Dynamical Generation of Baryons at the Electroweak Transition

Neil Turok and John Zadrozny
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544
(Received 26 June 1990; revised manuscript received 13 September 1990)

A new mechanism through which a net baryon asymmetry could be generated at the electroweak phase transition is discussed. It works efficiently in "nonminimal" extensions of the standard model such as occur in supersymmetric theories, where the Higgs sector involves several fields. In these theories it appears capable of producing the observed asymmetry. Conversely, with more detailed calculations this mechanism could be used to put important constraints on nonminimal extensions of the standard model.

PACS numbers: 98.80.Cq, 11.15.Ex, 12.15.Cc, 12.15.Ji

One of the most basic observational facts in cosmology is the apparent predominance of matter over antimatter in the Universe.[1] The baryon-to-photon ratio is required to be of order 10^{-9} in order for nucleosynthesis to produce the observed abundances of the light elements. Within the standard hot-big-bang theory there is no explanation for this small number: it is simply a required initial condition. In 1967 Sakharov proposed a simpler alternative, that the Universe began in a baryon-symmetric state but that particle interactions produced a net asymmetry. He pointed out that in addition to baryon-number violation this requires both *CP* violation and departure from thermal equilibrium. These conditions were met in many grand unified theories (GUT), and this was rightly regarded as a major success. However, more recent work[2] including numerical simulations[3] has established that baryon-number-violating processes occur rapidly in the electroweak theory at temperatures above the weak scale, and may in some cases interfere with any GUT-produced asymmetry.

The scenario we propose is instead based on far more conservative extensions of the standard model, which only have a slightly larger Higgs sector. Such "nonminimal" extensions occur in supersymmetric versions of the standard model which aim to explain the gauge hierarchy.[4] For an alternate scenario, see Ref. 5.

The possibility that baryon-number violation in the electroweak theory could lead to the observed baryon asymmetry has been previously discussed by Shaposhnikov[6] and McLerran.[7] While our scenario has some aspects in common with theirs (in particular, with one of Shaposhnikov's suggestions), our discussion of the mechanism is we feel clearer and more specific. The mechanism relies simply on the classical dynamics of the gauge and Higgs fields at the electroweak transition, when *CP*-violating effects from "integrating out" the fermions are included. We assume that the electroweak phase transition is at least weakly first order, as is reasonably generic in theories with several Higgs fields.[8,9] If tunneling rates are small, there is supercooling until the curvature of the potential becomes negative at the origin. This produces a departure from thermal equilibrium, and the transition proceeds by "spinodal decomposition," as the Higgs field rolls down the potential to the true vacuum. Our mechanism works during this brief rolling period. The gauge-Higgs system possesses a gauge-invariant winding number δN which we define below. This fluctuates by an amount of order unity per correlation volume. As the system relaxes toward the vacuum (in which $\delta N=0$), it can do so in two inequivalent ways. Either the Higgs field changes winding number or the gauge field does. In the latter case baryons are produced via the axial anomaly. We shall show that *CP*-violating terms in the equations of motion favor a change in the gauge-field winding number if δN has one sign, and a change in the Higgs-field winding number if δN has the other sign. The gauge-field winding changes more often in one direction, and a net baryon asymmetry is produced.

For the mechanism to work, we need to ensure that thermal sphaleron production *after* the phase transition does not wash out any baryons we generate. This requirement places an upper bound on the Higgs mass(es). Shaposhnikov[6] and, more recently, Bochkarev, Kuzmin, and Shaposhnikov,[9] have calculated this bound for both the standard model and the theory with two doublets, and it is compatible with current experimental limits (more easily so in the two-doublet case we are interested in).

