $d_e$ , but because  $d_n$  is proportional to the quark mass rather than to the electron mass, the constraint obtained using  $d_n$  is comparable to Eq. (71). Any baryogenesis scenario that relies on  $CP$  violation introduced via the operator  $\mathcal{O}$  must respect the bound (71).

#### D. General treatment

Common to both the above extensions of the standard model is the appearance of extra,  $CP$ -violating interactions, parametrized by a new quantity,  $\delta_{CP}$  (e.g.,  $\delta = \Delta\theta$  or  $\delta = b/M^2$ ). Although there are a number of other ways in which  $CP$  violation may appear in low-energy electroweak models (see, for example, Frere *et al.*, 1993; Reina and Tytgat, 1994), in many parts of this review, for definiteness, I shall focus on  $\Delta\theta$ . However, in the discussion of specific SUSY models I shall explain how  $CP$ -violating quantities arise. Whatever its origin, the effect of  $CP$  violation on anomalous baryon-number-violating processes is to provide a fixed direction for the net change in baryon number. When the bias is small, the equation governing this can be derived from detailed balance arguments (Khlebnikov and Shaposhnikov, 1998; Dine *et al.*, 1990) and may be written as

$$\frac{dn_B}{dt} = -3 \frac{\Gamma(T)}{T} \Delta F, \quad (72)$$

where  $\Delta F$  is the free-energy difference, induced by the  $CP$  violation, between producing baryons and antibaryons in a given process, and  $\Gamma$  is the rate per unit volume of baryon-number-violating events. I shall use this equation further when considering nonlocal baryogenesis.

#### IV. ELECTROWEAK PHASE TRANSITION

To begin with I shall lay down some definitions from thermodynamics. If a thermodynamic quantity changes discontinuously (for example as a function of temperature), then we say that a *first-order phase transition* has occurred. This happens because, at the point at which the transition occurs, there exist two separate thermodynamic states that are in equilibrium. Any thermodynamic quantity that undergoes such a discontinuous change at the phase transition is referred to as an *order parameter*, denoted by  $\varphi$ . Whether or not a first-order phase transition occurs often depends on other parameters that enter the theory. It is possible that, as another parameter is varied, the change in the order parameter at the phase transition decreases until it and all other thermodynamic quantities are continuous at the transition point. In this case we refer to a *second-order phase transition* at the point at which the transition becomes continuous, and a *continuous crossover* at the other

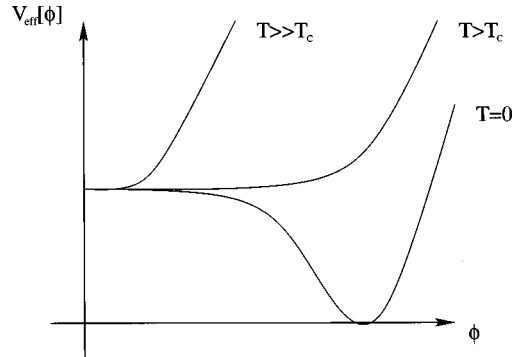


FIG. 4. Sketch of the finite-temperature effective potential for various values of temperature for a second-order phase transition.

points for which all physical quantities undergo no changes. In general, we are interested in systems for which the high-temperature ground state of the theory is achieved for  $\varphi = 0$  and the low-temperature ground state is achieved for  $\varphi \neq 0$ .

The question of the order of the electroweak phase (EWPT) transition is central to electroweak baryogenesis. Phase transitions are the most important phenomena in particle cosmology, since without them, the history of the universe is simply one of gradual cooling. In the absence of phase transitions, the only departure from thermal equilibrium is provided by the expansion of the universe. At temperatures around the electroweak scale, the expansion rate of the universe in thermal units is small compared to the rate of baryon-number-violating processes. This means that the equilibrium description of particle phenomena is extremely accurate at electroweak temperatures. Thus baryogenesis cannot occur at such low scales without the aid of phase transitions [for a treatment of this argument in the context of non-standard cosmologies, in which the universe is not radiation dominated at the electroweak scale, see Joyce and Prokopec (1998)].

For a continuous transition, the extremum at  $\varphi = 0$  becomes a local maximum at  $T_c$  and thereafter there is only a single minimum at  $\varphi \neq 0$  (see Fig. 4). At each point in space thermal fluctuations perturb the field which then rolls classically to the new global minimum of the finite temperature effective potential. Such a process is referred to as *spinodal decomposition*. If the EWPT is second-order or a continuous crossover, the associated departure from equilibrium is insufficient to lead to relevant baryon number production (Kuzmin *et al.*, 1985). This means that, for EWBG to succeed, we either need the EWPT to be strongly first order or need other methods of destroying thermal equilibrium, for example topological defects (see Sec. VIII), to be present at the phase transition.

For a first-order transition the extremum at  $\varphi = 0$  becomes separated from a second local minimum by an energy barrier (see Fig. 5). At the critical temperature  $T = T_c$  both phases are equally favored energetically and at later times the minimum at  $\varphi \neq 0$  becomes the global minimum of the theory. The dynamics of the phase tran-

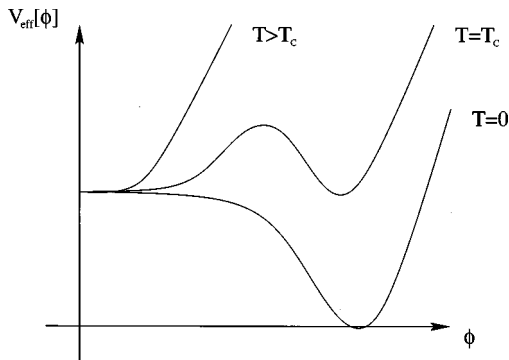


FIG. 5. Sketch of the finite-temperature effective potential for various values of temperature for a first-order phase transition.

sition in this situation are crucial to most scenarios of electroweak baryogenesis. The essential picture is that at temperatures around  $T_c$  quantum tunneling occurs and nucleation of bubbles of the true vacuum in the sea of false begins. Initially these bubbles are not large enough for their volume energy to overcome the competing surface tension and they shrink and disappear. However, at a particular temperature below  $T_c$  bubbles just large enough to grow nucleate. These are termed *critical* bubbles, and they expand, eventually filling all of space and completing the transition. As the bubble walls pass each point in space, the order parameter changes rapidly, as do the other fields, and this leads to a significant departure from thermal equilibrium. Thus, if the phase transition is strongly enough first order, it is possible to satisfy the third Sakharov criterion in this way.

There exists a simple equilibrium analysis of bubble nucleation in a first-order phase transition [see, for example, Rubakov and Shaposhnikov (1996)]. Write the bubble nucleation rate per unit volume at temperature  $T$  as  $R(T)$ . Further, note that if we assume that bubbles expand at constant speed  $v$ , then the volume occupied at time  $t$  by a bubble that nucleated at time  $t_0$  is

$$V(t, t_0) = \frac{4\pi}{3} (t - t_0)^3 v^3. \quad (73)$$

Then, the fraction of the total volume occupied by the broken phase at time  $t$  can be written as

$$P(t) = 1 - \exp[-\Sigma(t)], \quad (74)$$

where the quantity  $\Sigma(t)$  is given by

$$\Sigma(t) = \int_{t_c}^t dt_0 V(t, t_0) R(T_0) \quad (75)$$

and  $t_c$  and  $T_0$  are defined through  $T(t_c) = T_c$  and  $T(t_0) = T_0$ , respectively. In terms of this quantity, we say that the phase transition is complete when  $\Sigma \approx 1$ . In order to estimate  $\Sigma(t)$  note that, when bubble nucleation begins, the quantity

$$x \equiv \frac{T_c - T}{T_c} \quad (76)$$

is small. Making this change of variables in Eq. (75) and using the time-temperature relationship  $t = \tilde{M}/T^2$ , with  $\tilde{M} \sim 10^{18}$  GeV we obtain, for small  $x$ ,

$$\Sigma(t) = \frac{64\pi v^3}{3} \left( \frac{\tilde{M}}{T_c} \right)^4 \int_0^x dx x^3 \frac{1}{T_c^4} R(T). \quad (77)$$

From this expression it is clear that, when  $\Sigma$  is of order one, when the phase transition completes or *percolates*, the rate per unit volume is negligibly small.

In general, the calculation of the bubble nucleation rate is extremely complicated. The relevant time scale  $\tau_{\text{form}}$  for the formation of a critical bubble is of the form

$$\tau_{\text{form}} \propto \exp(F_c/T), \quad (78)$$

where  $F_c$  is the free energy of the bubble configuration. We may calculate this by writing the bubble as a spherically symmetric configuration  $\varphi(r)$  which extremizes the effective action (as defined in the next subsection) and satisfies

$$\lim_{r \rightarrow \infty} \varphi(r) = 0,$$

$$\varphi(0) = v(T),$$

where  $v(T)$  is the minimum of the finite-temperature effective potential  $V(\varphi, T)$  defined properly in the next subsection. Then the free energy is given by

$$F_c[\varphi(r)] \approx \int dr r^2 \left[ \frac{1}{2} (\varphi')^2 + V(\varphi, T) \right] \quad (79)$$

with the bubble configuration obtained by solving

$$\frac{1}{r^2} \frac{d}{dr} (r^2 \varphi') - \frac{\partial V}{\partial \varphi} = 0, \quad (80)$$

subject to the above boundary conditions. However, in the general situation these equations must be solved numerically, although some progress can be made in the *thin-wall limit* in which the typical size of the critical bubbles is much larger than the correlation length of the system.

The precise evolution of critical bubbles in the electroweak phase transition is a crucial factor in determining which regimes of electroweak baryogenesis are both possible and efficient enough to produce the baryon asymmetry of the universe. In essence, the bubble wall dynamics are governed by the interplay between surface tension and volume pressure.

The physics of a propagating phase boundary, or bubble wall, have been extensively investigated by many authors.<sup>1</sup> The crucial quantities that one wishes to estimate are the wall velocity  $v$  and the wall width  $\delta$ . As will become clear later, it is important to know whether the velocity is less than or greater than the speed of sound in the plasma, because different mechanisms of baryogenesis dominate in these regimes. As it turns out, the dy-

<sup>1</sup>See, for example, Dine, Huet, and Singleton, 1992; Enqvist *et al.*, 1992; Khlebnikov, 1992; Liu *et al.*, 1992; Turok, 1992; Arnold, 1993; Huet *et al.*, 1993; Ignatius *et al.*, 1994; Laine, 1994; Heckler, 1995; Kurki-Suonio and Laine, 1995a, 1996a, 1996b; Moore and Prokopec, 1995a, 1995b; Duari and Yajnik, 1996.

namics of the bubble wall are qualitatively different in these two regimes also (Steinhardt, 1982; Gyulassy *et al.*, 1983).

For example, consider a single bubble of broken electroweak symmetry expanding in a sea of symmetric phase plasma, and for simplicity assume that the bubble is large enough that we may neglect its curvature and idealize the wall as a planar interface. There are two relevant regimes:

- (1) The wall velocity is less than the speed of sound ( $v < c_s \approx 0.58$ ): In this case it is said that *deflagration* occurs. The plasma near the bubble wall in the unbroken phase accelerates away from the wall and a shock wave develops ahead of the wall moving at speed  $v_{sw} > v$ . This results in the heating of the plasma behind the shock front [for a detailed treatment of deflagration see Kurki-Suonio (1985)].
- (2) The wall velocity is greater than the speed of sound ( $v > c_s$ ): In this case we say that *detonation* occurs. In contrast to deflagration, the plasma ahead of the wall is now at rest, whereas that behind the wall in the broken phase accelerates away.

Which of these regimes is relevant for a given phase transition depends on a host of microphysical inputs, making an analytic approach extremely difficult. However, recent investigations of the bubble wall behavior in the standard model have been performed by Moore and Prokopec (1995a, 1995b). These authors find that if  $m_H < 90$  GeV (encompassing the whole region of physically allowed Higgs masses for which a strongly first-order phase transition is expected to occur), the phase transition proceeds through deflagration. They find a robust upper bound on bubble velocities

$$v < 0.45 < v_{sw} \approx c_s \approx 0.58, \quad m_H < 90 \text{ GeV}. \quad (81)$$

Other recent work (Moore and Turok, 1997b) has demonstrated the importance of friction from infrared W bosons on the wall velocity and suggests that velocities of the order

$$v \sim 0.1 - 0.2 \quad (82)$$

may be realistic. Of course, the methods used in this analysis cannot be extended past the point at which the shock waves originating from different bubbles would collide, since at that stage the approximation of an isolated wall is no longer valid. In fact, the dynamics of the phase transition from that point to completion are quite different from the simple picture I have described above. What seems clear is that significant heating of the plasma occurs, perhaps up to  $T_c$  (Kurki-Suonio and Laine, 1996b), as the latent heat of the transition is released. This should result in an appreciable deceleration of the bubble walls until, finally, the transition is completed by relatively slow-moving bubbles. In fact, some analytical progress may be made in the limit in which one assumes that the latent heat of the transition is released instantaneously into the plasma (Heckler, 1995).

While all these approaches are useful in understanding the nature of bubble walls, it is important to remember that they are performed in the minimal standard model. When a particular extension of the electroweak theory is used, the analysis should be repeated in that context.

In practice it can be a difficult task to determine the order of a given phase transition and thus whether, and how, bubble walls may arise. In the remainder of this section I shall review some of the analytical and numerical approaches that are used and discuss their application to the electroweak phase transition.

### A. Finite-temperature effective potential

A widely used tool in studying thermal phase transitions is the *finite-temperature effective potential* (FTEP) for the order parameter. I have already mentioned this in passing when discussing bubble nucleation. Here I shall define this object precisely and explain how it is calculated in a simple model. I shall then give the form of the potential for the electroweak theory and show how it is used.

Let us begin at zero temperature and consider a single scalar field  $\varphi$  with external source  $J$ . The generating functional for this theory is

$$Z[J] = \int \mathcal{D}\varphi \exp \left[ i \int d^4x (\mathcal{L}[\varphi] + J\varphi) \right]. \quad (83)$$

From this quantity the field theory analog  $E[J]$  of the Helmholtz free energy is defined by

$$e^{-iE[J]} \equiv Z[J]. \quad (84)$$

As is well known, functional differentiation of  $E[J]$  with respect to the source  $J$  defines the classical field through

$$\frac{\delta}{\delta J(x)} E[J] = -\varphi_{cl}. \quad (85)$$

Now, in thermodynamics it is usual to construct the Gibbs free energy of the theory by a Legendre transform of the Helmholtz free energy. In field theory we perform an analogous transformation to define the *effective action* by

$$\Gamma[\varphi_{cl}] \equiv -E[J] - \int d^4y J(y) \varphi_{cl}(y). \quad (86)$$

This functional has the useful property that

$$J(x) = -\frac{\delta}{\delta \varphi_{cl}(x)} \Gamma[\varphi_{cl}], \quad (87)$$

which means that, in the absence of an external source, the effective action defined by Eq. (86) satisfies

$$\frac{\delta}{\delta \varphi_{cl}(x)} \Gamma[\varphi_{cl}] = 0. \quad (88)$$

Thus the values of the classical fields in the vacua of the theory are obtained by solving this equation, i.e., by extremizing the effective action.

One further simplification is possible. If the vacua of the theory are translation and Lorentz invariant, then

$\varphi_{cl}$  is constant. In that case,  $\Gamma[\varphi_{cl}]$  contains no derivatives of  $\varphi_{cl}$  and Eq. (88) is an ordinary equation. It is therefore convenient to define the *effective potential* by

$$V_{\text{eff}}(\varphi_{cl}) \equiv -\frac{1}{\mathcal{V}_4} \Gamma[\varphi_{cl}], \quad (89)$$

where  $\mathcal{V}_4$  is the spacetime four-volume, so that the equation leading to the vacua of the theory reduces to

$$\frac{\partial}{\partial \varphi_{cl}} V_{\text{eff}}(\varphi_{cl}) = 0. \quad (90)$$

Note that this equation allows one, in principle, to compute the vacua of the theory exactly, taking into account all corrections to the bare potential from quantum fluctuations in the field  $\varphi$ .

Exact analytic calculations of the effective potential are difficult. Therefore it is usual to use perturbation theory. As an example, in the scalar field model above we choose the bare potential to be

$$V(\varphi) = -\frac{\mu^2}{2} \varphi^2 + \frac{\lambda}{4} \varphi^4, \quad (91)$$

with  $\mu$  an arbitrary mass scale and  $\lambda$  a parameter. We write  $\varphi$  as

$$\varphi(x) = \varphi_{cl} + \chi(x), \quad (92)$$

and the aim is to take account of the small quantum fluctuations in  $\chi$  around the classical vacuum state  $\varphi_{cl}$ . These fluctuations can contribute to both the energy of the vacuum state and the potential felt by  $\varphi_{cl}$ , since the mass of the field  $\chi$  is due to  $\varphi_{cl}$ . The perturbation-theory approach to calculating the effects of quantum fluctuations is usually phrased in the language of Feynman diagrams. At the one-loop level, there is only one diagram, which is a single  $\chi$  loop. This loop yields a contribution to the effective potential of

$$V^{(1)}(\varphi_{cl}) = \frac{1}{2(2\pi)^4} \int d^4k \ln[k^2 + 3\lambda\varphi_{cl}^2 - \mu^2]. \quad (93)$$

If we perform the  $k_0$  integration and remove an infinite constant, this becomes

$$V^{(1)}(\varphi_{cl}) = \frac{1}{(2\pi)^3} \int d^3k k [|\mathbf{k}|^2 + 3\lambda\varphi_{cl}^2 - \mu^2]^{1/2}. \quad (94)$$

This expression can now be seen to correspond to an integral over the energy of a  $\chi$  particle in all momentum modes. Note that, since this energy depends on  $\varphi_{cl}$ , normal ordering cannot remove this contribution to the energy.

Since Eq. (94) is divergent, we must choose renormalization conditions to perform the integration. If we choose to renormalize at the point  $\varphi = \mu\lambda^{-1/2}$ , and there set  $V'(\varphi) = V''(\varphi) = 0$ , then the one-loop zero-temperature effective potential becomes

$$V_{\text{eff}}^{(1)}(\varphi) = -\frac{\mu^2}{2} \varphi^2 + \frac{\lambda}{4} \varphi^4 + \frac{1}{64\pi^2} (3\lambda\varphi^2 - \mu^2)^2 \times \ln\left(\frac{3\lambda\varphi^2 - \mu^2}{2\mu^2}\right) + \frac{21\lambda\mu^2}{64\pi^2} \varphi^2 - \frac{27\lambda^2}{128\pi^2} \varphi^4. \quad (95)$$

We now wish to incorporate the effects of thermal fluctuations into this picture. Just as quantum fluctuations of fields lead to a modification of the potential, thermal fluctuations have an analogous effect. The potential resulting from taking account of thermal fluctuations is the finite-temperature effective potential (FTEP). The formalism we set up for the zero-temperature case is applicable here also because of the well-known connection between zero-temperature field theory and thermal field theory. The path-integral formulation of field theory at nonzero temperature  $T$ , describing the equilibrium structure of the theory, is formally equivalent to the zero-temperature formalism performed over a Euclidean time interval of length  $\beta \equiv 1/T$ . In addition, one must impose appropriate boundary conditions on the fields, periodic for bosons and antiperiodic for fermions. Thus any field  $\chi$  may be Fourier expanded over this Euclidean time interval, yielding the expression

$$\chi(x, \tau) = \sum_{n=-\infty}^{\infty} \chi_n(x) e^{i\omega_n \tau}, \quad (96)$$

with

$$\omega_n = \begin{cases} 2n\pi T & \text{bosons} \\ (2n+1)\pi T & \text{fermions.} \end{cases} \quad (97)$$

Since the Euclidean time coordinate is now finite, the zeroth component of a particle's four-momentum is now discrete. When we perform an integral over  $k_0$ , as in the quantum case above, this becomes a sum. This means that the one-loop temperature-dependent contribution to the FTEP is

$$V^{(1)}(\varphi, T) = \frac{T}{2(2\pi)^3} \sum_{n=-\infty}^{+\infty} \int d^3k \times \ln[(2\pi n T)^2 + |\mathbf{k}|^2 + 3\lambda\varphi^2 - \mu^2]. \quad (98)$$

If we define  $m^2(\varphi) \equiv 3\lambda\varphi^2 - \mu^2$ , then for  $m \ll T$  we may perform the sum and integral to first order in  $\lambda$ . After renormalization we add this to the bare potential to give the one-loop FTEP as

$$V_{\text{eff}}^{(1)}(\varphi, T) = -\frac{\mu^2}{2} \varphi^2 + \frac{\lambda}{4} \varphi^4 + \frac{m^2(\varphi)}{24} T^2 - \frac{\pi^2}{90} T^4. \quad (99)$$

This can be thought of as the contribution to the  $\varphi$  potential from the energy of, and interaction with, a thermal bath of particles at temperature  $T$ .