Most discussions so far have focused on the gauge-field winding number, the Chern-Simons number $N_{\rm CS}$. There is, however, another important winding number in the problem, the Higgs-field winding number $N_{\rm H}$:

$$N_{\rm CS}=\frac{g^2}{16\pi^2}\int d^3x\,\epsilon^{ijk}\,{\rm Tr}(F_{ij}A_k+\tfrac{2}{3}igA_iA_jA_k)\,,$$
$$N_{\rm H}=-\frac{1}{24\pi^2}\int d^3x\,\epsilon^{ijk}\,{\rm Tr}(\partial_i\Phi^\dagger\partial_j\Phi\partial_k\Phi^\dagger\Phi)\,, \qquad (1)$$

where $A_k=A_k^a\sigma^a/2$ and $F_{ij}=F_{ij}^a\sigma^a/2$ are the SU(2) gauge field and field strength, and Φ is a unitary matrix made from the standard doublet Higgs field ϕ: $\Phi=\hat\phi^4 1+i\hat{\boldsymbol\phi}\cdot\sigma$, with the real unit four-vector $\hat\phi^a$ given by $\phi=|\phi|(\hat\phi^2+i\hat\phi^1,\hat\phi^4-i\hat\phi^3)$.

$N_{\rm CS}$ plays the central role in baryon-number violation: If it changes, the number of baryons N_B changes accordingly, through the axial anomaly,[10]

$$\delta N_B=3\delta N_{\rm CS}\,, \qquad (2)$$

where the factor 3 comes from the number of families.[11] $N_{\rm CS}$ is not a very good measure of winding number—it is only necessarily an integer if A_i is everywhere pure gauge; i.e., in the gauge-field vacuum, it is not gauge invariant, changing by an integer under "big" gauge transformations. $N_{\rm H}$ is in contrast an integer whenever it is defined, i.e., when ϕ does not vanish anywhere in the integration region, as is nearly always the case, but is also not invariant under big gauge transformations. However, the difference

$$\delta N \equiv N_{\rm CS} - N_{\rm H} \tag{3}$$

is completely gauge invariant, and is therefore a good measure of nontrivial topology in excited configurations. In fact, δN is expressible as the integral of a local gauge-invariant density, the so-called Goldstone-Wilczek density.[12] In vacuo, zero energy density requires $(\nabla + ig\mathbf{A})\phi = 0$, which forces $N_{\rm CS} = N_{\rm H} \equiv N$ and thus $\delta N = 0$.

Our discussion will focus on configurations where δN is initially nonzero. Such configurations occur in abundance near the electroweak transition and relax toward $\delta N = 0$ as their energy dissipates. During this process the *changes* $\Delta N_{\rm CS}$ and $\Delta N_{\rm H}$ are also gauge invariant. The basic idea behind our mechanism is that if the dynamics of the gauge and Higgs fields violates *CP*, positive δN and negative δN regions evolve differently: so that for one sign the vacuum will be attained through $N_{\rm H}$ changing, whereas for other sign the vacuum will be attained through $N_{\rm CS}$ changing. For equal initial numbers of positive and negative δN regions, $N_{\rm CS}$ will tend to move in one direction, producing a net baryon number.

Ignoring the gauge fields, the electroweak scalar sector is actually the simplest example of a theory with "global texture,"[13] since the vacuum is simply a three-sphere. As was argued in Ref. 13, after the symmetry-breaking transition, the Higgs field falls toward the vacuum, but in such a way as to wind around the three-sphere of order once per correlation volume. Winding configurations collapse and unwind. This process continues, the correlation length growing at the speed of light in the process.[14]

In a gauge theory, the behavior of texture is more subtle. The gauge fields can cancel gradient energy in the Higgs field, "eating" the Goldstone modes. This competes with the tendency of the Higgs field to unwind, leading to a "bifurcation" in behavior at a critical scale L_B. As an example, consider the spherically symmetric configuration $\phi = g(\mathbf{x})\phi_0$, $\phi_0 = (0,\eta)$, and $\mathbf{A}_\mu = 0$, with $g(\mathbf{x})$ a gauge group element with winding number unity: $g(\mathbf{x}) = e^{-i\chi(r)\mathbf{x}\cdot\sigma/r}$, with $\chi(0) = 0$ and $\chi(\infty) = \pi$. This has $\Delta N = -1$, and by the gauge transformation $g^{-1}(x)$ is equivalent to a configuration with $N_{\rm CS} = -1$ and $N_{\rm H} = 0$. We evolved this configuration numerically with the full classical equations of motion (details will be given in Ref. 15)

$$\begin{aligned} (D_\mu F^{\mu\nu})^a &= -i\frac{g}{2}(\phi^\dagger \sigma^a D^\nu \phi - \text{H.c.})\,, \\ D_\mu D^\mu \phi &= -V'(\phi), \quad V(\phi) = \frac{\lambda}{2}(\phi^\dagger\phi - \eta^2)^2\,, \end{aligned} \tag{4}$$

and the general spherically symmetric ansatz for the fields,

$$\begin{aligned} A_i^a &= \frac{x^a x^i}{r^2} s + \delta^{ai} t + \epsilon^{aij}\frac{x^j}{r} u\,, \\ A_0^a &= \frac{x^a}{r} w, \quad \phi = \left(\mu + iv\frac{\mathbf{x}\cdot\sigma}{r}\right)\phi_0\,, \end{aligned} \tag{5}$$

where s, t, u, w, μ, and v are functions of r and t only, and $\phi_0 = (0,\eta)$ a constant. As usual, we rescale coordinates by $x'^\mu = x^\mu/g$ to remove the dependence on the gauge coupling constant. The only scales in the equations are now the gauge boson mass $m_W = (1/\sqrt{2})g\eta$, and $m_{\rm H} = \sqrt{2\lambda}\eta$. We use Lorentz gauge, $\partial_\mu A^{\mu a} = 0$, which determines the evolution of w. We choose as initial conditions $\chi(r) = \pi \tanh(r/L)$, with L a scale determining the "size" of the initial winding configuration.

A simplification occurs if we take $\lambda = \infty$: The Higgs evolution equation (4) becomes the "gauged nonlinear σ model" equation

$$D_\mu D^\mu \phi = \frac{(D_\mu\phi)^\dagger D^\mu\phi}{\eta^2}\phi \tag{6}$$

which is solved for all time in Lorentz gauge by $\phi = \phi_0$. So we perform a big gauge transformation using $g^{-1}(\mathbf{x})$ to make ϕ constant, and then have only the massive-gauge-boson equations to evolve. The only length scale in this case is m_W^{-1}. For $L > m_W^{-1}$ the field configuration oscillates and slowly spreads out, as one would expect for a cloud of massive gauge bosons. However, for $L < m_W^{-1}$, a collapse occurs, just as in the global texture case.

If λ is large but finite, a singularity does not occur. Instead, when the size of the configuration is smaller than $m_{\rm H}^{-1}$, the gradient term in the Higgs-field equation becomes large, and pulls the Higgs field over the potential barrier. So if $L < m_W^{-1}$, the Higgs field changes winding number to match $N_{\rm CS}$, whereas if $L > m_W^{-1}$, the gauge field changes winding number to match $N_{\rm H}$. This is consistent with the sphaleron picture. To change the value of $N_{\rm CS}$ by one unit, one has to have at least the energy of a sphaleron, which is between 1.5 and 2.7 times $4\pi\sqrt{2}\eta/g$ for λ between 0 and ∞. The energy in our initial configuration is $\approx 3.7 \times 4\pi\eta^2 L$, so for $L < m_W^{-1}$ there is not enough energy to carry the gauge fields over the sphaleron barrier, while for $L > m_W^{-1}$ there is enough. The process finally occurs through the gauge fields relaxing to cancel gradient energy in the Higgs field. We have calculated the bifurcation scale L_B for general λ, and display the results in Fig. 1. As the Higgs self-coupling gets smaller the bifurcation scale gets larger.