In general, in a theory with spontaneous symmetry breaking, the usual high-temperature behavior of the theory is that of the zero-temperature theory at high energies. That is to say that the full symmetry of the Lagrangian is restored at high temperatures. At high

temperatures the global minimum of the FTEP is at  $\varphi = 0$  and at zero temperature the global minimum is at  $\varphi = v$ , where  $v$  is the usual Higgs vacuum expectation value (VEV).

In the standard electroweak theory [ignoring the  $U(1)$  terms] the high-temperature form of the one-loop FTEP is

$$V_{\text{eff}}^{(1)}(\varphi; T) = \left( \frac{3}{32}g^2 + \frac{\lambda}{4} + \frac{m_t^2}{4v^2} \right) (T^2 - T_*^2) \varphi^2 - \frac{3g^2}{32\pi} T \varphi^3 + \frac{\lambda}{4} \varphi^4, \quad (100)$$

with  $\varphi \equiv \sqrt{\phi^\dagger \phi}$ , and  $m_t$  the top quark mass. Note that, in calculating this quantity for a gauge theory, one also takes into account loop diagrams corresponding to fluctuations in the gauge fields. The lowest-order thermal correction to the zero-temperature effective potential is a temperature-dependent mass. It is this contribution that causes the extremum at  $\varphi = 0$  to be the global minimum at high temperatures. Note also the presence of a cubic term in the effective potential. This term arises from the gauge-field fluctuations; it is responsible for the existence of a barrier separating two degenerate vacua at the critical temperature and hence for the prediction that the electroweak phase transition must be first order. Using this one-loop potential, we may estimate the critical temperature at which the vacua are degenerate to be

$$T_c = m_H \left( \frac{3}{8}g^2 + \lambda - \frac{9}{256\pi^2}g^6 + \frac{m_t^2}{v^2} \right)^{-1/2}. \quad (101)$$

Calculations of quantities such as the critical temperature can be refined if the effective potential is calculated to higher orders in perturbation theory. Recently the two-loop calculation has been performed (Arnold and Espinosa, 1993; Farakos *et al.*, 1994a, 1994b; Fodor and Hebecker, 1994; Buchmuller *et al.*, 1995).

It is worthwhile including a cautionary note concerning the FTEP. If the interactions are weak, i.e., if

$$\frac{g^2 T}{\pi m_W(\varphi)} \ll 1, \quad (102)$$

then the effective-potential approach is equivalent to solving the full equations of motion for the gauge and Higgs fields with all fluctuations taken into account. However, in a perturbative approximation it is very difficult to go beyond two loops, which limits the accuracy of the method. Further, when the phase transition is weak, the infrared degrees of freedom become strongly coupled (Linde, 1980; Gross *et al.*, 1981), the inequality (102) is no longer satisfied, and we must turn to numerical approaches.

Note that the higher the Higgs mass is, the weaker the phase transition gets. The current lower bound on the mass of the Higgs in the minimal standard model (MSM) comes from combining the results of the DELPHI, L3, and OPAL experiments at LEP and is

$$m_H > 89.3 \text{ GeV} \quad (103)$$

(de Jong, 1998), so the perturbation expansion might not be expected to converge.

## B. Dimensional reduction

As I have already explained, equilibrium field theory at nonzero temperature can be formulated as zero-temperature field theory performed over a finite interval of Euclidean time with period  $1/T$ .

If we expand the fields as in Eq. (96), we may think of the theory as a three-dimensional system of an infinite number of fields  $\chi_n$ . These fields comprise a tower of states of increasing masses proportional to the Matsubara frequencies  $\omega_n$ . If the theory is weakly coupled (as we expect), then we may make a simplifying approximation. For momenta much less than the temperature, we may perturbatively account for the effects of all fields with nonzero thermal masses. Thus, since all the fermions are massive, what remains is an effective theory of only the bosonic fields with  $n=0$ , the zero modes.

To see how the remaining fields contribute to the dynamics of such a theory consider the following equivalent construction. Write down the most general, renormalizable Lagrangian for the zero modes in three dimensions. For this to be the effective theory we seek, it is necessary that the free parameters be determined by matching the one-particle irreducible (1PI) Green's functions with those of the full theory in four dimensions. The massive degrees of freedom are important for this matching condition and thus contribute to the masses and couplings in the three-dimensional effective theory.

The three-dimensional (3D) effective theory is much simpler than the full theory. For the minimal standard model, the particle content is the Higgs doublet, the 3D  $SU(2) \times U(1)$  gauge fields, and extra bosonic fields corresponding to the temporal components of the gauge fields in the full theory. For this effective theory, perturbative calculations of the coupling constants to one loop (Farakos *et al.*, 1994a) and masses to second order (Kajantie *et al.*, 1996a) have been performed.

In the region of the phase transition itself, a further simplification is possible. In this regime, the masses of the extra bosons in the theory are proportional to  $gT$  or to  $g'T$  and are heavy compared to the effective Higgs mass of the theory. Thus we may integrate out these heavy degrees of freedom and obtain a simple, effective 3D theory that describes the system near the phase transition (Ginsparg, 1980; Applequist and Pisarski, 1981; Nadkarni, 1983; Landsman, 1989; Farakos *et al.*, 1994a, 1994b; Laine, 1995). The appropriate Lagrangian is

$$\mathcal{L} = \frac{1}{4} W_{ij}^a W_{ij}^a + \frac{1}{4} F_{ij} F_{ij} + (D_i \phi)^\dagger D_i \phi + m_3^2 \phi^\dagger \phi + \lambda_3 (\phi^\dagger \phi)^2, \quad (104)$$

where  $m_3$  and  $\lambda_3$  are the effective 3D Higgs mass and self-coupling in the theory near the phase transition.

This theory is *superrenormalizable* since  $m_3^2$  has only linear and log divergences and, in the  $\overline{\text{MS}}$  scheme,  $\lambda_3$

and  $g_3$ , the three-dimensional gauge coupling, do not run at all. Thus the theory described by Eq. (104) is very powerful, since the perturbative calculations of the parameters involved do not suffer from any infrared divergences. As a result, the theory is valid over a large range,

$$30 \text{ GeV} < m_H < 240 \text{ GeV}, \quad (105)$$

of Higgs masses. In addition, removal of the ultraviolet divergences of the theory is simple. Further, the vastly decreased number of degrees of freedom makes the analysis of this theory much more manageable. One might wonder if the results from this approach can be checked against those from the full 4D theory in any regimes. In fact, at high temperatures, calculations of the 4D effective potential in which hard thermal loops are “resummed” (Arnold and Espinosa, 1993; Fodor and Hebecker, 1994; Farakos *et al.*, 1994a) agree well with the 3D results (for a nonperturbative approach to checking this agreement see Laine, 1996b). For an alternative treatment of the effective 3D theory, in which a cutoff is introduced rather than explicitly using the superrenormalizability of the theory, see Karsch *et al.* (1996).

While we may construct the 3D effective theory without worrying about divergences in its parameters, perturbative calculations using the theory are infrared divergent in the symmetric phase of the system. This means that perturbation theory is not a particularly effective tool and nonperturbative approaches are necessary. It is usual to treat the simpler  $SU(2)$  system, in which the  $U(1)_Y$  interactions are ignored, to gain insight into the nonperturbative dynamics of the electroweak phase transition. In the context of the 3D effective theory I have described, a number of nonperturbative methods have been used by different groups.

### C. Numerical simulations—lattice approaches

The second approach, which has been very successful for analyzing the nature of phase transitions at nonzero temperatures, is to simulate the systems numerically on a lattice. These approaches are technically difficult to perform but have the advantage of covering the entire range of parameter space that might be interesting for electroweak baryogenesis. Here I shall provide an outline of how these approaches are performed and give the essential results. Numerical simulations of the electroweak phase transition have been performed in both the 3D theory described above and in the full 4D theory (Csikor *et al.*, 1996). However, in the 4D simulations the situation is complicated in the physically allowed range of Higgs masses. Nevertheless, considerable progress has been made.

Reliable quantitative results for the order of the electroweak phase transition have been obtained from lattice Monte Carlo simulations of the effective 3D theory (Kajantie *et al.*, 1993, 1996a, 1996b, 1996c, 1996d; Farakos *et al.*, 1994b, 1994c; Dosch *et al.*, 1996, 1997; Gurtler, Ilgenfritz, Kripfgang, *et al.*, 1997; Gurtler, Ilgenfritz, and Schiller, 1997; Gurtler *et al.*, 1998a, 1998b). This ap-

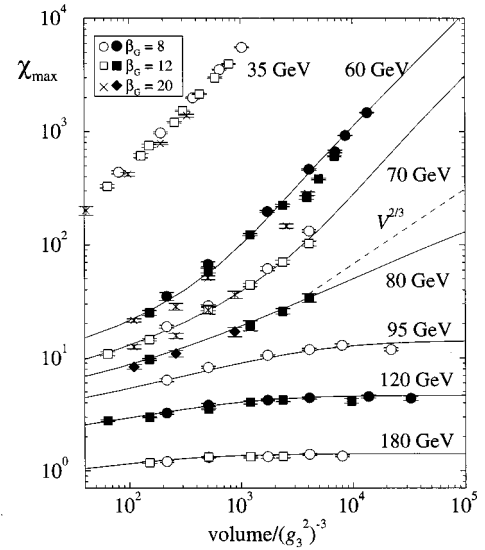


FIG. 6. The maximum susceptibility  $\chi_{\max}$  as a function of volume for different values of the Higgs mass  $m_H$ : Solid lines, mean-field fits; dashed line, the mean-field critical exponent. Thanks to Kari Rummukainen for supplying this figure.

proach is technically complicated and one must take great care to ensure that the observed features are not artifacts of the lattice implementation of the theory. For a discussion of these matters and an excellent comprehensive treatment of the method, I refer the reader to Kajantie *et al.* (1996b). The Monte Carlo calculations are performed over a range of system volumes  $V$  and lattice spacings  $a$ . The important results are obtained in the limit of extrapolation to infinite volume ( $V \rightarrow \infty$ ) and zero lattice spacing ( $a \rightarrow 0$ ). Here, I shall just sketch how these lattice Monte Carlo calculations are performed.

The most widely studied lattice model is the pure  $SU(2)$  theory. The parameters of the 3D theory are matched with those of the 4D continuum theory by renormalization-group methods. In particular, the top-quark mass is matched to the known value  $m_t \approx 175$  GeV. More recently, the full bosonic 3D  $SU(2) \times U(1)$  theory has been simulated (Kajantie *et al.*, 1996c).

The approach of Kajantie *et al.* is to measure the volume (finite-size) behavior of the susceptibility  $\chi$  of the operator  $\phi^\dagger \phi$ . The susceptibility is defined in terms of quantities in the 3D effective Lagrangian (104) by

$$\chi \equiv g_3^2 V \langle (\phi^\dagger \phi - \langle \phi^\dagger \phi \rangle)^2 \rangle. \quad (106)$$

For a given Higgs mass, the procedure is as follows. First, at fixed volume  $V$ , one determines the maximum value  $\chi_{\max}$  of  $\chi$  as a function of temperature. This procedure is then performed over a range of  $V$  to give  $\chi_{\max}(V)$ . From finite-size scaling arguments, the following behaviors are expected:

$$\chi_{\max}(V) \propto \begin{cases} V & \text{first-order transition} \\ V^{\gamma/3} & \text{second-order transition} \\ \text{const} & \text{no transition,} \end{cases} \quad (107)$$

with  $\gamma \neq 0, 3$  a critical exponent. Using this criterion, it is



possible to infer the nature of the phase transition from the numerical simulations. A summary of the results (Kajantie *et al.*, 1996d) is shown in Fig. 6. The general result of the Monte Carlo approach is that the phase transition is first order for Higgs masses  $m_H \leq 80$  GeV and becomes a smooth crossover for  $m_H \geq 80$  GeV. The existence of an end point of first-order phase transitions was shown by Kajantie *et al.* (1996d), and the position of the end point was identified by Gurtler, Ilgenfritz, and Schiller (1997). This end point, at  $m_H \sim 80$  GeV, falls into the universality class of the 3D Ising model. The end point of the transition has also been examined in the 4D simulations by Aoki (1997), with results in agreement with the 3D simulations.

Although the results I quote here are obtained in the  $SU(2)$  theory [equivalently  $\sin^2 \theta_W = 0$  in  $SU(2) \times U(1)$ ], the effects of the additional  $U(1)$  of the true electroweak theory have been considered by Kajantie *et al.* (1996c). However, while the effects of hypercharge strengthen the phase transition in general, the position of the end point is relatively insensitive to the value of  $\sin^2 \theta_W$ . I shall mention briefly in the next section how these results compare with those obtained from other approaches.

As the Higgs mass is increased to 80 GeV from below, the strength of the first-order phase transition decreases. This behavior is quantified by the tension of the phase interface at  $T_c$  and by the latent heat  $L$  of the transition. Both these quantities are measured in the simulations and the expected weakening of the transition is observed. The Clausius-Clapeyron equation

$$\frac{d\Delta p}{dT} = \frac{L}{T}, \quad (108)$$

where  $\Delta p$  is the difference in pressures between the symmetric and broken phases at temperature  $T$ , is particularly useful here. The equation may be written as

$$\Delta \langle \phi^\dagger \phi \rangle \left( \frac{m_H^2}{T_c^3} \right) = \frac{L}{T_c^4}, \quad (109)$$

where  $\Delta \langle \phi^\dagger \phi \rangle$  is the jump in the squared order parameter. As we shall use later, the measurement of the latent heat from the simulations yields a value for  $\Delta \langle \phi^\dagger \phi \rangle$  that we shall find very useful when discussing local electroweak baryogenesis.

#### D. Other approaches

The application of  $\varepsilon$ -expansion techniques to the electroweak phase transition was suggested by Gleiser and Kolb (1993) and has been investigated in detail by Arnold and Yaffe (1994). Since it is difficult to solve the 3D theory systematically using analytic methods, the approach here is to replace the three spatial dimensions by  $4 - \varepsilon$  dimensions. For example, in massive  $\phi^4$  theory, the 4D action is transformed to

$$S = \int d^{4-\varepsilon}x [(\partial\phi)^2 + m^2\phi^2 + \mu^\varepsilon\lambda\phi^4], \quad (110)$$

where a mass scale  $\mu$  has been introduced to keep couplings dimensionless. This class of theories can be solved perturbatively in  $\varepsilon$  by making use of renormalization-group-improved perturbation theory. The procedure is to perform the calculation of a given quantity in this perturbative scheme to a given order in  $\varepsilon$  and then to take the 3D limit by taking  $\varepsilon \rightarrow 1$ . This approach provides an interesting way to study phase transitions. However, its application to the electroweak theory is complicated, and the authors estimate that the results may be accurate at the 30% level for Higgs masses below around 150 GeV and less reliable for higher values. Nevertheless, the prediction of  $\varepsilon$ -expansion techniques is that the phase transitions become weaker but remain first order as  $m_H$  increases, even to values higher than 80 GeV.

In another approach (Buchmuller and Philipsen, 1995) the one-loop Schwinger-Dyson equations are studied. This method predicts that the first-order transitions become a smooth crossover for Higgs masses above 100 GeV. Although this result is in reasonable agreement with that obtained from numerical approaches, it does rely on perturbation theory, which, as I have emphasized, should not really be trusted for  $m_H \geq 80$  GeV.

Finally, exact renormalization-group approaches have been applied to the electroweak theory by Wetterich and co-workers (Reuter and Wetterich, 1993; Bergerhoff and Wetterich, 1995) and by Buchmuller and Fodor (1994). Again, these analyses support the results of the numerical approaches I described earlier.

With the exception of the  $\varepsilon$  expansion, the qualitative conclusion of the above approaches is that there exists a critical value of the Higgs mass, below which the EWPT is first order, and above which the transition is continuous. Since, as I have mentioned, results from LEP (de Jong, 1998) now give  $m_H > 89.3$  GeV, it seems that the standard electroweak theory undergoes a continuous transition at high temperatures.

#### E. Subcritical fluctuations

A central assumption of the analysis of the phase transitions that I have followed above is that the initial state is homogeneous. If this is the case then the existence of a cubic term in the finite-temperature effective potential clearly implies that the phase transition proceeds by the nucleation and propagation of critical bubbles of true vacuum in the homogeneous sea of false. This result is obtained within the vacuum decay formalism, which relies on the semiclassical expansion of the effective action. This approach is valid if the system begins homogeneously in the false vacuum state so that one may consistently sum over small-amplitude fluctuations about this state, as I described earlier. However, this picture may not be valid if the initial state of the system is highly inhomogeneous.

The effects of initial-state inhomogeneities have been investigated using a variety of different approaches by Gleiser and collaborators (Gleiser, 1993, 1994, 1995;

Gleiser and Ramos, 1994; Borrill and Gleiser, 1995, 1997; Gleiser and Heckler, 1996; Gleiser *et al.*, 1997). At very high temperatures, the order parameter is certainly well localized about the symmetric vacuum state. However, as we approach the critical temperature we must be sure that thermal fluctuations do not lead to significant inhomogeneities such that the saddle-point approximation breaks down. In particular, if there is a sufficient probability for the order parameter to reach the true vacuum by purely classical thermal processes (sub-critical fluctuations) near the critical point, then we may not apply the nucleation calculation and might expect that the transition proceeds similarly to spinodal decomposition despite the first-order nature of the effective potential.

Although quantifying this statement is highly non-trivial, recent progress has been made numerically by Borrill and Gleiser (1995). These authors use a Langevin approach to track the dynamics of the order parameter in a spontaneously broken theory of a real scalar field in 3+1 dimensions. They use a potential that is identical in form to the electroweak potential and their initial conditions are that the scalar field is localized around the false vacuum. The potential is fixed at the critical temperature, the field evolves in time, and then the final spatial distribution of the scalar is determined for a range of values of the mass. This distribution is measured by defining  $f_0(t)$  to be the fraction of the total volume in the false vacuum at time  $t$  and evaluating the ensemble average of this quantity for different values of the scalar field mass.

With this method the authors are making all the assumptions usually made in studies of the transition (validity of the effective potential, homogeneity of the initial state). However, since the dynamics are governed by a Langevin equation (with rescaled order parameter  $\varphi$ ),

$$\ddot{\varphi} - \nabla^2 \varphi + \eta \dot{\varphi} + \frac{\partial V}{\partial \varphi} = \xi(\mathbf{x}, t), \quad (111)$$

the effects of noise  $\xi$  and viscosity  $\eta$  are taken into account and are related by the fluctuation-dissipation equation

$$\langle \xi(\mathbf{x}, t) \xi(\mathbf{x}', t') \rangle = 2\eta T \delta(t-t') \delta^3(\mathbf{x}-\mathbf{x}'). \quad (112)$$

The results of this simulation are that, for all scalar masses in the experimentally allowed range of Higgs mass, the authors expect phase mixing to occur at the critical temperature (see Fig. 7). In other words, by the time of the transition, half the universe is classically in the false vacuum and half is classically in the true.

If this is indeed the case in the electroweak theory then we might expect the standard bubble nucleation picture not to hold, irrespective of the validity of the finite-temperature effective potential. In fact, the dynamics of the phase transition may closely resemble that of spinodal decomposition. It is important to stress, however, that the electroweak theory itself is not being simulated in these approaches, and when the transition

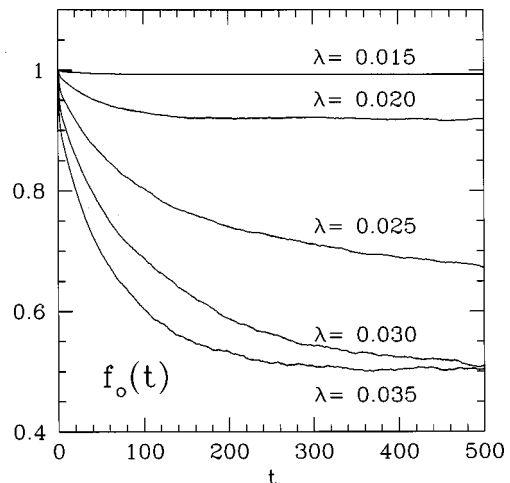


FIG. 7. Evolution of the ensemble-averaged fraction  $f_0(t)$  for several values of the scalar field self-coupling  $\lambda$ . Thanks to Marcelo Gleiser for permission to use this figure, which first appeared in Borrill and Gleiser, 1995.

from a real scalar field to the modulus of a complex field is made, and gauge fields are included, the dynamics may be quite different.

#### F. Erasure of the baryon asymmetry: Washout

There are essentially two criteria related to the strength of the phase transition that are important for electroweak baryogenesis. First, as I have commented, if the phase transition is second order or a crossover, then the only departure from equilibrium arises from the expansion of the universe. At the electroweak scale, this is a negligible effect and so no baryon production results. In such a situation, electroweak baryogenesis may only proceed if other physics is responsible for displacing the system from equilibrium. Such a situation is realized if TeV-scale topological defects are present at the phase transition, as I shall discuss in detail in Sec. VIII. If the phase transition is weakly first order, then the dynamics can be very complicated, as I have indicated above.

However, if the phase transition is sufficiently strongly first order, then we may be confident that widely separated critical bubbles nucleate and propagate as I have described. In this case, electroweak baryogenesis may proceed, but there is a further criterion to be satisfied. Consider a point in space as a bubble wall passes by. Initially, the point is in the false vacuum and sphaleron processes are copious, with rate per unit volume given by Eq. (52). As the wall passes the point, the Higgs fields evolve rapidly and the Higgs VEV changes from  $\langle \phi \rangle = 0$  in the unbroken phase to

$$\langle \phi \rangle = v(T_c) \quad (113)$$

in the broken phase. Here,  $v(T)$  is the value of the order parameter at the symmetry-breaking global minimum of the finite-temperature effective potential. That is,  $v(T_c)$  minimizes the free-energy functional at  $T = T_c$ . Now,  $CP$  violation and the departure from equi-

librium, and hence biased baryon production, occur while the Higgs field is changing. Afterwards, the point is in the true vacuum, baryogenesis has ended, and the rate per unit volume of baryon-number-violating events is given by Eq. (48) with  $\langle\phi\rangle$  given by Eq. (113). Since baryogenesis is now over, it is imperative that baryon number violation be negligible at this temperature in the broken phase, otherwise any baryonic excess generated will be equilibrated to zero. Such an effect is known as *washout* of the asymmetry. The condition that the Boltzmann suppression factor in Eq. (48) be sufficiently large to avoid washout may be roughly stated as (Shaposhnikov, 1986a, 1986b)

$$\frac{v(T_c)}{T_c} \geq 1. \quad (114)$$

This is the traditionally used criterion that the baryon asymmetry survive after the wall has passed, and in this article, this is the criterion I shall use. However, it is worth pointing out that there are a number of nontrivial steps that lead to this simple criterion. The actual bound is one on the energy of the sphaleron configuration at the bubble nucleation temperature. In order to arrive at the criterion (114), one must translate this into a bound at  $T_c$  and then write the sphaleron energy in terms of the VEV  $v(T_c)$ . Finally, as I commented earlier, if the evolution of the scale factor of the universe is nonstandard, then this bound can change. For an analysis of these issues see Joyce and Prokopec (1998). It is necessary that this criterion, or one similar to it, be satisfied for any electroweak baryogenesis scenario to be successful.

## G. Summary

I have given arguments that the minimal standard model has neither enough  $CP$  violation nor a sufficiently strong phase transition to allow electroweak baryogenesis to take place. These issues have been investigated in depth (Farrar and Shaposhnikov, 1994a, 1994b) and a detailed analysis was presented by Farrar and Shaposhnikov (1995a, 1995b), in which the authors found a potential enhancement of the Kobayashi-Maskawa  $CP$  violation by performing a detailed calculation of the effects of particle reflection and transport. However, it was later argued (Huet and Sather, 1995; Gavela *et al.*, 1995a, 1995b) that this potential effect would be erased by quantum decoherence. Separate attempts to enhance the  $CP$  violation present in the standard model through dynamical effects (Nasser and Turok, 1994) have thus far also been unsuccessful. It is therefore necessary to turn to extensions of the standard model. I shall do this in some detail for the MSSM in Sec. VII but in the next two sections I shall focus on the dynamics of electroweak baryogenesis.

## V. LOCAL ELECTROWEAK BARYOGENESIS

In the previous few sections, I have laid out the necessary criteria for a particle physics model to produce a

baryonic excess and have demonstrated that these criteria are satisfied by the standard model of particle physics and its modest extensions. In the present section I shall describe how these separate ingredients can come together dynamically to produce baryon asymmetry as the universe evolves through the electroweak phase transition.

Historically, the ways in which baryons may be produced as a bubble wall, or phase boundary, sweeps through space have been separated into two categories:

- (1) *Local baryogenesis*: baryons are produced when the baryon-number-violating processes and  $CP$ -violating processes occur together near the bubble walls.
- (2) *Nonlocal baryogenesis*: particles undergo  $CP$ -violating interactions with the bubble wall and carry an asymmetry in a quantum number other than baryon number into the unbroken phase region away from the wall. Baryons are then produced as baryon-number-violating processes convert the existing asymmetry into a baryon number asymmetry.

In general, both local and nonlocal baryogenesis will occur and the baryon asymmetry of the universe (BAU) will be the sum of that generated by the two processes. In this section I shall discuss local baryogenesis; nonlocal baryogenesis will be treated in the next section.

Models of local baryogenesis are some of the earliest viable models in the field. In this section I want to describe in detail the two major semianalytical approaches to calculating the BAU produced by this mechanism. It turns out that both approaches fall short of providing a reliable quantitative measure of the BAU. However, these attempts provide interesting ways of viewing the microphysics behind local baryogenesis. This is especially important since at present it appears that future numerical simulations may be our best hope of obtaining a reasonable estimate of this quantity. The most recent analysis of methods of treating local baryogenesis is due to Lue *et al.* (1997) and it is that analysis that I shall describe here.

To be specific, I shall assume  $CP$  violation via the operator  $\mathcal{O}$ , and assume a strongly first-order phase transition and thin bubble walls. I shall first look at an approach (Ambjørn *et al.*, 1989; Turok and Zadrozny, 1990, 1991; McLerran *et al.*, 1991) that attempts to estimate the baryon asymmetry by considering the relaxation of topologically nontrivial field configurations produced during the phase transition. Second, I shall turn to a method introduced by Dine *et al.* (1991; see also Dine, Leigh, *et al.*, 1992). In this treatment, one considers configurations that happen to be near the crest of the ridge between vacua as the wall arrives, and one estimates the extent to which their velocity in configuration space is modified by the operator  $\mathcal{O}$ , defined in Eq. (68), during the passage of the wall.

The mechanisms I consider in this section involve the dynamical evolution of configurations such as Eq. (31) which are released from rest and end up as outgoing

radiation. This complicates somewhat the simple situation described in Sec. II, since in this case it is dangerous to use the anomaly Eq. (33) because  $\int d^4x \text{Tr}(W\tilde{W})$  is not well defined as an integral and any answer can be obtained for the change in fermion number (Farhi *et al.*, 1995). Nonetheless, careful calculation shows that the results of Sec. II still apply. If the configuration (31) is released and falls apart without ever going through a zero of the Higgs field, then the analysis of Farhi *et al.* (1996) is directly applicable and one net antifermion is produced just as in the second interpolation. If the Higgs field unwinds by going through a zero, then one can use arguments presented in Farhi *et al.* (1995) to demonstrate that the presence of outgoing radiation in the final configuration does not affect the result above, namely, that there is no fermion number violation.

### A. Local baryogenesis through unwinding

Consider using the method of Turok and Zdrozny to estimate the baryon asymmetry produced by local baryogenesis in a scenario in which the electroweak phase transition is strongly first order and the bubble walls are thin (Lin *et al.*, 1997). In their original work, Turok and Zdrozny studied the classical dynamics of topologically nontrivial gauge and Higgs field configurations in the presence of  $CP$  violation. The first step is to consider spherically symmetric nonvacuum configurations of the form (31) with Higgs winding  $N_H = \pm 1$  and discuss their dynamics when they are released from rest and evolve according to the equations of motion. Solutions to the equations of motion typically approach a vacuum configuration uniformly throughout space at late times, and these solutions are no exception. There are, however, two qualitatively different possible outcomes of the evolution, as I mentioned in Sec. II. Without  $CP$  violation, for every  $N_H = +1$  configuration that relaxes in a baryon-producing fashion there is an  $N_H = -1$  configuration that produces antibaryons. With the inclusion of the  $CP$ -violating operator  $\mathcal{O}$ , the hope of this approach is that there will be some configurations that produce baryons whose  $CP$  conjugate configurations relax to the  $N_H = 0$  vacuum without violating baryon number.

The scenario described above is set in the following dynamical context. Imagine that the (thin) bubble wall has just passed, leaving in its wake a particular configuration, but that this configuration has not yet had time to relax to equilibrium. The goal is a qualitative understanding of the dynamics of this relaxation. The first-order electroweak phase transition can be characterized by the change in the gauge-invariant quantity  $\langle \sigma^2 \rangle$ , defined in Eq. (15). When  $\langle \sigma^2 \rangle$  is renormalized such that it is equal to  $v^2$  at zero temperature, then for a strongly first-order phase transition it is close to  $v^2$  just below  $T_c$  in the low-temperature phase, and is much smaller just above  $T_c$  in the high-temperature phase. This is a slight motivation for considering initial configurations in which  $\sigma = v$  throughout space, even though in reality this is not a good description of the nonequilibrium configurations

left in the wake of the wall and is not maintained during the subsequent evolution. Similarly, there is no justification for choosing a spherically symmetric configuration, or one with  $A_\mu = 0$ , or one that is initially at rest.

Note that the analysis of Lue *et al.* (1997) that I follow here treats the nonequilibrium conditions after the passage of a thin wall. The Lagrangian is simply Eq. (12) plus Eq. (68). This means that energy is conserved (to better than half a percent in numerical simulations) during the evolution of the gauge and Higgs fields.

The strategy now is to solve the equations of motion obtained from the action (12) augmented by the addition of the operator (68). This is performed in the spherical ansatz (Witten, 1977; Ratra and Yaffe, 1988) in which all gauge-invariant quantities are functions only of  $r$  and  $t$ , and the equations are solved numerically. First consider initial conditions of the form (31) with  $\eta(r)$  of (32) given by

$$\eta(r) = -\pi \left[ 1 - \tanh\left(\frac{r}{R}\right) \right], \quad (115)$$

where  $R$  is a constant parametrizing the size of the configuration. This configuration satisfies the boundary condition (17) and has Higgs winding number  $N_H = +1$ . It is used as the initial condition for the equations of motion, setting all time derivatives to zero at  $t=0$ . Initially,  $b$  is set to zero in Eq. (68) and thus there is no  $CP$  violation. In agreement with Turok and Zdrozny, this analysis finds that there is a critical value  $R_c^+$  of  $R$ , defined as follows. For all  $R < R_c^+$  the configuration evolves toward a vacuum configuration with  $N_H = 0$ . No fermions are produced in this background. For all  $R > R_c^+$ , the configuration evolves toward a vacuum configuration with  $N_H = +1$ , and fermions are produced. The values of  $R_c^+$  obtained in simulations with several different values of  $m_H/m_W$  are in quantitative agreement with those obtained by Turok and Zdrozny (1991). Repeating this exercise beginning with  $\eta(r)$  given by  $-1$  times that in Eq. (115), that is, beginning with the  $CP$  conjugate configuration having  $N_H = -1$ , results in an analogously defined  $R_c^-$ . Since  $b = 0$ , as expected  $R_c^- = R_c^+$ . The entire procedure is now repeated with  $b \neq 0$ , that is, with  $CP$  violation present, and yields  $R_c^- \neq R_c^+$ . Unfortunately, this is not the end of the story.

Consider slightly more general initial conditions, namely, exactly as above except that the time derivative of  $\sigma$  is nonzero and given by

$$\dot{\sigma}(r) = \gamma v^2 [1 - \tanh(r/R)], \quad (116)$$

with  $\gamma$  some constant. Performing the simulations again reveals that, for some values of  $\gamma$ ,  $R_c^- < R_c^+$  whereas for other values of  $\gamma$ ,  $R_c^- > R_c^+$ . This dooms an analysis in terms of the single parameter  $R$ . Clearly, a more general framework is needed.

Consider a family of initial configurations with  $N_H = +1$ , much more general than considered to this point, with a set of parameters  $\beta_i$ . It would be more general still to go beyond the spherical ansatz and eventually to work towards an analysis involving an infinite set of  $\beta$ 's,

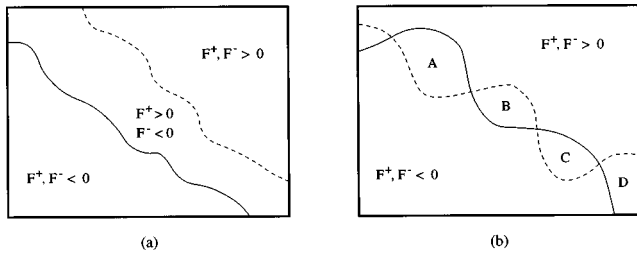


FIG. 8. Sketch of two qualitatively different possible behaviors of the critical surfaces  $F^+ = 0$  (represented by a solid line) and  $F^- = 0$  (represented by a dashed line) in the space  $(\beta_1, \beta_2, \dots)$  describing initial configurations.

but it seems reasonable to start with some finite set  $\beta_i$ . For any fixed  $b$ , define a function  $F^+(\beta_1, \beta_2, \dots)$  that has the following properties:

- (1)  $F^+(\beta_1, \beta_2, \dots) > 0$  for all points in  $\beta$  space that describe configurations which evolve toward the  $N_H = +1$  vacuum, thereby producing fermions;
- (2)  $F^+(\beta_1, \beta_2, \dots) < 0$  for all points in  $\beta$  space describing configurations that evolve towards the  $N_H = 0$  vacuum.

Considering only initial configurations described by Eqs. (31), (32), and (115), which are parametrized by the single parameter  $R$ , implies  $F^+(R) = R - R_c^+$ . Completely analogously, define a function  $F^-(\beta_1, \beta_2, \dots)$  such that the hypersurface  $F^-(\beta_1, \beta_2, \dots) = 0$  divides the  $\beta$  space of  $N_H = -1$  configurations into those which evolve towards the  $N_H = -1$  and  $N_H = 0$  vacua.

In the absence of  $CP$  violation,  $F^+(\beta_1, \beta_2, \dots) = 0$  and  $F^-(\beta_1, \beta_2, \dots) = 0$  define the same hypersurface. In this case, imagine allowing a  $CP$ -symmetric ensemble of configurations with  $N_H = +1$  and  $N_H = -1$  to evolve. (In this context,  $CP$ -symmetric means that the probability for finding a particular  $N_H = +1$  configuration in the ensemble is equal to that for finding its  $CP$  conjugate  $N_H = -1$  configuration.) Since the configurations that anomalously produce fermions are exactly balanced by those that anomalously produce antifermions, the net fermion number produced after relaxation would be zero. The aim is to investigate the behavior in the presence of the  $CP$ -violating term  $\mathcal{O}$  of Eq. (68). The hope is that  $\mathcal{O}$  will affect the dynamics of  $N_H = +1$  configurations and  $N_H = -1$  configurations in qualitatively different ways and that after relaxation to vacuum a net fermion number will result, even though the initial ensemble of configurations was  $CP$  symmetric. As described above, when  $b \neq 0$  the two hypersurfaces  $F^+ = 0$  and  $F^- = 0$  are indeed distinct. The configurations represented by points in  $\beta$  space between the two hypersurfaces yield a net baryon asymmetry. There are two qualitatively different possibilities, however, illustrated schematically in Fig. 8.

In Fig. 8(a), the hypersurfaces  $F^+ = 0$  and  $F^- = 0$  do not cross. The sign of  $b$  has been chosen such that a net baryon number is produced in the region between the two surfaces. In Fig. 8(b), the hypersurfaces cross and a net baryon number is produced in regions  $B$  and  $D$ , and

net antibaryons in regions  $A$  and  $C$ . If the hypersurfaces do not cross, as in Fig. 8(a), then a simple estimate of the fraction of configurations that yield a net baryon asymmetry is possible. This fraction would be proportional to the separation between the two hypersurfaces measured in any direction in  $\beta$  space with a component perpendicular to the hypersurfaces—for example, it would be proportional to  $(R_c^+ - R_c^-)$ —and it would be proportional to  $b/M^2$ , the coefficient of  $\mathcal{O}$ . Unfortunately, as described earlier, in the two-parameter space of  $(R, \gamma)$  the hypersurfaces  $F^+ = 0$  and  $F^- = 0$  do in fact cross. Thus, in the more general space  $(\beta_1, \beta_2, \dots)$ , the picture cannot look like that sketched in Fig. 8(a) and must look like that sketched in Fig. 8(b).

Hence, although the dynamics of the unwinding of topological configurations after the phase transition in the presence of the operator  $\mathcal{O}$  may lead to a baryon asymmetry, there is at present no way to make a simple analytical estimate of this asymmetry. This may well be a valid way of looking at the microphysics of electroweak baryogenesis, but it seems that large-scale (3+1)-dimensional numerical simulations of the kind recently pioneered by Moore and Turok (1997a) [but including  $CP$  violation via Eq. (68) and working in a setting in which the bubble walls are thin and rapidly moving] are required in order to estimate the contribution to the baryon asymmetry of the universe. I shall return shortly to a brief discussion of the large-scale numerical simulations that seem necessary. Before that, however, let us examine a different attempt at obtaining a semianalytical estimate of the magnitude of the effect.