During the transition configurations of varying scales L will be produced with both positive and negative δN. With the above *CP*-invariant classical equations, $N_{\rm CS}$ changes as frequently in the positive direction as in the negative direction. What we require to favor a particular sign of $N_{\rm CS}$ is a dynamics which violates *CP*. The terms lowest in derivatives that can occur in the effective action for the gauge and Higgs fields once the fermions are integrated out are, at temperature T,

$$\Delta L = -\int d^4x \tfrac{1}{8}\,[f(\phi,T)\delta^{ab}+g(\phi,T)^{ab}]F^a_{\mu\nu}F^b_{\rho\lambda}\epsilon^{\mu\nu\rho\lambda}\,, \quad (7)$$

where $f(\phi,T)$ is a dimensionless SU(2) scalar constructed from the Higgs fields present in the theory and $g^{ab}(\phi,T)$ a traceless symmetric tensor. Such terms appear even in the minimal electroweak theory, but are very tiny, of order 10^{-20},[6] due to Glashow-Iliopoulos-Maiani cancellation. However, in nonminimal models, with CP violation in the Higgs sector, there is no reason for them to be small. For example, in the simplest two-doublet model with the soft CP-breaking $\Phi_1^\dagger\Phi_2-\mathrm{H.c.}$ term in the potential[16] and the "minimally nonminimal" model described by Ellis *et al.*,[4] f is proportional to $g^2\theta$, where θ is the phase of $\Phi_1^\dagger\Phi_2$. As the Higgs fields roll, θ evolves as a consequence of CP-violating phases in the potential.[15] In fact, all that will matter here will be that f changes in a particular direction as the Higgs fields roll towards the vacuum. The sign of the time derivative of f determines the sign of the baryon number produced, and we shall henceforth assume it is positive.

Since (7) violates parity, positive and negative winding configurations evolve differently. Its effect is most simply seen if one considers spatially homogeneous ϕ configurations. After performing the spatial integral and integrating by parts in time, one finds (7) is equivalent to

$$\Delta L = \tfrac{1}{4}\int dt\,\frac{df}{dt}N_{\mathrm{CS}}\,, \quad (8)$$

which is simply a linear potential for N_{CS}. During the phase transition, when df/dt is positive, (7) provides a force term driving N_{CS} positive.

Ignoring spatial gradients in f, the gauge-field equations become

$$D_\mu F^{\mu i} = \zeta g\eta B^i - i\frac{g}{2}(\phi^\dagger\sigma^a D^i\phi - \mathrm{H.c.})\,,\quad \zeta = \frac{1}{g\eta}\frac{df}{dt}\,. \quad (9)$$

Equation (9) has been investigated numerically by Grigoriev, Rubakov, and Shaposhnikov[17] for the 1+1 Abelian Higgs model, and by Ambjorn and coworkers[3,18] in the course of investigating whether a Chern-Simons condensate could develop in the electroweak theory at high temperatures—the extra term was added as a source, to probe the response of the system (this issue is still unresolved, but we assume there is no condensate in our work). In our case it is present intrinsically in the classical dynamics of the theory.

We have evolved Eqs. (4) numerically with the addition of the extra parity-violating term in (9). We treated ζ as a constant for these purposes, ignoring the "backreaction" on the rolling of the Higgs field. This is reasonable for small ζ, but we intend to evolve the full equations in forthcoming work.[15] A typical example is shown in Fig. 2, for $\lambda=1000$. For ζ equal to zero (dashed lines) the configuration collapses, leading to an unwinding of the Higgs field. The collapse is even faster if ζ is negative. But for ζ positive (solid lines), the configuration "bounces," and the Higgs field does not change winding number. Therefore N_{CS} must change. In this case, exactly three baryons are produced. Of