## B. Kicking configurations across the barrier

The ideas presented in this subsection were originally considered by Shaposhnikov (1988) (see also Dine *et al.*, 1991). The basic idea is to consider the dynamics of configurations that are near the crest of the ridge between vacua as the bubble wall of the first-order electroweak phase transition arrives. The objective is to understand how the dynamics might be biased by the presence of  $CP$  violation in the theory and hence how a baryon asymmetry might result.

In a sense, this discussion is more general than that of the previous section, because it attempts to treat baryon-number-violating processes of a type more general than the unwinding of winding-number-one configurations. On the other hand, the treatment of these more general processes is, of necessity, greatly oversimplified. This discussion follows those of Dine *et al.* (1991) and Lue *et al.* (1997) to some extent, though they are not always parallel.

In the high-temperature phase, baryon-number-violating processes are not exponentially suppressed. The barrier-crossing configurations typically (Arnold and McLerran, 1987; Shaposhnikov, 1988; Arnold *et al.*, 1997; Huet and Son, 1997) have sizes given by the magnetic correlation length

$$\xi \sim (\alpha_W T)^{-1}. \quad (117)$$

Imagine dividing space up into cells of this size and looking at configurations cell by cell. The energy in gauge-field oscillations with wavelength  $\xi$  is of order  $T$ , but the total energy in a cell is much larger, as it is presumably of order  $T^4 \xi^3$ . Indeed, this energy is much larger than the sphaleron energy, which is  $E_{sph} \sim v/g$ . Most of the energy is in oscillations of the gauge and Higgs fields on length scales shorter than  $\xi$ . These configurations are crossing the barrier between vacua via regions of the barrier far above the lowest point on the barrier, that is far above the sphaleron, and they look nothing like the sphaleron. Recall that it is now believed that in each cell of volume  $\xi^3$ , the sphaleron barrier is crossed once per time  $\xi/\alpha_W$ , leading to  $\kappa \sim \alpha_W$ . Now consider what happens when the bubble wall hits the configurations just described.

Focus on the configuration in one cell. It traverses a path through configuration space, which may be parametrized by  $\tau$ . Dine *et al.* (1991) consider the special case in which this path is the path in configuration space that an instanton follows as a function of Euclidean time  $\tau$ , but this is not essential, and it is clear that they were thinking of more general circumstances also. The configurations discussed in Sec. V.A can be seen as special cases of those described here. The energy of the configuration has a maximum at some  $\tau$  (at which the configuration crosses the barrier) which is defined to be  $\tau=0$ . Now write down a Lagrangian that is intended to describe the dynamics of  $\tau$  as a function of time for  $\tau$  near  $\tau=0$ :

$$\mathcal{L}(\tau, \dot{\tau}) = \frac{c_1}{2\xi} \dot{\tau}^2 + \frac{c_2}{2\xi^3} \tau^2 + \frac{c_3}{\xi} \frac{b}{M^2} \sigma^2 \dot{\tau}. \quad (118)$$

In this expression,  $c_1$ ,  $c_2$ , and  $c_3$  are dimensionless constants, different for each of the infinitely many possible barrier-crossing trajectories. The factors of  $\xi$  have been put in by dimensional analysis, treating  $\tau$  as a quantity of dimension  $-1$ . (However, rescaling  $\tau$  by a dimensionful constant does not change the final result.) This Lagrangian should be seen as the first few terms in an expansion in powers of  $\tau$  and  $\dot{\tau}$ . Because by assumption  $\tau=0$  is a maximum of the energy as a function of  $\tau$ , no odd powers of  $\tau$  can appear. In the absence of  $CP$  violation, there can be no odd powers of  $\dot{\tau}$ , since they make the dynamics for crossing the barrier from left to right different than from right to left. The  $\text{Tr } F\tilde{F}$  in the operator  $\mathcal{O}$  includes a term that is proportional to the time derivative of the Chern-Simons number, and this means that  $\mathcal{O}$  must contribute a term in  $\mathcal{L}$  which is linear in  $\dot{\tau}$ . It is obviously quite an oversimplification to treat barrier crossing as a problem with one degree of freedom. As described earlier, the complete dynamics can be very complicated, even for relatively simple initial conditions. However, let us forge ahead with Eq. (118).

The momentum conjugate to  $\tau$  is given by

$$p = \frac{c_1}{\xi} \dot{\tau} + \frac{c_3}{\xi} \frac{b}{M^2} \sigma^2 \quad (119)$$

and the Hamiltonian density is therefore

$$\mathcal{H} = \frac{\xi}{2c_1} \left( p - \frac{c_3}{\xi} \frac{b}{M^2} \sigma^2 \right)^2 - \frac{c_2}{2\xi^3} \tau^2. \quad (120)$$

Before considering the thin-wall case described in detail here, it is worth pausing to consider the thick-wall limit in which  $\langle \sigma^2 \rangle$  is changing slowly and other quantities evolve adiabatically in this slowly changing background. A reasonable assumption is that the variables  $(\tau, p)$  are Boltzmann distributed with respect to the Hamiltonian (120) at each instant, treating  $\sigma^2$  as approximately constant. This implies that the distribution of  $p$  is centered at

$$p_0 = \frac{c_3}{\xi} \frac{b}{M^2} \sigma^2. \quad (121)$$

However, from Eq. (119), this means that the velocity  $\dot{\tau}$  is Boltzmann distributed with center  $\dot{\tau}=0$ . Thus the presence of the  $CP$ -violating operator (68) does not bias the velocity of trajectories in configuration space in the thick-wall limit. This conclusion disagrees with that of Dine *et al.* (1991) and Dine, Leigh, *et al.* (1992). There is nevertheless an effect. Integrating the third term in Eq. (118) by parts, one obtains a term linear in  $\tau$  proportional to the time derivative of  $\sigma^2$ . This changes the shape of the potential-energy surface in configuration space during the passage of the wall and yields an asymmetry. In this limit, in which the wall is thick and departure from equilibrium is small, the problem is much more easily treated in the language of spontaneous baryogenesis, as I shall discuss later—the operator  $\mathcal{O}$  acts like a chemical potential for baryon number.

Now let us return to the thin-wall case. Immediately after the wall strikes, the fields are not yet in equilibrium. The idea here is to use an impulse approximation to estimate the kick that  $\dot{\tau}$  receives as the wall passes, and from this to estimate the baryon asymmetry that results. The equation of motion for  $\tau$  obtained from Eq. (118) is

$$\ddot{\tau} = \frac{c_2}{c_1 \xi^2} \tau - \frac{c_3}{c_1} \frac{b}{M^2} \frac{d}{dt} \sigma^2. \quad (122)$$

During the passage of a thin wall, the first term on the right-hand side can be neglected relative to the second. In the impulse approximation the passage of the wall kicks  $\dot{\tau}$  by an amount

$$\Delta \dot{\tau} = - \frac{c_3}{c_1} \frac{b}{M^2} \Delta \sigma^2, \quad (123)$$

where  $\Delta \sigma^2$  is the amount by which  $\sigma^2$  changes at the phase transition. The kick  $\Delta \dot{\tau}$  has a definite sign. Thus, in the thin-wall limit, the distribution of the velocities in configuration space of barrier-crossing trajectories is biased and a baryon asymmetry results. How might one calculate the magnitude of this asymmetry?

If  $\Delta \dot{\tau}$  is large compared to  $\dot{\tau}_0$ , the velocity the configuration would have had as it crossed  $\tau=0$  in the absence of the action of the wall, then  $\Delta \dot{\tau}$  will kick the configuration over the barrier in the direction it favors and will produce, say, baryons rather than antibaryons.

If  $\Delta\tau$  is small compared to  $\tau_0$ , it will have no qualitative effect. The fraction of the distribution of configurations with  $\tau_0 < \Delta\tau$  is proportional to  $\Delta\tau$ . Note that in this calculation it was not necessary for  $\tau$  to be precisely at  $\tau = 0$  when the wall hits. It was only necessary for  $\tau$  to be close enough to  $\tau = 0$  that the Lagrangian (118) is a good approximation. It is difficult to quantify what fraction  $f$  of configurations satisfy this criterion of being “close enough to  $\tau = 0$ ,” although it is worth noting that  $f$  does not depend on the time it takes configurations to traverse the barrier. Nevertheless, the net number density of baryons produced may be estimated as

$$n_B \sim \Delta\tau f \xi^{-3}, \tag{124}$$

where the constants  $c_i, c_2, c_3$  have been absorbed into  $f$ .

At the time of the electroweak phase transition, the entropy density of the universe is  $s \sim 45T^3$ , and the baryon-to-entropy ratio is therefore<sup>2</sup>

$$\frac{n_B}{s} \sim f \frac{\alpha_W^3}{45} \frac{b}{M^2} \Delta\sigma^2. \tag{125}$$

The size of the effect clearly depends on  $\Delta\sigma^2$ . It was suggested by Dine *et al.* (1991) that  $\Delta\sigma^2$  corresponds to increasing  $\sigma^2$  up to that value at which baryon-number-violating processes become exponentially suppressed in thermal equilibrium. Whereas, in the thick-wall case, baryon-number-violating processes stop when  $\sigma^2$  reaches this value, this is not the case in the thin-wall scenario. In this setting, thermal equilibrium is not maintained even approximately, and it can be seen from the above discussion that what matters is the net change in  $\sigma^2$  as the wall passes. Once one picks an extension of the standard model which makes the transition strongly first order, one can compute  $\Delta\sigma^2$ . It is simplest just to take  $\Delta\sigma^2 = v^2/2$ , which is approximately what is obtained in the minimal standard model with a 35-GeV Higgs mass (Kajantie *et al.*, 1996b). Putting it all together yields

$$\frac{n_B}{s} \sim f(1 \times 10^{-9}) b \frac{(5 \text{ TeV})^2}{M^2}. \tag{126}$$

If, for example,  $b \sim \alpha_W$  and  $M \sim 1 \text{ TeV}$ , the bound (71) can be satisfied and Eq. (126) suggests that a cosmologically relevant baryon asymmetry may be generated. If  $CP$  violation is introduced via the operator  $\mathcal{O}$  with a coefficient  $b/M^2$  satisfying Eq. (71), and if the bubble walls are thin, then the contribution to the baryon asymmetry of the universe from local electroweak baryogenesis can be at an interesting level so long as the quantity  $f$  is not smaller than about a tenth.

There are many contributions to  $f$ , since the treatment leading to the estimate (126) is greatly oversimplified. First, there are the constants  $c_i$  ( $i = 1, 2, 3$ ), which of course differ for the different configurations in different

cells of volume  $\xi^3$  and must somehow be averaged over. Second, using the impulse approximation is not really justified. In reality, the wall does not have zero thickness. More important, even if the wall *is* thin, the time during which it can affect a configuration of size  $\xi$  is at least  $\xi$ . Third, the treatment in terms of the Lagrangian (118) only has a chance of capturing the physics near  $\tau = 0$ , and it is not at all clear what fraction of configurations satisfy this. Configurations that happen to be farther away from the crest of the ridge between vacua when the wall hits do receive a kick from the wall. However, even if this kick is large, it may not be in a suitable direction in configuration space to be effective. Configurations far from  $\tau = 0$  can contribute to  $n_B$ , but their contribution is hard to compute, because far from  $\tau = 0$ , there is no way to reduce the problem to one of one degree of freedom. Fourth, even near the crest of the ridge for a given trajectory the problem does not really reduce to one degree of freedom. For the configurations of interest,  $\sigma$  is a function of space and time, and the operator  $\mathcal{O}$  and the bubble wall conspire to affect its dynamics. In the method discussed here, the effect is described by treating  $\sigma$  as constant in space and time on either side of the wall and only changing at the wall. Fifth, we must face up to the specific difficulties discussed in the treatment of the Turok and Zdrozny mechanism. After the wall has passed, the fields are not yet in thermal equilibrium and their dynamics are complicated. This may in fact yield a further contribution to  $n_B$ . It may also, however, negate some of the contribution estimated in Eq. (126) because some configurations kicked across the barrier in one direction by the passage of the wall may at a later time wander back across the barrier whence they came. As mentioned earlier, an estimate of the magnitude of these sorts of effects is difficult even for a restricted class of configurations. To sum up,  $f$  is almost certainly less than 1. Hence it would be best to use Eq. (126) as an upper bound on  $n_B/s$ , rather than as an estimate.

Even if Eq. (126) is used only as an upper bound it is still interesting. Combined with the experimental bound (71) on the coefficient of  $\mathcal{O}$ , the result (126) shows that if the experimental sensitivity to the electric dipole moment of the electron or the neutron can be improved by about an order of magnitude, and if these experiments continue to yield results consistent with zero, then the baryon asymmetry of the universe produced by local electroweak baryogenesis is smaller than that observed, even if future numerical simulations were to demonstrate that  $f$  is as large as 1. Such a result would rule out the operator (68) as the source of  $CP$  violation for electroweak baryogenesis.

### C. Making progress

If the bubble walls are thick, conditions remain close to thermal equilibrium during the passage of the wall, and the nonequilibrium physics can be captured by assigning nonzero chemical potentials to various quantum numbers including baryon number. As I have men-

<sup>2</sup>This result agrees with that of Dine *et al.* (1991) although the discussion and results of Lue *et al.* (1997) are somewhat different.

tioned, analytic estimates for the baryon asymmetry of the universe produced in this setting exist in the literature (Cohen *et al.*, 1991b; Dine *et al.*, 1991; McLerran *et al.*, 1991; Cohen and Nelson, 1992; Dine, Leigh, *et al.*, 1992; Dine and Thomas, 1994), and the need for a numerical treatment is not pressing. If the bubble walls are thin, however, or if (as is no doubt the case) they are comparable in thickness to other length scales in the problem, the situation is unsettled and it seems that a large-scale numerical treatment is necessary.

Moore and Turok (1997a) have recently taken a big step in this direction. They have performed (3+1)-dimensional classical simulations in which a bubble wall moves through a box converting the high-temperature phase to the low-temperature phase. To date, they have focused more on computing quantities like the wall thickness, the wall velocity, the surface tension, and the drag on the wall and have only begun their treatment of local electroweak baryogenesis. To this point, they have introduced  $CP$  violation only by “mocking up” the effects of Eq. (68) by first computing the average wall profile  $\langle\sigma\rangle(z)$  for an ensemble of walls, where  $z$  is the direction perpendicular to the wall, and then doing a simulation in which one measures the distance of a given point to the nearest bubble wall and adds a chemical potential for Chern-Simons number at that point proportional to the spatial derivative of the average wall profile at that distance. This chemical potential is only nonzero on the wall, as it would be if it were proportional to  $d(\sigma^2)/dt$  for a moving wall. Nevertheless, by imposing the chemical potential as an external driving force instead of simply introducing Eq. (68) in the Lagrangian and letting the dynamics do their thing self-consistently, one risks missing a lot of the difficulties (and potential effects) discussed earlier. The simulations of Moore and Turok suggest that a large-scale numerical assault on the problem of local electroweak baryogenesis is now possible.

## VI. NONLOCAL BARYOGENESIS

If  $CP$  violation leads to an asymmetry in a quantum number other than baryon number and this asymmetry is subsequently transformed into a baryon excess by sphaleron effects in the symmetric phase, then this process is referred to as *nonlocal* baryogenesis. Nonlocal baryogenesis typically involves the interaction of the bubble wall with the various fermionic species in the unbroken phase. The main picture is that as a result of  $CP$  violation in the bubble wall, particles with opposite chirality interact differently with the wall, resulting in a net injected chiral flux. This flux thermalizes and diffuses into the unbroken phase, where it is converted to baryons. In this section, for definiteness when describing these effects, I shall assume that the  $CP$  violation arises because of a two-Higgs-doublet structure. There are typically two distinct calculational regimes that are appropriate for the treatment of nonlocal effects in elec-

troweak baryogenesis. Which regime is appropriate depends on the particular fermionic species under consideration.

- (1) The thin-wall regime (Cohen *et al.*, 1990, 1991a; Joyce *et al.*, 1994b): If the mean free path  $l$  of the fermions being considered is much greater than the thickness  $\delta$  of the wall, i.e., if

$$\frac{\delta}{l} < 1, \quad (127)$$

then we may neglect scattering effects and treat the fermions as free particles in their interactions with the wall.

- (2) The thick-wall regime: If the mean free path of the fermions is of the same order or less than the wall thickness, then scattering effects become important and the noninteracting picture is no longer applicable. Here there are two effects, classical force baryogenesis (Joyce *et al.*, 1994c) and nonlocal spontaneous baryogenesis (Joyce *et al.*, 1994c; Cohen *et al.*, 1994). In the former scenario, as a result of  $CP$  violation, an axial field emerges on the wall, leading to a classical force that perturbs particle densities, thus biasing baryon number. In the latter, hypercharge-violating processes in the presence of an axial field on the wall are responsible for perturbing particle densities in a  $CP$ -violating manner. When the effects of particle transport are taken into account, both cases give rise to nonlocal baryogenesis.

Both these mechanisms lead to an increase in the net effective volume contributing to baryogenesis over that of local baryogenesis since it is no longer necessary to rely on anomalous interactions taking place in the narrow region of the face of the wall where the changing Higgs fields provide  $CP$  violation.

The chiral asymmetry which is converted to an asymmetry in baryon number is carried by both quarks and leptons. However, the Yukawa couplings of the top quark and the  $\tau$  lepton are larger than those of the other quarks and leptons, respectively. Therefore it is reasonable to expect that the main contribution to the injected asymmetry comes from these particles and to neglect the effects of the other particles (for an alternative scenario see Davoudiasl *et al.*, 1998).

When considering nonlocal baryogenesis it is convenient to write the equation for the rate of production of baryons in the form (Joyce *et al.*, 1994b)

$$\frac{dn_B}{dt} = -\frac{n_f \Gamma(T)}{2T} \sum_i \mu_i, \quad (128)$$

where the rate per unit volume for electroweak sphaleron transitions is given by Eq. (52). Here,  $n_f$  is again the number of families and  $\mu_i$  is the chemical potential for left-handed particles of species  $i$ . An accurate evaluation of the chemical potentials that bias baryon number production is crucial in applying this equation.



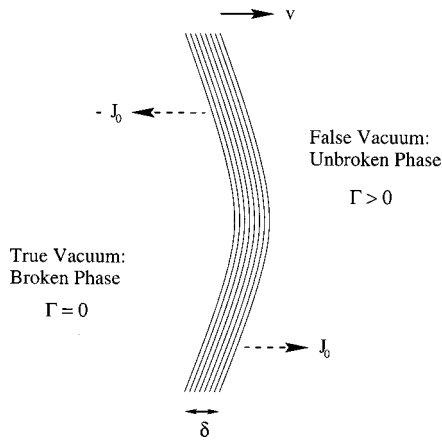


FIG. 9. Relevant quantities for nonlocal baryogenesis as a bubble wall propagates through the electroweak plasma, converting false vacuum to true.

### A. Thin bubble walls

Let us first consider the case in which the Higgs fields change only in a narrow region at the face of the bubble wall. We refer to this as the *thin-wall* case. In this regime, effects due to local baryogenesis are heavily suppressed because  $CP$ -violating processes take place only in a very small volume in which the rate for baryon-number-violating processes is nonzero. However, in contrast, we shall see that nonlocal baryogenesis produces an appreciable baryon asymmetry due to particle transport effects (Cohen *et al.*, 1990, 1991a; Joyce *et al.*, 1994b).

In the rest frame of the bubble wall, particles see a sharp potential barrier and undergo  $CP$ -violating interactions with the wall due to the gradient in the  $CP$ -odd Higgs phase. As a consequence of  $CP$  violation, there will be asymmetric reflection and transmission of particles, thus generating an injected current into the unbroken phase in front of the bubble wall. As a consequence of this injected current, asymmetries in certain quantum numbers will diffuse both behind and in front of the wall due to particle interactions and decays (Cohen *et al.*, 1990, 1991a; Joyce *et al.*, 1994b). In particular, the asymmetric reflection and transmission of left- and right-handed particles will lead to a net injected chiral flux from the wall (see Fig. 9). However, there is a qualitative difference between the diffusion occurring in the interior and exterior of the bubble.

Exterior to the bubble the electroweak symmetry is restored and weak sphaleron transitions are unsuppressed. This means that the chiral asymmetry carried into this region by transport of the injected particles may be converted to an asymmetry in baryon number by sphaleron effects. In contrast, particles injected into the phase of broken symmetry interior to the bubble may diffuse only by baryon-number-conserving decays, since the electroweak sphaleron rate is exponentially suppressed in this region. Hence I shall concentrate only on those particles injected into the unbroken phase.

The net baryon to entropy ratio that results via nonlocal baryogenesis in the case of thin walls has been cal-

culated in several different analyses (Cohen *et al.*, 1990, 1991a; Joyce *et al.*, 1994b). In the following I shall give a brief outline of the logic of the calculation, following Joyce *et al.* (1994b). The baryon density produced is given by Eq. (128) in terms of the chemical potentials  $\mu_i$  for left-handed particles. These chemical potentials are a consequence of the asymmetric reflection and transmission off the walls and the resulting chiral particle asymmetry. Baryon number violation is driven by the chemical potentials for left-handed leptons or quarks. To be concrete, I shall focus on leptons (Joyce *et al.*, 1994b); for quarks see, for example, Cohen *et al.* (1990). If there is local thermal equilibrium in front of the bubble walls—as I am assuming—then the chemical potentials  $\mu_i$  of particle species  $i$  are related to their number densities  $n_i$  by

$$n_i = \frac{T^2}{12} k_i \mu_i, \quad (129)$$

where  $k_i$  is a statistical factor which equals 1 for fermions and 2 for bosons. In deriving this expression, it is important (Joyce *et al.*, 1994a) to correctly impose the constraints on quantities that are conserved in the region in front of and on the wall.

Using the above considerations, the chemical potential  $\mu_L$  for left-handed leptons can be related to the left-handed lepton number densities  $L_L$ . These are in turn determined by particle transport. The source term in the diffusion equation is the flux  $J_0$  resulting from the asymmetric reflection and transmission of left- and right-handed leptons off the bubble wall.

For simplicity let us assume a planar wall. If  $|p_z|$  is the momentum of the lepton perpendicular to the wall (in the wall frame), the analytic approximation used by Joyce *et al.* (1994b) allows the asymmetric reflection coefficients for lepton scattering to be calculated in the range

$$m_l < |p_z| < m_H \sim \frac{1}{\delta}, \quad (130)$$

where  $m_l$  and  $m_H$  are the lepton and Higgs masses, respectively. This results in

$$\mathcal{R}_{L \rightarrow R} - \mathcal{R}_{R \rightarrow L} \simeq 2\Delta\theta_{CP} \frac{m_l^2}{m_H |p_z|}. \quad (131)$$

The corresponding flux of left-handed leptons is

$$J_0 \simeq \frac{vm_l^2 m_H \Delta\theta_{CP}}{4\pi^2}. \quad (132)$$

Note that in order for the momentum interval in Eq. (130) to be nonvanishing, the condition  $m_l \delta < 1$  needs to be satisfied.

The injected current from a bubble wall will lead to a “diffusion tail” of particles in front of the moving wall. In the approximation in which the persistence length of the injected current is much larger than the wall thickness we may to a good approximation model it as a delta-function source and search for a steady-state solution. In addition, assume that the decay time of leptons

is much longer than the time it takes for a wall to pass so that we may neglect decays. Then the diffusion equation for a single particle species becomes

$$D_L L_L'' + v L_L' = \xi_L J_0 \delta(z), \quad (133)$$

where  $D_L$  is the diffusion constant for leptons and a prime denotes the spatial derivative in the direction  $z$  perpendicular to the wall. This equation contains a parameter  $\xi^L$  that is called the *persistence length* of the current in front of the bubble wall. This describes, and contains all uncertainties about how the current thermalizes in the unbroken phase. Equation (133) can be immediately integrated once, with the integration constant specified by the boundary condition

$$\lim_{|z| \rightarrow \infty} L_L(z) = 0. \quad (134)$$

This leads easily to the solution

$$L_L(z) = \begin{cases} J_0 \frac{\xi_L}{D_L} e^{-\lambda_D z} & z > 0, \\ 0 & z < 0 \end{cases} \quad (135)$$

with the diffusion root

$$\lambda_D = \frac{v}{D_L}. \quad (136)$$

Note that in this approximation the injected current does not generate any perturbation behind the wall. This is true provided  $\xi^L \gg \delta$  is satisfied. If this inequality is not true, the problem becomes significantly more complex (Joyce *et al.*, 1994b).

In the massless approximation the chemical potential  $\mu_L$  can be related to  $L_L$  by

$$\mu_L = \frac{6}{T^2} L_L \quad (137)$$

(for details see Joyce *et al.*, 1994b). Inserting the sphaleron rate and the above results for the chemical potential  $\mu$  into Eq. (128), we find that the final baryon-to-entropy ratio becomes

$$\frac{n_b}{s} = \frac{1}{4\pi^2} \kappa \alpha_W^4 (g^*)^{-1} \Delta \theta_{CP} \left( \frac{m_l}{T} \right)^2 \frac{m_H}{\lambda_D} \frac{\xi^L}{D_L}. \quad (138)$$

The diffusion constant is proportional to  $\alpha_W^{-2}$  (see Joyce *et al.*, 1994b):

$$\frac{1}{D_L} \approx 8 \alpha_W^2 T. \quad (139)$$

Hence, provided that sphalerons do not equilibrate in the diffusion tail,

$$\frac{n_b}{s} \sim 0.2 \alpha_W^2 (g^*)^{-1} \kappa \Delta \theta_{CP} \frac{1}{v} \left( \frac{m_l}{T} \right)^2 \frac{m_H}{T} \frac{\xi^L}{D_L}. \quad (140)$$

Since we may estimate that

$$\frac{\xi^L}{D_L} \sim \frac{1}{T \delta}, \quad (141)$$

the baryon-to-entropy ratio obtained by nonlocal baryogenesis is proportional to  $\alpha_W^2$  and not  $\alpha_W^4$  as in the result for local baryogenesis.

Now consider the effects of top quarks scattering off the advancing wall (Cohen *et al.*, 1990, 1991a). Several effects tend to decrease the contribution of the top quarks relative to that of tau leptons. First, for typical wall thicknesses the thin-wall approximation does not hold for top quarks. This is because top quarks are much more strongly interacting than leptons and so have a much shorter mean free path. An important effect is that the diffusion tail is cut off in front of the wall by *strong sphalerons* (Mohapatra and Zhang, 1992; Giudice and Shaposhnikov, 1994). There is an anomaly in the quark axial-vector current in QCD. This leads to chirality-nonconserving processes at high temperatures. These processes are relevant for nonlocal baryogenesis since it is the chirality of the injected current that is important in that scenario. In an analogous expression to that for weak sphalerons, we may write the rate per unit volume of chirality-violating processes due to strong sphalerons in the unbroken phase as

$$\Gamma_s = \kappa_s (\alpha_s T)^4, \quad (142)$$

where  $\kappa_s$  is a dimensionless constant (Moore, 1997). Note that the uncertainties in  $\kappa_s$  are precisely the same as those in  $\kappa$  defined in Eq. (52). As such,  $\kappa_s$  could easily be proportional to  $\alpha_s$ , in analogy with Eq. (55), perhaps with a logarithmic correction. These chirality-changing processes damp the effect of the injected chiral flux and effectively cut off the diffusion tail in front of the advancing bubble wall. Second, the diffusion length for top quarks is intrinsically smaller than that for tau leptons, thus reducing the volume in which baryogenesis takes place. Although there are also enhancement factors, e.g., the ratio of the squares of the masses  $m_t^2/m_\tau^2$ , it seems that leptons provide the dominant contribution to nonlocal baryogenesis.

## B. Thick bubble walls

If the mean free path of the fermions being considered is smaller than the width of the wall, we refer to the thick-wall, or adiabatic, limit, and the analysis is more complicated. In the case of thin bubble walls, the plasma within the walls undergoes a sharp departure from equilibrium. When the walls are thick, however, most interactions within the wall will be almost in thermal equilibrium. The equilibrium is not exact because some interactions, in particular baryon number violation, take place on a time scale much slower than the rate of passage of the bubble wall. These slowly varying quantities are best treated by the method of chemical potentials. For nonlocal baryogenesis, it is still useful to follow the diffusion equation approach. However, whereas in the thin-wall case it was sufficient to model the source as a  $\delta$  function, here we must consider sources that extend over the wall. There are a number of possible sources.

To be definite, consider the example in which  $CP$  violation is due to a  $CP$ -odd phase  $\theta$  in the two-Higgs-doublet model.

## 1. Spontaneous baryogenesis

Let us begin by describing how the dynamics of  $\theta$  might bias baryon number production. In order to see how  $\theta$  couples to the fermionic sector of the theory (to produce baryons) we may remove the  $\theta$  dependence of the Yukawa couplings arising from the Higgs terms. We do this by performing an anomaly-free hypercharge rotation on the fermions (Nelson *et al.*, 1992), inducing a term in the Lagrangian density of the form

$$\mathcal{L}_{CP^\infty} \partial_\mu \theta \left[ \frac{1}{6} \bar{U}_L \gamma^\mu U_L + \frac{1}{6} \bar{D}_L \gamma^\mu D_L + \frac{2}{3} \bar{U}_R \gamma^\mu U_R - \frac{1}{3} \bar{D}_R \gamma^\mu D_R - \frac{1}{2} \bar{l}_L \gamma^\mu l_L - \bar{E}_R \gamma^\mu E_R \right]. \quad (143)$$

Here  $U_R$  and  $D_R$  are the right-handed up and down quarks, respectively,  $l_L$  are the left-handed leptons, and  $E_R$  are the right-handed charged leptons. The quantity in the square brackets is proportional to the fermionic part of the hypercharge current, and therefore changes in  $\theta$  provide a preferential direction for the production of quarks and leptons, in essence a chemical potential for baryon number  $\mu_B$  similar to that described in Eq. (65).

Of course, strictly speaking, this is not a chemical potential, since it arises dynamically rather than by imposing a constraint. For this reason, the quantity  $\mu_B$  is sometimes referred to as a ‘‘charge potential.’’ The effect of the charge potential is to split the otherwise degenerate energy levels of the baryons with respect to the antibaryons. This results in a free-energy difference which we may feed into Eq. (72) to obtain

$$\frac{dn_B}{dt} = -9 \frac{\Gamma(T)}{T} \mu_B. \quad (144)$$

The relevant baryon number produced by spontaneous baryogenesis is then calculated by integrating this equation.

Initially, spontaneous baryogenesis was considered as an example of local baryogenesis. However, it has become clear that diffusion effects can lead to an appreciable enhancement of the baryon asymmetry produced by this mechanism (Cohen *et al.*, 1994), and this effect has been investigated by a number of authors (Joyce *et al.*, 1994; Comelli *et al.*, 1995). In addition, it has been shown that the final result for the baryon-to-entropy ratio must be suppressed by a factor of  $m^2/T^2$  where  $m$  is the relevant fermion mass (Comelli *et al.*, 1995). As we saw in the case of thin walls, the major part of the work is to use the diffusion equation to calculate the number densities of particle species, and then to relate these densities to the chemical potentials. In the case of thick walls the situation is complicated, since one must consider extended sources for the appropriate diffusion equations. Nevertheless, the calculations can be performed and I will sketch one such approach as part of the next subsection.

## 2. Classical force baryogenesis

An alternative way of generating a source for the diffusion equation was suggested by Joyce *et al.* (1994c).

The essential idea is that there exists a purely classical chiral force that acts on particles when there is a  $CP$ -violating field on the bubble wall. I shall briefly sketch this idea here. The Lagrangian for a fermion  $\psi$  in the background of a bubble wall in the presence of the  $CP$ -odd phase  $\theta$  is

$$\tilde{\mathcal{L}} = i \bar{\psi} \gamma^\mu \left( \partial_\mu + \frac{i}{2} \frac{v_2^2}{v_1^2 + v_2^2} \gamma^5 \partial_\mu \theta \right) \psi - m \bar{\psi} \psi, \quad (145)$$

where  $v_1$  and  $v_2$  are the VEVs of the Higgs fields and  $m$  is the fermion mass. As usual we assume a planar wall propagating at speed  $v$  in the  $z$  direction. Further, if interactions of the Higgs and gauge fields with the wall have reached a stationary state, then these fields are functions only of  $z - vt$ . Then, in the wall rest frame plane-wave fermion solutions  $\psi \propto e^{-ip \cdot x}$  have the dispersion relation

$$E = \left[ p_\perp^2 + \left( \sqrt{p_z^2 + m^2} \pm \frac{1}{2} \frac{v_2^2}{v_1^2 + v_2^2} \partial_z \theta \right)^2 \right]^{1/2}, \quad (146)$$

where  $\pm$  corresponds to  $s_z = \pm 1/2$ , with  $s_z$  the  $z$  component of the spin.

In the WKB approximation, the Hamilton equations derived from this relation allow calculation of the acceleration of a WKB wave packet (i.e., the chiral force). This force acts as a potential well that induces an excess of chiral charge on the wall. In the diffusion tail in front of the wall, there exists a corresponding deficit of chiral charge. This acts as a chemical potential that is the source of baryon number.

This calculation is performed by solving the associated Boltzmann equation,

$$\frac{\partial f}{\partial t} + \dot{z} \frac{\partial f}{\partial z} + \dot{p}_z \frac{\partial f}{\partial p_z} = -C(f), \quad (147)$$

in which the collision integral  $C(f)$  accounts for the fact that in the thick-wall regime we may no longer use the noninteracting picture, and  $\dot{z}$  and  $\dot{p}_z$  are obtained from the Hamilton equations.

Joyce *et al.* (1994c) follow a fluid approach to solving this equation. By concentrating on particles with  $|p_z| \sim T \gg m$ , they are able to treat the particle and antiparticle excitations (146) as separate fluids in the WKB approximation. This permits an analytic solution to the diffusion equation, which yields a relation between the chiral force and the chemical potential on the wall for the fermion species considered as

$$\mu_\psi = \frac{2 \ln(2)}{3 \zeta(3)} \frac{v}{2} \frac{v_2^2}{v_1^2 + v_2^2} \partial_z \theta \left( \frac{m}{T} \right)^2, \quad (148)$$

where  $\zeta$  is again the Riemann function. Integrating this in front of the wall yields a chemical potential that may then be fed into Eq. (128) to yield an approximate answer for the baryon asymmetry of the universe.

## C. Summary and progress

The advantage of nonlocal baryogenesis over local baryogenesis is that the effective volume in which

baryon number generation can occur is enhanced in the former case. This is because the effects of transport in the plasma external to the expanding bubble allow baryon-number-violating transitions in the unbroken phase to transform a chiral asymmetry produced on the wall into a baryon asymmetry. This means that in the case of nonlocal baryogenesis we do not need to rely on baryon-number-violating processes occurring in the region where the Higgs fields are changing.

The diffusion equation approach to the problem of nonlocal baryogenesis has been very successful. In the thin-wall case, it is a valid approximation to assume that the source for the diffusion equation is essentially a  $\delta$  function on the wall. This is because one may ignore the effects of particle scattering in this picture. However, in the case of thick walls, significant particle scattering occurs, and as a result it is necessary to consider sources for the diffusion equations that extend over the wall.

New approaches to this scenario continue to appear as we try to understand the detailed predictions of specific extensions of the standard model to compare with upcoming accelerator tests (Cline, Joyce, and Kainulainen, 1998; Enqvist *et al.*, 1997; Riotto, 1998b, 1998c).

## VII. A REALISTIC MODEL OF ELECTROWEAK BARYOGENESIS: THE MSSM

As we have seen, if we are to produce enough baryons at the electroweak phase transition, we must go beyond the minimal Glashow-Salam-Weinberg theory. This is true both to obtain strong enough  $CP$  violation and to ensure a sufficient departure from thermal equilibrium. It has been common to invoke general two-Higgs-doublet models to satisfy these conditions (Bochkarev *et al.*, 1990; Arnold *et al.*, 1992; Parwani, 1992; Davies *et al.*, 1994; Losada, 1996b), and the behavior of the phase transition has most recently been investigated in that context by Cline and Lemieux (1997). Here I shall focus on a more restrictive model with good particle physics motivations—the minimal supersymmetric standard model (MSSM).

Supersymmetry requires that each fermion have a bosonic superpartner and vice versa. The essential feature of the MSSM, therefore, is that it contains twice the particle content of the MSM plus an extra Higgs doublet (and superpartner). Naturally, introducing a host of new free parameters into the theory allows us to relax the constraints that were derived for the MSM. However, many of the new parameters are constrained by supersymmetry and by existing accelerator measurements.

### A. Electroweak phase transition in the MSSM

A strongly first-order phase transition can result if the theory contains light scalars, which are strongly coupled to the Higgs field. The MSSM contains two Higgs doublets  $H_1$  and  $H_2$ , one linear combination of which is a  $CP$ -even Higgs boson  $h$ . The mass  $m_h$  of this particle is constrained to be less than of order 125 GeV (Carena *et al.*, 1996; Haber *et al.*, 1997). Further, the superpart-

ners of the top quark, the stops  $\tilde{t}$ , couple to the Higgs with strength of the order of the top-quark Yukawa coupling. These light scalars are precisely the particles necessary to enable the phase transition to be more strongly first order than in the MSM (Guidice, 1992; Brignole *et al.*, 1994; Delepine *et al.*, 1996). In fact, for their effects to be useful, it is necessary that the lightest stop be no heavier than the top quark itself (see, for example, Espinosa, 1996).

The phase transition in supersymmetric electroweak theories has been investigated very recently (Cline and Kainulainen, 1996, 1998; Farrar and Losada, 1996; Laine, 1996a; Losada, 1996a, 1996b; Bodeker *et al.*, 1997; Carena *et al.*, 1998). In addition to the parameters mentioned above, the other important quantity is

$$\tan\beta \equiv \frac{\langle H_2 \rangle}{\langle H_1 \rangle}, \quad (149)$$

the ratio of the vacuum expectation values of the Higgs fields. The methods available to estimate the region of parameter space in which EWBG is possible in the minimal standard model can also be applied here. However, since there is more structure in the MSSM, the calculation is more involved. Nevertheless, the calculation of the allowed region of the  $m_h$ - $m_{\tilde{t}}$  space has been performed using the finite-temperature effective potential computed up to all finite-temperature two-loop corrections (Carena *et al.*, 1998). The two-loop contributions are crucial to the accuracy of this calculation and yield the value

$$\tan\beta \approx 2 \quad (150)$$

and the constraints

$$75 \text{ GeV} \leq m_h \leq 105 \text{ GeV}, \quad (151)$$

$$100 \text{ GeV} \leq m_{\tilde{t}} \leq m_t. \quad (152)$$

These results have been confirmed by lattice numerical simulations by Laine and Rummukainen (1998). In the ranges (152) the phase transition in the MSSM is strong enough that washout of the baryon asymmetry produced is avoided.