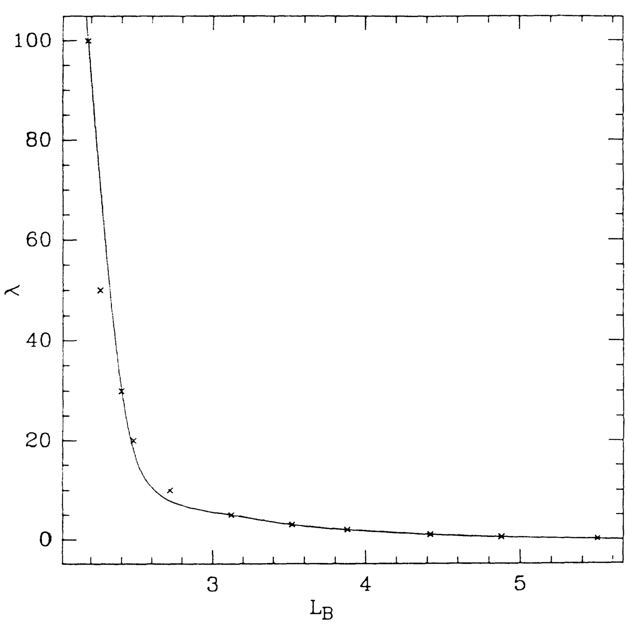

FIG. 1. The bifurcation scale L_B between collapse and relaxation for spherical configurations with $\delta N=1$, as described in the text. It is plotted against the Higgs self-coupling λ/g^2: For scales to the left of the line the Higgs field unwinds, but for scales to the right of the line the gauge fields wind up. Length units are such that $g\eta=1$.

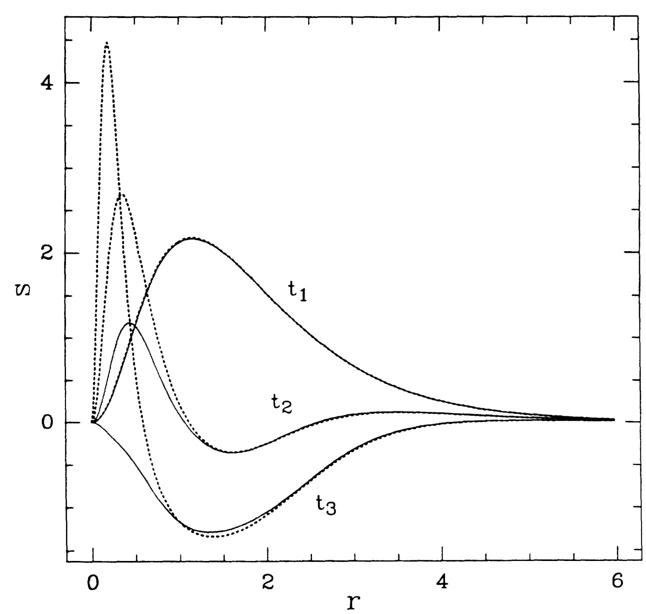

FIG. 2. The function s, defined in (5), plotted vs radius for several times, starting at an early time t_1, and ending at t_3. The dashed lines show the evolution for $\zeta=0$, where the configuration collapses, causing the Higgs field to unwind. The solid lines show the evolution for $\zeta=0.2$, where the gauge-field configuration "bounces," and the Higgs field does not unwind.

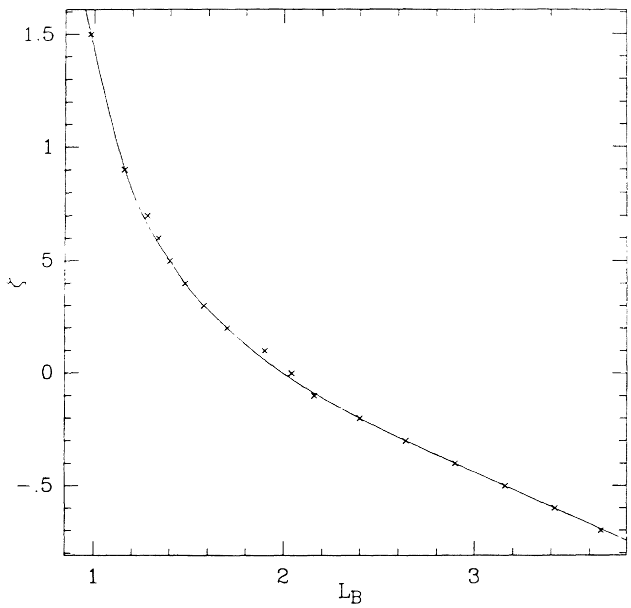

FIG. 3. The bifurcation scale L_B plotted as a function of ζ. As in Fig. 1, for scales to the left of the line the Higgs field unwinds, but for scales to the right the gauge field winds up. In any theory ζ is fixed, but the opposite sign of ζ describes configurations of opposite winding number δN. The asymmetry on this diagram between positive and negative ζ results in a net asymmetry in the change in N_{CS} during the phase transition and therefore a net baryon number.