In quoting these bounds, there is an issue I have neglected. It is possible to choose values of the parameters in the MSSM such that the absolute energy minimum in field space is a vacuum that breaks the color symmetry. Although this would normally be forbidden, it is possible that the vacuum in which we live is metastable with an extremely long lifetime, and hence such a minimum would be consistent with observations. If we allow such parameter values, then the above constraints may be relaxed a little. Here I have chosen to adopt the most conservative assumption, namely that the color-preserving vacuum state is the global minimum at all temperatures.

### B. Extra $CP$ violation in the MSSM

In the MSSM there are extra sources of  $CP$  violation beyond that contained in the Cabibbo-Kobayashi-

Maskawa matrix of the standard model. First, when supersymmetry breaking occurs, as we know it must, the interactions of the Higgs fields  $H_1$  and  $H_2$  with charginos and neutralinos at the one-loop level lead to a  $CP$ -violating contribution to the scalar potential of the form

$$V_{CP} = \lambda_7 (H_1 H_2)^2 + \lambda_8 |H_1|^2 H_1 H_2 + \lambda_9 |H_2|^2 H_1 H_2 + (\text{H.c.}) \quad (153)$$

The interaction of the Higgs fields with the stops is a potential further effect that contributes to the  $CP$ -violating part of the potential. However, in the range (152) of stop masses required to maintain a strong enough phase transition, this effect is suppressed. The nature of supersymmetry breaking can lead to the parameters' being complex,

$$\lambda_7 = |\lambda_7| e^{2i\alpha}, \quad (154)$$

$$\lambda_8 = |\lambda_8| e^{i\alpha}, \quad (155)$$

$$\lambda_9 = |\lambda_9| e^{i\alpha}. \quad (156)$$

The phase  $\alpha$  breaks  $CP$  and can bias baryon production in a way similar to that described for the two-Higgs model.

Second, in SUSY extensions of the minimal standard model, there exists a mass mixing matrix for the charginos. This mass matrix has a similar structure to the quark mixing matrix in the MSM. In particular, the chargino mixing matrix contains a phase  $\phi_B$  that parametrizes  $CP$  violation in this sector. Similarly, there exists a mixing matrix for the neutralinos, which also contains this same phase. Finally, there is a further mass mixing matrix for the top squarks. This contains a phase that is a linear combination of  $\phi_B$  and a second phase  $\phi_A$ . These extra phases can provide enough  $CP$  violation to be useful for baryogenesis (Huet and Nelson, 1996). However, in some regimes it is possible to constrain their sizes through dipole moment calculations such as I described earlier.

The electroweak phase transition and  $CP$  violation are not the only features of electroweak baryogenesis that change when one considers supersymmetric models. It is necessary to reanalyze the reflection of particles from bubble walls within the context of these theories, and to reexamine the ranges of validity of the various approximations used to obtain estimates for the baryon asymmetry of the universe. Serious attempts to do just this have only been made in the last two years (Davies *et al.*, 1996; Carena *et al.*, 1997; Worah, 1997; Cline, Joyce, and Kainulainen *et al.*, 1998; Riotto, 1998a, 1998b, 1998c).

The regions of parameter space I have mentioned above are still experimentally allowed. Thus the MSSM remains a viable candidate theory to explain the baryon asymmetry of the universe. It is quite possible that the full parameter range will be covered in the next decade. At that time we should know whether MSSM baryogenesis explains the BAU or whether we must turn to more complicated models of the electroweak scale.

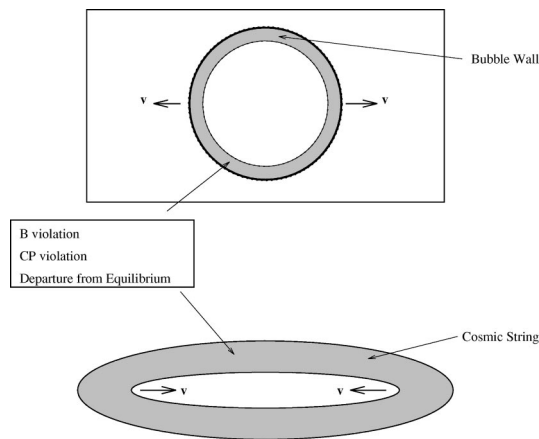


FIG. 10. Analogy between the bubble-wall-mediated baryogenesis scenario and the defect-mediated scenario.

## VIII. TOPOLOGICAL DEFECTS AND THE DEPARTURE FROM THERMAL EQUILIBRIUM

If the dynamics of the electroweak phase transition are such that the traditional scenarios for electroweak baryogenesis that I have described are inefficient, then it is interesting to explore alternative implementations of baryogenesis at the electroweak scale. In this section, I describe a particular realization of the third Sakharov condition suggested by myself and collaborators (Brandenberger *et al.*, 1994; Trodden *et al.*, 1995; Brandenberger *et al.*, 1996; Prokopec *et al.*, 1996). This implementation uses the out-of-equilibrium evolution of a network of topological defects to realize the third Sakharov criterion, instead of the evolution of bubble walls (see Fig. 10). Different scenarios making use of topological defects formed at that temperature have been suggested by other authors (Vachaspati and Field, 1994; Barriola, 1995).

Topological defects are regions of trapped energy density that can remain after a cosmological phase transition if the topology of the vacuum of the theory is nontrivial. Typically, cosmological phase transitions occur when a gauge symmetry of a particle physics theory is spontaneously broken. In that case, the cores of the topological defects formed are regions in which the symmetry of the unbroken theory is restored. There exist many excellent reviews of the physics of topological defects and I refer the reader to one of those, rather than provide a lengthy aside here. For the purposes of this section, it is sufficient to specialize to the case of cosmic strings; linelike, solitonic solutions to spontaneously broken field theories, for which there exist noncontractible loops in the vacuum manifold. If these objects arise from the breakdown of a gauge symmetry, then they can have interesting microphysics, as I shall briefly describe.

### A. Electroweak symmetry restoration and the baryogenesis volume

Let us review the scenario for electroweak baryogenesis proposed by Brandenberger *et al.* (1994). I shall begin by briefly explaining the physical principles behind

electroweak symmetry restoration around ordinary (nonsuperconducting) defects. This is an essential ingredient of defect-mediated electroweak baryogenesis.

Consider a cosmic string, formed at a scale  $\eta > \eta_{EW}$ , that couples to the Glashow-Salam-Weinberg model. Further, assume that the gauge fields corresponding to this higher symmetry scale acquire an extra mass at the electroweak scale. The simplest example is to introduce an extra gauged  $U(1)$  symmetry that breaks to the standard model,

$$SU(2)_L \times U(1)_Y \times U(1) \rightarrow SU(2)_L \times U(1)_Y. \quad (157)$$

This is the model originally considered by Perkins and Davis (1993). Let the string's scalar and gauge fields be  $S$  (which breaks the symmetry) and  $R_\mu$ , respectively. The coupling between the string and the electroweak sector is through  $R_\mu$  and the electroweak Higgs field  $\Phi$ , in the covariant derivative

$$D_\mu \Phi = \left( \partial_\mu - \frac{1}{2} ig \tau \cdot \mathbf{W}_\mu - \frac{1}{2} ig' B_\mu - \frac{1}{2} ig'' R_\mu \right) \Phi. \quad (158)$$

Since  $\eta > \eta_{EW}$ , we may consistently treat  $R_\mu$  as a background with Nielsen-Olesen characteristics. We now minimize the energy of this configuration. Writing the Higgs field in the unitary gauge

$$\Phi = (0, \varphi(r))^T, \quad (159)$$

we find that the minimal energy configuration is achieved for

$$\varphi(r) = \frac{\eta}{\sqrt{2}} \begin{cases} \left( \frac{r}{R_s} \right)^{a/2} & r < R_s \\ 1 & r > R_s, \end{cases} \quad (160)$$

where  $a$  is a constant. With this ansatz the scale of electroweak symmetry restoration  $R_s$  is determined to be

$$R_s \sim \lambda^{-1/4} \tilde{G}^{-1/2} \eta_{EW}^{-1}, \quad (161)$$

where  $\eta_{EW}$  is the electroweak scale and  $\tilde{G}^2 = g^2 + g'^2$ .

Thus the electroweak symmetry is restored out to the inverse electroweak scale around such a higher-scale ordinary defect. If certain consistency conditions are satisfied [e.g., sphalerons fit inside the defect so that their rate is not significantly suppressed (Perkins, 1995)], then within this region the rate of baryon number violation is still given by Eq. (52) after the electroweak phase transition.

Once again, I shall assume that  $CP$  violation is due to a  $CP$ -odd relative phase,  $\theta$ , between two electroweak Higgs doublets, and that this phase changes by  $\Delta\theta_{CP}$  during the transition from false to true vacuum, and by  $-\Delta\theta_{CP}$  in the reverse transition. Thus, if  $\Delta\theta_{CP} > 0$  for a given process, then baryon number is driven positive (an excess of baryons over antibaryons is generated) and vice versa.

As a defect moves, certain regions of the background space enter the core of the defect—i.e., make the transition from true to false vacuum—while others leave the core and make the transition from false to true vacuum. There are certain types of motion of defects and evolu-

tions of defect networks that can provide an asymmetry such that an overall baryon excess is created in the universe.

An important quantity that enters the calculation is the suppression factor

$$\Lambda \sim \left( \frac{V_{BG}}{V} \right), \quad (162)$$

where  $V_{BG}$  is the volume in which baryogenesis occurs and  $V$  is the total volume.  $\Lambda$  is the factor by which defect-mediated baryogenesis is weaker than baryogenesis with bubble walls. In the original work of Brandenberger *et al.* (1994), the processes responsible for the generation of the baryon asymmetry were purely local. For a collapsing topological defect, purely local baryogenesis restricts the baryogenesis volume to be the initial volume of the defect because the effects of  $\dot{\theta} > 0$  on one side of the string are canceled by the effects of  $\dot{\theta} < 0$  on the other. In fact, in order for  $\Lambda$  not to be a prohibitively small suppression, it is necessary that the scale at which the defects are formed be extremely close to the electroweak scale. For definiteness, we shall restrict ourselves to local baryogenesis here, but note that nonlocal effects lead to an appreciable enhancement and allow the defects to be formed at scales further above the electroweak scale.

## B. Local baryogenesis and diffusion in defects

In this section I shall obtain an estimate for the baryon asymmetry produced by a topological defect as a consequence of local mechanisms.

As a topological defect passes each point in space, a number density of antibaryons is produced by local baryogenesis at the leading face of the defect, and then an equal number density of baryons is produced as the trailing edge passes. Naively we would expect that these effects would cancel each other, so that any time-symmetric motion of the defect, such as translation, would yield no net baryon asymmetry. For this reason, the analysis of Brandenberger *et al.* (1994) was restricted to the time-asymmetric motion of cosmic-string loop collapse. In that case, the cancellation effects led to suppression of the strength of the mechanism by the factor

$$\Lambda = \lambda^3 \left( \frac{\eta_{EW}}{\eta} \right)^3. \quad (163)$$

However, this neglects an important effect and thus underestimates the strength of the mechanism. The antibaryons produced at the leading edge of the defect at a fixed point in space spend a time interval  $\tau$  inside the defect during which they may decay before the trailing edge passes by and produces baryons at the same point. The core passage time  $\tau$  is given by

$$\tau = \frac{L}{v_D}, \quad (164)$$

where  $L$  is the width of the defect and  $v_D$  is its velocity.

Thus, if  $n_b^0$  is the number density of baryons (or antibaryons) produced at either edge, we may estimate the net baryon asymmetry  $B$  produced after the defect has passed a given point once to be

$$B = n_b^0(1 - e^{-\bar{\Gamma}\tau}), \quad (165)$$

where  $\bar{\Gamma}$  is the rate at which antibaryons decay and may be related to the electroweak sphaleron rate by (Joyce *et al.*, 1994b)

$$\bar{\Gamma} = 6n_f \frac{\Gamma}{T^3} = 6n_f \kappa \alpha_W^4 T. \quad (166)$$

The resulting average baryon number density  $n_b$  can be estimated from Eq. (165), taking into account Eq. (166) and the volume suppression [see Eq. (162)]:

$$n_b \approx 3n_f \frac{\Gamma}{T} \mu \frac{\delta}{v_D} (1 - e^{-\bar{\Gamma}\tau}) \Lambda, \quad (167)$$

where  $\delta$  is the thickness of the defect wall. The derivatives of  $\theta_{CP}$  and  $\delta/v_D$  combine to give  $\Delta\theta_{CP}$ , and hence the resulting net baryon-to-entropy ratio becomes

$$\frac{n_b}{s} \approx 4\kappa \alpha_W^4 g_*^{-1} \left(\frac{m}{T}\right)^2 \Delta\theta_{CP} \frac{V_{BG}}{V} (1 - e^{-\bar{\Gamma}\tau}), \quad (168)$$

where  $g_*$  is the number of spin degrees of freedom that enters into the equation for the entropy density.

Note that, although there is a volume suppression factor, in some cases it is  $\mathcal{O}(v_D)$  because the defect network can sweep out that fraction of the total volume in one Hubble expansion time. That is, we are no longer restricted to the initial volume of the defect, as we shall see shortly in an example. However, even if  $V_{BG}/V \sim 1$  there is still a suppression of this mechanism over the usual bubble wall scenarios by the factor

$$(1 - e^{-\bar{\Gamma}L/v_D}).$$

This clearly distinguishes two cases. When the defects are “thin,” defined as  $L < v_D/\bar{\Gamma}$ , there is a suppression factor of approximately  $\bar{\Gamma}L/v_D$ . However, when the defects are “thick,”  $L > v_D/\bar{\Gamma}$ , then there is negligible suppression due to this effect.

Let us now examine how these conditions are related to the microphysical parameters of the models. First consider nonsuperconducting defects. The electroweak symmetry is restored out to a distance given by Eq. (161). The defects are considered “thin” if the Higgs self-coupling  $\lambda$  satisfies

$$\lambda > \left(\frac{\bar{\Gamma}}{v_D \eta_{EW}}\right)^4 \frac{1}{\tilde{G}^2}, \quad (169)$$

and they are considered “thick” otherwise. This quantity may be estimated by evaluating  $\bar{\Gamma}$  at the electroweak temperature, using  $\tilde{G} \sim \mathcal{O}(\frac{1}{30})$  and  $v_D \sim 0.1 - 1$ . This results in the condition

$$\lambda > 10^{-23} - 10^{-27}, \quad (170)$$

an inequality that includes most of the parameter space of the theory. Thus one concludes that, for the case of ordinary defects, the suppression factor  $\bar{\Gamma}L/v_D$  almost always applies.

In some cases, it is possible for the region of symmetry restoration around a string to be enhanced. It was shown by Witten (1995) that some strings can carry supercurrents. In the simplest examples, these currents are due to the presence of a scalar condensate field on the string. The supercurrent is proportional to the winding of this condensate field. There is a large gauge flux associated with the supercurrent, and this flux, coupling to the electroweak Higgs field, leads to an increased region of electroweak symmetry restoration. If, as in Brandenberger *et al.* (1991), we estimate the current on the defects by assuming a random walk of the winding of the condensate field, then we may estimate the size of the symmetry restoration region to be (Ambjørn *et al.*, 1988; Perkins and Davis, 1993)

$$R_s \sim \left(\frac{1}{2\lambda}\right)^{1/2} \frac{1}{2\pi\eta_{EW}} \left(\frac{\eta}{\eta_{EW}}\right)^{3/4}, \quad (171)$$

where  $\eta$  is the scale at which the defects are formed. A similar result has been shown to hold in the two-Higgs-doublet model relevant here (Trodden, 1994). Thus, in this case, the defects are considered “thick” if

$$\eta > \left(\frac{v_D \eta_{EW} \sqrt{2\lambda} 2\pi}{\bar{\Gamma}_s}\right)^{4/3} \eta_{EW} \quad (172)$$

and they are considered “thin” otherwise. Using  $\lambda \sim 1$  and estimating  $\Gamma$  from Eq. (9), we obtain  $\eta > 10^8 - 4 \times 10^{10}$  GeV. Therefore, if the scale of the defects is in this range, there is no additional suppression beyond the volume suppression. If the scale lies below this, then the factor  $\Gamma L/v_D$  applies as in the case of ordinary defects.

The above considerations allow the computation of the asymmetry in the baryon number density at every point swept out by a topological defect of a given type. In order to make a specific prediction one must consider a particular type of defect in a given configuration and have knowledge of the evolution of the defect network. This then provides a reliable estimate for the volume suppression  $\Lambda$  and hence the total baryon asymmetry.

### C. Specific geometry and examples

Let us assume that the network is in the friction-dominated epoch at  $t_{EW}$ . In this case it is reasonable to make the approximation that all string loops have the same radius. Also, note that, for this example, the strings are thin enough (for local baryogenesis) that the additional suppression factor  $\bar{\Gamma}_s L/v_D$  mentioned above applies in both ordinary and superconducting cases (for a large range of Higgs self-coupling). For nonlocal baryogenesis with defects, the suppression factor is linear in  $L v_D/D$ . Further, assume that there is one string loop per correlation volume at formation, via the Kibble mechanism. In one horizon volume the total volume taking part in baryogenesis is

$$V_{BG} = R_s \xi(t)^2 \left( \frac{t}{\xi(t)} \right)^3 v_D, \quad (173)$$

where I have used the largest strings with radius equal to the correlation length  $\xi(t)$  and the last factor is the number of string loops per horizon volume. Thus dividing by the horizon volume  $t^3$  yields the volume suppression factor

$$\Lambda = \frac{V_{BG}}{V} = \frac{R_s}{\xi(t)}. \quad (174)$$

Using  $\xi(t_f) \approx \lambda^{-1} \eta^{-1}$  (Kibble, 1976), where  $t_f$  is the formation time of the string network (Everett, 1981; Kibble, 1982; Hindmarsh, 1986)

$$\xi(t) \sim \xi(t_f) \left( \frac{t}{t_f} \right)^{5/4} \quad (175)$$

gives

$$\Lambda = \lambda \left( \frac{\eta_{EW}}{\eta} \right)^{3/2} v_D \quad \text{ordinary strings} \quad (176)$$

$$= \lambda \left( \frac{\eta_{EW}}{\eta} \right)^{3/4} v_D \quad \text{superconducting strings.} \quad (177)$$

These equations take into account only the dynamics during the first Hubble expansion time after  $t_{EW}$ . In later expansion times, the density of strings is diluted, and hence the above results are a good approximation of the total effect of strings.

Now consider briefly the case in which the strings are formed at a scale much higher than the electroweak scale. If the strings are ordinary, one still expects the  $\bar{\Gamma}_s L / v_D$  suppression, but for superconducting defects we shall see that this is absent since the electroweak symmetry is restored out to such a large radius that all the antibaryons may decay before the baryons are created.

Focus again on string loops. By the time of the electroweak phase transition the defect network is well described by a *scaling solution*. This solution is characterized by the fact that the distribution of string loops looks the same when viewed on all scales. Quantitatively, the number density of string loops with radii in the range  $[R, R+dR]$  is given by (Zel'dovich, 1980; Vilenkin, 1981; Turok and Brandenberger, 1986)

$$n(R, t) = \begin{cases} \nu R^{-5/2} t^{-3/2} & \gamma t < R < t \\ \nu \gamma^{-5/2} t^{-4} & R < \gamma t, \end{cases} \quad (178)$$

where  $\gamma \ll 1$  is a constant determined by the strength of electromagnetic radiation from the string. Loops with radius  $R = \gamma t$  decay in one Hubble expansion time. In the above I have assumed that electromagnetic radiation dominates over gravitational radiation. If this is not the case, then  $\gamma$  must be replaced by  $\gamma_g G \mu$ ,  $\mu$  being the mass per unit length of the string ( $\mu \approx \eta^2$ ) and (Turok, 1984; Burden, 1985; Vachaspati and Vilenkin, 1985)  $\gamma_g \sim 100$ . In other words,  $\gamma$  is bounded from below,

$$\gamma > \gamma_g G \mu. \quad (179)$$

The suppression factor  $\Lambda$  can be estimated by integrating over all the string loops present at  $t_{EW}$ ,

$$\Lambda \approx \pi \int_0^{\eta_{EW}} dR R^2 R_s n(R, t_{EW}) = \frac{\pi}{3} \nu \gamma^{1/2} \left( \frac{R_s}{t_{EW}} \right). \quad (180)$$

Without superconductivity the suppression factor for Grand unified theory (GUT) strings ( $\eta = 10^{16}$  GeV) is so small ( $\sim 10^{-32}$ ) that the contribution is negligible. However, for superconducting strings the suppression is

$$\Lambda \sim \nu \gamma^{1/2} \left( \frac{\eta}{m_{pl}} \right), \quad (181)$$

so that the final baryon-to-entropy ratio generated by this mechanism is given by

$$\frac{n_B}{s} = \frac{n_B^0}{s} \Lambda, \quad (182)$$

with  $\Lambda$  given by Eq. (180) and  $n_B^0/s$  proportional to  $\alpha_W^4$  for local baryogenesis and to  $\alpha_W^2$  for nonlocal baryogenesis. Clearly, this lies below the observed value, and therefore strings formed at high-energy scales are not viable candidates for mediating electroweak baryogenesis.

#### D. Particle physics model

The results above indicate that particle physics models that admit cosmic strings at or around the TeV scale are perhaps the best candidates for implementing the defect-mediated scenario. The particle physics literature contains many such examples. Here I shall give just one.

The particular supersymmetric model I shall consider (Suematsu and Yamagishi, 1995) has been proposed as a solution to the  $\mu$  problem of the MSSM and the cosmological solar neutrino problem.

In the MSSM there exists a mixing term of the form

$$\mathcal{L}_\mu = \mu \bar{H} H, \quad (183)$$

where  $H$  is the supersymmetric Higgs field. In order to obtain radiative SUSY breaking at the weak scale it is necessary that  $\mu \sim \mathcal{O}(G_F^{-1/2})$  where  $G_F$  is the Fermi constant. However, there is no natural scale in the MSSM to ensure that this is the case.

In the model under consideration the MSSM is supplemented by two  $U(1)$  symmetries. One of the extra  $U(1)$ 's breaks at a high scale ( $\sim 10^{15}$  GeV) and is concerned with the implementation of the Mikheyev, Smirnov, and Wolfenstein (MSW, Wolfenstein, 1979; Mikheyev and Smirnov, 1985) solution of the solar neutrino problem via the seesaw mechanism. I shall not discuss that aspect of the model any further. The  $\mu$  term in this model is given in terms of a Yukawa coupling  $\lambda'$  and a scalar  $S$  which is a singlet under the standard-model gauge group but charged under the low-energy extra  $U(1)$ . Thus the term (183) is forbidden, since it is not invariant under the extra  $U(1)$  symmetry, and in its place we have a term

$$\mathcal{L}_\mu = \lambda' S \bar{H} H. \quad (184)$$



Therefore if the low-energy  $U(1)$  breaks at a scale  $\eta$  of the order of 1 TeV, then  $S$  gets a VEV of this order and the  $\mu$  problem is resolved.

The symmetry-breaking scheme of the model is

$$\begin{aligned}
 & SU(3)_c \times SU(2)_L \times U(1)_Y \times U(1) \times U(1) \\
 & \rightarrow SU(3)_c SU(2)_L \times U(1)_Y \times U(1) \\
 & \xrightarrow{\eta} SU(3)_c \times SU(2)_L U(1)_Y \\
 & \xrightarrow{\eta_{EW}} SU(3)_c \times U(1)_{em}. \tag{185}
 \end{aligned}$$

Clearly we obtain TeV-scale ordinary cosmic strings from this final  $U(1)$  breaking, and this model would exhibit defect-mediated electroweak baryogenesis.

### E. Summary

I have described an alternative scenario to traditional electroweak baryogenesis, in which the out-of-equilibrium evolution of a network of topological defects satisfies the third Sakharov condition. The advantage of such a scenario for electroweak baryogenesis is that it does not depend in any way on the order of the electroweak phase transition. A further advantage of using topological defects to seed baryogenesis is that the volume in defects decreases only as a power of time below the phase-transition temperature. Therefore, as pointed out by Brandenberger *et al.* (1991), defect-mediated baryogenesis remains effective even if sphalerons are in thermal equilibrium just below the electroweak phase-transition temperature. This alleviates the problem of washout of the asymmetry. However, a potential drawback for specific implementations is the requirement that defects be formed at a scale rather close to the electroweak scale in order to avoid large volume suppression factors (for further constraints in the specific case of ordinary cosmic strings see Cline, Espinosa, *et al.*, 1998).

One might worry that, since the baryon production in these models occurs inhomogeneously, nucleosynthesis might proceed inhomogeneously, leading to a conflict with observation. This possibility has been investigated (Brandenberger *et al.*, 1995) and, for defects formed at low enough temperatures that the volume suppression factor is not prohibitively small, the baryons produced are sufficiently homogeneous at the time of nucleosynthesis.

### IX. CONCLUDING REMARKS AND A LOOK TO THE FUTURE

Modern particle cosmology consists of the union of the hot big-bang model with quantum field theories of elementary particles. Under mild assumptions, it is a consequence of this structure that, when the universe was extremely young and hot, the net baryon number of the universe was zero. That is, the number of particles carrying a given baryon number in any region was equal

on average to the number of the appropriate antiparticles carrying the opposite baryon number. However, on the other hand, it is a clear observational fact that the universe is maximally baryon-antibaryon asymmetric. This fact is quantified from the considerations of primordial nucleosynthesis. These calculations are perhaps the most impressive success of the standard cosmology, accurately predicting the abundances of the light elements from the single input parameter of the baryon-to-entropy ratio, which is constrained as in Eq. (2). Until recently, there were two possible explanations for this. First, the universe as a whole could be baryon number symmetric, but baryon-antibaryon separation could have resulted in an apparent baryon asymmetry in the local universe. The second possibility is that some dynamical process took place as the universe evolved, causing baryons to be preferentially produced over antibaryons. A recent analysis (Cohen *et al.*, 1998) has ruled out the former possibility over scales up to the size of the observable universe. It therefore appears that the latter option, *baryogenesis*, must have taken place.

In this article I have tried to describe how a number of different physical effects, all present in the standard electroweak theory at nonzero temperature, can come together in the context of the expanding universe to implement baryogenesis. It is an amazing fact about the Glashow-Salam-Weinberg model and its modest extensions that they satisfy all three Sakharov criteria for producing a baryon excess. As a result, over the last decade, *electroweak baryogenesis* has been a very popular scenario for the generation of the baryon asymmetry of the universe.

There are four technical issues to be investigated when considering models of electroweak baryogenesis. These are

- (1) How is the departure from equilibrium realized? Is the electroweak phase transition strongly first order or are topological defects necessary?
- (2) How is sufficient  $CP$  violation obtained?
- (3) What is the rate of baryon-number-violating processes? Is this rate high enough in the unbroken phase and can washout of any asymmetry be avoided?
- (4) What are the actual dynamics of baryon number production?

I hope I have described how each of these issues has been addressed in the literature on electroweak baryogenesis. Detailed analyses of the phase transition, coupled with the smallness of the  $CP$  violation due to phases in the Cabibbo-Kobayashi-Maskawa matrix, have made it clear that an extension of the standard model is required to make the scenarios viable. While general two-Higgs-doublet theories have been considered, perhaps the most appealing candidate from particle physics motivations is the minimal supersymmetric standard model. The implementations of electroweak baryogenesis in this model have recently been investigated and the range of parameters of the model for which an appreciable baryon asymmetry of the universe

can be generated have been calculated. This range should be accessible to future particle colliders, and thus the scenario of electroweak baryogenesis in the MSSM should be experimentally testable. If the MSSM or two-Higgs models are not chosen by nature, then other models of electroweak baryogenesis may be relevant. If the topological structure of the relevant theory admits gauge solitons, then defect-mediated electroweak baryogenesis may be important, irrespective of the order of the phase transition.

Although I have argued that there exist viable scenarios of electroweak baryogenesis, particularly in the context of supersymmetric models, there remain a number of open questions and directions for future research. I shall list the ones that I feel are most important.

- (1) At present, the best quantitative understanding of the electroweak phase transition comes from lattice Monte Carlo simulations. While the results from these are impressive, they do not provide an intuitive understanding of the microphysics of the phase transition. The numerical results are partially supported by some analytical approaches, but these often cannot be trusted in the physical range of Higgs masses. An analytic understanding of the nonperturbative dynamics of the phase transition would be an important step forward.
- (2) The chemical potential and “fluid” approaches to nonlocal baryogenesis provide good analytical tools for understanding the nonlocal production of baryons. However, in the case of local baryogenesis, analytical methods that yield believable quantitative results have yet to be found. Again, it is encouraging that numerical calculations, for example of Chern-Simons number diffusion, are providing quantitative predictions, but an appropriate analytical model is very desirable.
- (3) While it will strongly support electroweak baryogenesis if supersymmetry is verified with parameters in the correct range, it is important to ensure that there is sufficient  $CP$  violation. To this end, phenomenological predictions of  $CP$  violation in SUSY models and the corresponding experimental tests are crucial hurdles that the scenario must pass.

If electroweak baryogenesis is correct, then no matter what the relevant electroweak model is, the physics involved is extremely beautiful and diverse. The topology of gauge-field theories, the physics of phase transitions,  $CP$  violation, plasma dynamics, and thermal field theory all play a part in generating the baryon asymmetry of the universe. However, perhaps the most attractive feature of electroweak baryogenesis scenarios is that they should be testable at the next generation of particle colliders. It is an exciting possibility that we may soon understand the origin of the matter that makes up our universe.

#### ACKNOWLEDGMENTS

I first want to thank my collaborators, Robert Brandenberger, Anne Davis, Arthur Lue, Tomislav Proko-

pec, and Krishna Rajagopal, for the pleasure of working with them, for helping shape my ideas about baryogenesis, and for their permission to use parts of our joint work in this article. I also owe thanks to Robert Brandenberger, James Cline, Marcelo Gleiser, Michael Joyce, Mikko Laine, Guy Moore, Mikhail Shaposhnikov, and Tanmay Vachaspati for a critical reading of the first draft of this work and for their advice. Their many suggestions and comments were invaluable and undoubtedly made this a much better article. Finally, I should acknowledge further advice and discussions during the time I have spent thinking about electroweak baryogenesis with Andrew Cohen, Edward Farhi, Jeffrey Goldstone, Alan Guth, Kenneth Johnson, Lawrence Krauss, Lisa Randall, Glenn Starkman, and Neil Turok. This work was supported by the U.S. Department of Energy, the National Science Foundation, and by funds provided by Case Western Reserve University. I would also like to acknowledge the Center for Theoretical Physics at MIT, where I began this work.

#### REFERENCES

- Adler, S., 1969, *Phys. Rev.* **177**, 2426.  
 Affleck, I., 1981a, *Nucl. Phys. B* **191**, 429.  
 Affleck, I., 1981b, *Phys. Rev. Lett.* **46**, 388.  
 Ahlen, S. P., *et al.*, 1988, *Phys. Rev. Lett.* **61**, 145.  
 Altarev, I. S., *et al.*, 1992, *Phys. Lett. B* **276**, 242.  
 Ambjørn, J., T. Askgaard, H. Porter, and M. Shaposhnikov, 1990, *Phys. Lett. B* **244**, 479.  
 Ambjørn, J., T. Askgaard, H. Porter, and M. Shaposhnikov, 1991, *Nucl. Phys. B* **353**, 346.  
 Ambjørn, J., and A. Krasnitz, 1995, *Phys. Lett. B* **362**, 97.  
 Ambjørn, J., and A. Krasnitz, 1997, *Nucl. Phys. B* **506**, 387.  
 Ambjørn, J., M. Laursen, and M. Shaposhnikov, 1989, *Nucl. Phys. B* **316**, 483.  
 Ambjørn, J., N. Nielsen, and P. Olesen, 1988, *Nucl. Phys. B* **310**, 625.  
 Anderson, C. D., 1932, *Science* **76**, 238.  
 Anderson, C. D., 1933, *Phys. Rev.* **43**, 491.  
 Aoki, Y., 1997, *Phys. Rev. D* **56**, 3860.  
 Applequist, T., and R. Pisarski, 1981, *Phys. Rev. D* **23**, 2305.  
 Arnold, P., 1993, *Phys. Rev. D* **48**, 1539.  
 Arnold, P., 1997, *Phys. Rev. D* **55**, 7781.  
 Arnold, P., E. Braaten, and S. Vokos, 1992, *Phys. Rev. D* **46**, 3576.  
 Arnold, P., and O. Espinosa, 1993, *Phys. Rev. D* **47**, 3546.  
 Arnold, P., and L. McLerran, 1987, *Phys. Rev. D* **36**, 581.  
 Arnold, P., D. T. Son, and L. G. Yaffe, 1997, *Phys. Rev. D* **55**, 6264.  
 Arnold, P., and L. Yaffe, 1994, *Phys. Rev. D* **49**, 3003.  
 Baacke, J., and S. Junker, 1993, *Mod. Phys. Lett. A* **8**, 2869.  
 Baacke, J., and S. Junker, 1994a, *Phys. Rev. D* **49**, 2055.  
 Baacke, J., and S. Junker, 1994b, *Phys. Rev. D* **50**, 4227.  
 Barriola, M., 1995, *Phys. Rev. D* **51**, 300.  
 Bell, J. S., and R. Jackiw, 1969, *Nuovo Cimento A* **51**, 47.  
 Bergerhoff, B., and C. Wetterich, 1995, *Nucl. Phys. B* **440**, 171.  
 Bochkarev, A. I., S. V. Kuzmin, and M. E. Shaposhnikov, 1990, *Phys. Lett. B* **244**, 275.  
 Bochkarev, A. I., and M. E. Shaposhnikov, 1987, *Mod. Phys. Lett. A* **2**, 417.  
 Bödeker, D., 1998, *Phys. Lett. B* **426**, 351.

- Bödeker, D., P. John, M. Laine, and M. G. Schmidt, 1997, Nucl. Phys. B **497**, 387.
- Borrill, J., and M. Gleiser, 1995, Phys. Rev. D **51**, 4111.
- Borrill, J., and M. Gleiser, 1997, Nucl. Phys. B **483**, 416.
- Brandenberger, R. H., and A.-C. Davis, 1993, Phys. Lett. B **308**, 79.
- Brandenberger, R. H., A.-C. Davis, and M. Hindmarsh, 1991, Phys. Lett. B **263**, 239.
- Brandenberger, R. H., A.-C. Davis, T. Prokopec, and M. Trodden, 1996, Phys. Rev. D **53**, 4257.
- Brandenberger, R. H., A.-C. Davis, and M. J. Rees, 1995, Phys. Lett. B **349**, 329.
- Brandenberger, R. H., A.-C. Davis, and M. Trodden, 1994, Phys. Lett. B **335**, 123.
- Brignole, A., J. R. Espinosa, M. Quiros, and F. Zwirner, 1994, Phys. Lett. B **324**, 181.
- Buchmuller, W., and Z. Fodor, 1994, Phys. Lett. B **331**, 124.
- Buchmuller, W., Z. Fodor, and A. Hebecker, 1995, Nucl. Phys. B **447**, 317.
- Buchmuller, W., and O. Philipsen, 1995, Nucl. Phys. B **443**, 47.
- Burden, C., 1985, Phys. Lett. B **164**, 277.
- Carena, M., M. Quiros, A. Riotto, I. Vilja, and C. E. M. Wagner, 1997, Nucl. Phys. B **503**, 387.
- Carena, M., M. Quiros, and C. E. M. Wagner, 1996, Nucl. Phys. B **461**, 407.
- Carena, M., M. Quiros, and C. E. M. Wagner, 1999, Nucl. Phys. B **524**, 3.
- Carson, L., X. Li, L. McLerran, and R. Wang, 1990, Phys. Rev. D **42**, 2127.
- Carson, L., and L. McLerran, 1990, Phys. Rev. D **41**, 647.
- Christ, N., 1980, Phys. Rev. D **43**, 1781.
- Christenson, J. H., J. W. Cronin, V. L. Fitch, and R. Turlay, 1964, Phys. Rev. Lett. **13**, 138.
- Cline, J. M., 1994, Phys. Lett. B **338**, 263.
- Cline, J., J. Espinosa, G. Moore, and A. Riotto, 1999, Phys. Rev. D **59**, 065014.
- Cline, J. M., M. Joyce, and K. Kainulainen, 1998, Phys. Lett. B **417**, 79.
- Cline, J. M., and K. Kainulainen, 1996, Nucl. Phys. B **482**, 73.
- Cline, J. M., and K. Kainulainen, 1998, Nucl. Phys. B **510**, 88.
- Cline, J. M., K. Kainulainen, and A. P. Vischer, 1996, Phys. Rev. D **54**, 2451.
- Cline, J. M., and P.-A. Lemieux, 1997, Phys. Rev. D **55**, 3873.
- Cohen, A. G., A. De Rujula, and S. L. Glashow, 1998, Astrophys. J. **495**, 539.
- Cohen, A. G., and D. B. Kaplan, 1987, Phys. Lett. B **199**, 251.
- Cohen, A. G., and D. B. Kaplan, 1988, Nucl. Phys. B **308**, 913.
- Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1990, Phys. Lett. B **245**, 561.
- Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1991a, Nucl. Phys. B **349**, 727.
- Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1991b, Phys. Lett. B **263**, 86.
- Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1992, Phys. Lett. B **294**, 57.
- Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1993, Annu. Rev. Nucl. Part. Sci. **43**, 27.
- Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1994, Phys. Lett. B **336**, 41.
- Cohen, A. G., and A. E. Nelson, 1992, Phys. Lett. B **297**, 111.
- Coleman, S., 1979, "The Uses of Instantons," in *The Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum, New York).
- Coleman, S., and E. J. Weinberg, 1973, Phys. Rev. D **7**, 1888.
- Comelli, D., M. Pietroni, and A. Riotto, 1995, Phys. Lett. B **354**, 91.
- Commins, E. D., S. B. Ross, D. DeMille, and B. C. Regan, 1994, Phys. Rev. A **50**, 2960.
- Copi, C. J., D. N. Schramm, and M. S. Turner, 1995, Science **267**, 192.
- Csikor, F., Z. Fodor, J. Hein, A. Jaster, and I. Montvay, 1996, Nucl. Phys. B **474**, 421.
- Dashen, R., B. Hasslacher, and A. Neveu, 1974, Phys. Rev. D **10**, 4138.
- Davies, A. T., C. D. Froggatt, G. Jenkins, and R. G. Moorhouse, 1994, Phys. Lett. B **336**, 464.
- Davies, A. T., C. D. Froggatt, and R. G. Moorhouse, 1996, Phys. Lett. B **372**, 88.
- Davoudiasl, H., K. Rajagopal, and E. Westphal, 1998, Nucl. Phys. B **515**, 384.
- de Jong, S., 1998, "Higgs Searches at LEP," review talk delivered at the XXXIIIrd Rencontres de Moriond, Les Arcs, France, unpublished.
- Delepine, D., J. Gerard, R. Gonzalez Filipe, and J. Weyers, 1996, Phys. Lett. B **386**, 183.
- D'Hoker, E., and E. Farhi, 1984, Nucl. Phys. B **241**, 109.
- Dimopoulos, S., and L. Susskind, 1978, Phys. Rev. D **18**, 4500.
- Dine, M., P. Huet, and R. Singleton, Jr., 1992, Nucl. Phys. B **375**, 625.
- Dine, M., P. Huet, R. Singleton, Jr., and L. Susskind, 1991, Phys. Lett. B **257**, 351.
- Dine, M., O. Lechtenfeld, B. Sakita, W. Fischler, and J. Polchinski, 1990, Nucl. Phys. B **342**, 381.
- Dine, M., R. G. Leigh, P. Huet, A. D. Linde, and D. Linde, 1992, Phys. Rev. D **46**, 550.
- Dine, M., and S. Thomas, 1994, Phys. Lett. B **328**, 73.
- Dirac, P. A. M., 1930, Proc. R. Soc. London, Ser. A **126**, 360.
- Dirac, P. A. M., 1931, Proc. R. Soc. London, Ser. A **133**, 60.
- Dolan, L., and R. Jackiw, 1974, Phys. Rev. D **9**, 3320.
- Dolgov, A., 1992, Phys. Rep. **222**, 309.
- Dosch, H., J. Kripfganz, A. Laser, and M. Schmidt, 1996, Phys. Lett. B **365**, 213.
- Dosch, H., J. Kripfganz, A. Laser, and M. Schmidt, 1997, Nucl. Phys. B **507**, 519.
- Duari, S. B., and U. A. Yajnik, 1996, Mod. Phys. Lett. A **11**, 2481.
- Enqvist, K., J. Ignatius, K. Kajantie, and K. Rummukainen, 1992, Phys. Rev. D **45**, 3415.
- Enqvist, K., A. Riotto, and I. Vilja, 1998, Phys. Lett. B **438**, 273.
- Espinosa, J. R., 1996, Nucl. Phys. B **475**, 273.
- Everett, A., 1981, Phys. Rev. D **24**, 858.
- Farakos, K., K. Kajantie, K. Rummukainen, and M. Shaposhnikov, 1994a, Nucl. Phys. B **425**, 67.
- Farakos, K., K. Kajantie, K. Rummukainen, and M. Shaposhnikov, 1994b, Nucl. Phys. B **442**, 317.
- Farakos, K., K. Kajantie, K. Rummukainen, and M. Shaposhnikov, 1994c, Phys. Lett. B **336**, 494.
- Farhi, E., J. Goldstone, S. Gutmann, K. Rajagopal, and R. Singleton, Jr., 1995, Phys. Rev. D **51**, 4561.
- Farhi, E., J. Goldstone, A. Lue, and K. Rajagopal, 1996, Phys. Rev. D **54**, 5336.
- Farrar, G. R., and M. Losada, 1996, Phys. Lett. B **406**, 60.
- Farrar, G. R., and M. E. Shaposhnikov, 1994a, Phys. Rev. D **50**, 774.

- Farrar, G. R., and M. E. Shaposhnikov, 1994b, "Note Added to "Baryon Asymmetry of the Universe in the Standard Model," hep-ph/9406387.
- Farrar, G. R., and M. E. Shaposhnikov, 1995a, Phys. Rev. Lett. **70**, 2833.
- Farrar, G. R., and M. E. Shaposhnikov, 1995a, Phys. Rev. Lett. **71**, 210.
- Fodor, Z., and A. Hebecker, 1994, Nucl. Phys. B **432**, 127.
- Frere, J. M., L. Houart, J. Moreno, J. Orloff, and M. Tytgat, 1993, Phys. Lett. B **314**, 289.
- Funakubo, K., A. Kakuto, S. Otsuki, K. Takenaga, and F. Toyoda, 1997a, Prog. Theor. Phys. **94**, 845.
- Funakubo, K., A. Kakuto, S. Otsuki, and F. Toyoda, 1997b, Prog. Theor. Phys. **98**, 427.
- Gavela, M. B., P. Hernandez, J. Orloff, O. Pene, and C. Quimbay, 1995a, Mod. Phys. Lett. A **9**, 795.
- Gavela, M. B., P. Hernandez, J. Orloff, O. Pene, and C. Quimbay, 1995b, Nucl. Phys. B **430**, 382.
- Gavela, M. B., M. Lozano, J. Orloff, and O. Pène, 1994, Nucl. Phys. B **430**, 345.
- Ginsparg, P., 1980, Nucl. Phys. B **170**, 388.
- Giudice, G. F., 1992, Phys. Rev. D **45**, 3177.
- Giudice, G. F., and M. E. Shaposhnikov, 1994, Phys. Lett. B **326**, 118.
- Glashow, S. L., 1961, Nucl. Phys. **22**, 579.
- Gleiser, M., 1993, Phys. Rev. D **49**, 2978.
- Gleiser, M., 1994, Phys. Rev. Lett. **73**, 3495.
- Gleiser, M., 1995, Phys. Rev. Lett. **75**, 589.
- Gleiser, M., and A. F. Heckler, 1996, Phys. Rev. Lett. **76**, 180.
- Gleiser, M., A. F. Heckler, and E. W. Kolb, 1997, Phys. Lett. B **405**, 121.
- Gleiser, M., and E. W. Kolb, 1992, Phys. Rev. Lett. **69**, 1304.
- Gleiser, M., and E. W. Kolb, 1993, Phys. Rev. D **48**, 1560.
- Gleiser, M., and R. Ramos, 1994, Phys. Rev. D **50**, 2441.
- Goldstone, J., and F. Wilczek, 1981, Phys. Rev. Lett. **47**, 986.
- Grigorev, D., M. Shaposhnikov, and N. Turok, 1992, Phys. Lett. B **275**, 395.
- Gross, D., R. Pisarski, and L. Yaffe, 1981, Rev. Mod. Phys. **53**, 43.
- Gunion, J. F., H. E. Haber, G. L. Kane, and S. Dawson, 1989, *The Higgs Hunter's Guide* (Addison-Wesley, Redwood City, CA).
- Gurtler, M., E. M. Ilgenfritz, J. Kripfganz, J. Perlt, and A. Schiller, 1997, Nucl. Phys. B **483**, 383.
- Gurtler, M., E. M. Ilgenfritz, and A. Schiller, 1997, Phys. Rev. D **56**, 3888.
- Gurtler, M., E. M. Ilgenfritz, and A. Schiller, 1998a, Nucl. Phys. B, Proc. Suppl. **63**, 563.
- Gurtler, M., E. M. Ilgenfritz, and A. Schiller, 1998b, Eur. Phys. J. C **1**, 363.
- Gyulassy, M., K. Kajantie, H. Kurki-Suonio, and L. McLerran, 1983, Nucl. Phys. B **237**, 477.
- Haber, H. E., R. Hempfling, and H. Hoang, 1997, Z. Phys. C **75**, 539.
- Heckler, A. F., 1995, Phys. Rev. D **51**, 405.
- Hindmarsh, M., 1986, Ph.D. thesis, University of London.
- Hogan, C., 1997, "Extragalactic Abundances of Hydrogen, Deuterium and Helium: New Steps, Missteps and Next Steps," talk given at the *Primordial Nuclei and their Galactic Evolution Workshop (ISSI)*, astro-ph/9712031.
- Hu, C., and B. Mueller, 1997, Phys. Lett. B **409**, 377.
- Huet, P., K. Kajantie, R. G. Leigh, B.-H. Liu, and L. McLerran, 1993, Phys. Rev. D **48**, 2477.
- Huet, P., and A. E. Nelson, 1995, Phys. Lett. B **355**, 229.
- Huet, P., and A. E. Nelson, 1996, Phys. Rev. D **53**, 4578.
- Huet, P., and E. Sather, 1995, Phys. Rev. D **51**, 379.
- Huet, P., and D. T. Son, 1997, Phys. Lett. B **393**, 94.
- Iancu, E., 1997, "How to Compute the Anomalous Baryon Number Violation at High Temperature," hep-ph/9710543.
- Ignatius, J., K. Kajantie, H. Kurki-Suonio, and M. Laine, 1994, Phys. Rev. D **49**, 3854.
- Joyce, M., 1994, "A Note on Spontaneous Baryogenesis," in *Sintra Electroweak 1994*.
- Joyce, M., and T. Prokopec, 1998, Phys. Rev. D **57**, 6022.
- Joyce, M., T. Prokopec, and N. Turok, 1994a, Phys. Lett. B **338**, 269.
- Joyce, M., T. Prokopec, and N. Turok, 1994b, Phys. Lett. B **339**, 312.
- Joyce, M., T. Prokopec, and N. Turok, 1994c, Phys. Rev. Lett. **75**, 1695; **75**, 3375(E).
- Joyce, M., T. Prokopec, and N. Turok, 1996a, Phys. Rev. D **53**, 2930.
- Joyce, M., T. Prokopec, and N. Turok, 1996b, Phys. Rev. D **53**, 2958.
- Kajantie, K., K. Rummukainen, and M. Shaposhnikov, 1993, Nucl. Phys. B **407**, 356.
- Kajantie, K., M. Laine, K. Rummukainen, and M. Shaposhnikov, 1996a, Nucl. Phys. B **458**, 90.
- Kajantie, K., M. Laine, K. Rummukainen, and M. Shaposhnikov, 1996b, Nucl. Phys. B **466**, 189.
- Kajantie, K., M. Laine, K. Rummukainen, and M. Shaposhnikov, 1996c, Nucl. Phys. B **493**, 413.
- Kajantie, K., M. Laine, K. Rummukainen, and M. Shaposhnikov, 1996d, Phys. Rev. Lett. **77**, 2887.
- Karsch, F., T. Neuhaus, A. Patkos, and J. Rank, 1996, Nucl. Phys. B **474**, 217.
- Khlebnikov, S. Yu, 1992, Phys. Rev. D **46**, 3223.
- Khlebnikov, S. Yu, and M. Shaposhnikov, 1988, Nucl. Phys. B **308**, 885.
- Kibble, T. W. B., 1976, J. Phys. A **9**, 1387.
- Kibble, T. W. B., 1982, Acta Phys. Pol. B **13**, 723.
- Kirzhnits, D. A., 1972, Zh. Eksp. Teor. Fiz., Pis'ma Red. **15**, 745 [JETP Lett. **15**, 529].
- Kirzhnits, D. A., and A. D. Linde, 1972, Phys. Lett. B **42**, 471.
- Klinkhamer, F. R., and N. S. Manton, 1984, Phys. Rev. D **30**, 2212.
- Kobayashi, M., and K. Maskawa, 1973, Prog. Theor. Phys. **49**, 652.
- Krauss, L., and P. Kernan, 1995, Phys. Lett. B **347**, 347.
- Kurki-Suonio, H., 1985, Nucl. Phys. B **255**, 231.
- Kurki-Suonio, H., and M. Laine, 1995, Phys. Rev. D **51**, 5431.
- Kurki-Suonio, H., and M. Laine, 1996a, Phys. Rev. D **54**, 7163.
- Kurki-Suonio, H., and M. Laine, 1996b, Phys. Rev. Lett. **77**, 3951.
- Kuzmin, V. A., V. A. Rubakov, and M. E. Shaposhnikov, 1985, Phys. Lett. B **155**, 36.
- Laine, M., 1994, Phys. Rev. D **49**, 3847.
- Laine, M., 1995, Nucl. Phys. B **451**, 484.
- Laine, M., 1996a, Nucl. Phys. B **481**, 43.
- Laine, M., 1996b, Phys. Lett. B **385**, 249.
- Laine, M., and K. Rummukainen, 1998, Phys. Rev. Lett. **80**, 5259.
- Landsman, N. P., 1989, Nucl. Phys. B **322**, 498.
- Langer, J., 1967, Ann. Phys. (N.Y.) **41**, 108.
- Linde, A. D., 1980, Phys. Lett. B **96**, 289.
- Linde, A. D., 1981, Phys. Lett. B **100**, 37.

- Liu, B.-H., L. McLerran, and N. Turok, 1992, *Phys. Rev. D* **46**, 2668.
- Losada, M., 1996a, “The Electroweak Phase Transition in the Minimal Supersymmetric Standard Model,” preprint, hep-ph/9612337.
- Losada, M., 1996b, *Phys. Rev. D* **56**, 2893.
- Lue, A., K. Rajagopal, and M. Trodden, 1997, *Phys. Rev. D* **55**, 1250.
- MacKenzie, R., and F. Wilczek, 1984a, *Phys. Rev. D* **30**, 2194.
- MacKenzie, R., and F. Wilczek, 1984b, *Phys. Rev. D* **30**, 2260.
- MacKenzie, R., F. Wilczek, and A. Zee, 1984, *Phys. Rev. Lett.* **53**, 2203.
- Manton, N. S., 1983, *Phys. Rev. D* **28**, 2019.
- McLerran, M., 1989, *Phys. Rev. Lett.* **62**, 1075.
- McLerran, L., M. Shaposhnikov, N. Turok, and M. Voloshin, 1991, *Phys. Lett. B* **256**, 451.
- Mikheyev, S., and A. Smirnov, 1985, *Yad. Fiz.* **42** [*Sov. J. Nucl. Phys.* **42**, 913].
- Mohapatra, R., and X. Zhang, 1992, *Phys. Rev. D* **46**, 5331.
- Moore, G. D., 1996, *Nucl. Phys. B* **480**, 657.
- Moore, G. D., 1997, *Phys. Lett. B* **412**, 359.
- Moore, G. D., 1998a, *Phys. Lett. B* **439**, 357.
- Moore, G. D., 1998b, “The Sphaleron Rate: Bödeker’s Leading Log,” preprint, hep-ph/9810313.
- Moore, G. D., C. Hu, and B. Mueller, 1998, *Phys. Rev. D* **58**, 045001.
- Moore, G. D., and T. Prokopec, 1995a, *Phys. Rev. Lett.* **75**, 777.
- Moore, G. D., and T. Prokopec, 1995b, *Phys. Rev. D* **52**, 7182.
- Moore, G. D., and N. Turok, 1997a, *Phys. Rev. D* **55**, 6538.
- Moore, G. D., and N. Turok, 1997b, *Phys. Rev. D* **56**, 6533.
- Nadkarni, S., 1983, *Phys. Rev. D* **27**, 388.
- Nasser, S., and N. Turok, 1994, “Z Condensation and Standard Model Baryogenesis,” hep-ph/9406270 (unpublished).
- Nelson, A. E., D. B. Kaplan, and A. G. Cohen, 1992, *Nucl. Phys. B* **373**, 453.
- Parwani, R., 1992, *Phys. Rev. D* **45**, 4695.
- Perkins, W. B., 1995, *Nucl. Phys. B* **449**, 265.
- Perkins, W. B., and A.-C. Davis, 1993, *Nucl. Phys. B* **406**, 377.
- Prokopec, T., R. H. Brandenberger, A.-C. Davis, and M. Trodden, 1996, *Phys. Lett. B* **384**, 175.
- Ratra, B., and L. G. Yaffe, 1988, *Phys. Lett. B* **205**, 57.
- Reina, L., and M. Tytgat, 1994, *Phys. Rev. D* **50**, 751.
- Reuter, M., and C. Wetterich, 1993, *Nucl. Phys. B* **408**, 91.
- Riotto, A., 1998a, *Phys. Rev. D* **58**, 095009.
- Riotto, A., 1998b, *Nucl. Phys. B* **518**, 339.
- Riotto, A., 1998c, *Int. J. Mod. Phys. D* **7**, 815.
- Rubakov, V. A., and M. E. Shaposhnikov, 1996, *Phys. Usp.* **39**, 461.
- Rummukainen, K., 1997, *Nucl. Phys. B, Proc. Suppl.* **53**, 30.
- Sakharov, A. D., 1967, *Zh. Eksp. Teor. Fiz., Pis'ma Red.* **5**, 32 [*JETP Lett.* **5**, 24 (1967)].
- Salam, A., 1968, in *Proceedings of the Eighth Nobel Symposium*, edited by N. Svartholm (Wiley-Interscience, New York), p. 367.
- Shaposhnikov, M., 1986a, *JETP Lett.* **44**, 465.
- Shaposhnikov, M., 1986b, *Nucl. Phys. B* **287**, 757.
- Shaposhnikov, M., 1988, *Nucl. Phys. B* **299**, 797.
- Smith, K. F., *et al.*, 1990, *Phys. Lett. B* **234**, 191.
- Son, D., 1997, “Effective Nonperturbative Real Time Dynamics of Soft Modes in Hot Gauge Theories,” preprint, hep-ph/9707351.
- Stecker, F. W., 1985, *Nucl. Phys. B* **252**, 25.
- Steigman, G., 1976, *Annu. Rev. Astron. Astrophys.* **14**, 336.
- Steinhardt, P. J., 1982, *Phys. Rev. D* **25**, 2074.
- Suematsu, D., and Y. Yamagishi, 1995, *Int. J. Mod. Phys. A* **10**, 4521.
- Tang, W.-H., and J. Smit, 1996, *Nucl. Phys. B* **482**, 265.
- 't Hooft, G., 1976, *Phys. Rev. Lett.* **37**, 8.
- Trodden, M., 1994, *Mod. Phys. Lett. A* **9**, 2649.
- Trodden, M., A.-C. Davis, and R. H. Brandenberger, 1995, *Phys. Lett. B* **349**, 131.
- Turok, N., 1984, *Nucl. Phys. B* **242**, 520.
- Turok, N., 1992, *Phys. Rev. Lett.* **68**, 1803.
- Turok, N., 1993, “Electroweak Baryogenesis,” in *Perspectives on Higgs Physics*, edited by G. L. Kane (World Scientific, Singapore), p. 300.
- Turok, N., and R. Brandenberger, 1986, *Phys. Rev. D* **37**, 2075.
- Turok, N., and J. Zadrozny, 1990, *Phys. Rev. Lett.* **65**, 2331.
- Turok, N., and J. Zadrozny, 1991, *Nucl. Phys. B* **358**, 471.
- Tytler, D., X.-M. Fan, and S. Burles, 1996, *Nature (London)* **381**, 207.
- Vachaspati, T., and G. B. Field, 1994, *Phys. Rev. Lett.* **73**, 373; **74**, 1258(E) (1995).
- Vachaspati, T., and A. Vilenkin, 1985, *Phys. Rev. D* **31**, 3052.
- Vilenkin, A., 1981, *Phys. Rev. Lett.* **46**, 1169.
- Weinberg, S. W., 1967, *Phys. Rev. Lett.* **19**, 1264.
- Witten, E., 1977, *Phys. Rev. Lett.* **38**, 121.
- Witten, E., 1985, *Nucl. Phys. B* **249**, 557.
- Wolfenstein, L., 1979, *Phys. Rev. D* **20**, 2634.
- Worah, M., 1997, *Phys. Rev. D* **56**, 2010.
- Zel'dovich, Ya. B., 1980, *Mon. Not. R. Astron. Soc.* **192**, 663.
- Zhang, X., S. Lee, K. Whisnant, and B. Young, 1994, *Phys. Rev. D* **50**, 7042.
- Zhang, X., and B.-L. Young, 1994, *Phys. Rev. D* **49**, 563.