course, changing the sign of ζ is exactly the same as considering a configuration of the opposite winding number, so we have shown that in this case N_{CS} would increase more often than decrease. The dependence of the bifurcation scale L_B on the value of ζ is shown in Fig. 3. The important region is near $\zeta=0$, where the dependence is clearly linear, and the slope of order unity. The results are shown for large $\lambda=1000$. We also calculated the slope for $\lambda=1$ and again found it to be of order unity.[15] This means that in the "realistic" case, where ζ is a small number, we would expect the difference between collapse and expansion to be seen for configurations whose scale was within $\pm\zeta L_B$ of the bifurcation scale L_B. So for small ζ only a small fraction of winding configurations would behave differently for positive and negative N_{CS}.

It is clear that for small ζ the net baryon-to-photon ratio is linear in ζ, and in fact of order ζ for coupling constants of order unity. There is some probability per correlation volume of having a winding configuration of scale L. This probability distribution, being thermal in origin, will not be sharply peaked, and will be largest around the inverse weak scale m_W^{-1}. The probability that L lies within ΔL_B of the "bifurcation scale," $L_B \approx m_W^{-1}$, is then of order Δ since the probability function is of order unity there. As we have seen, for $\Delta = \pm\zeta$, a net baryon number is produced. Thus a fraction of order ζ correlation volumes will produce baryons. Since there is roughly one photon per correlation volume, the final baryon-to-photon ratio is simply of order ζ. It is clear that it is easy to obtain an asymmetry of the required magnitude for reasonable couplings, and a preliminary survey of the existing constraints on flavor-changing neutral-Higgs exchange indicates these are not a problem,[19] for the rather small values of ζ that we require.

We thank E. d'Hoker, M. Shaposhnikov, P. Steinhardt, and D. Spergel for discussions. N.T. acknowledges the support of NSF Contract No. PHY80-19754 and the Alfred Sloan Foundation.

[1] For a review, see G. Steigman, Annu. Rev. Astron. Astrophys. **14**, 339 (1976).

[2] F. Klinkhamer and N. Manton, Phys. Rev. D **30**, 2212 (1984); V. Kuzmin, V. Rubakov, and M. Shaposhnikov, Phys. Lett. **155B**, 36 (1985); P. Arnold and L. McLerran, Phys. Rev. D **37**, 1020 (1988); for a clear review of these arguments, see M. Dine, O. Lechtenfeld, B. Sakita, W. Fischler, and J. Polchinski, City College of New York, Report No. CCNY-HEP-89/18, 1990 (to be published).

[3] J. Ambjorn, T. Askgaard, H. Porter, and M. Shaposhnikov, Niels Bohr Institute Report No. HE-90-18 (to be published).

[4] J. Ellis, J. F. Gunion, H. E. Haber, L. Roszkowski, and F. Zwirner, Phys. Rev. D **39**, 844 (1989).

[5] A. Cohen, D. Kaplan, and A. Nelson, Institute for Theoretical Physics Report No. NSF-ITP-90-85 (to be published).

[6] M. Shaposhnikov, Nucl. Phys. **B287**, 757 (1987); **B299**, 797 (1988).

[7] L. McLerran, Phys. Rev. Lett. **62**, 1075 (1989).

[8] M. Sher, Phys. Rep. **179**, 273 (1989).

[9] A. Bochkarev, S. Kuzmin, and M. Shaposhnikov, Niels Bohr Institute Report No. HE-90-08 (to be published).

[10] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976).

[11] We use the approximation $\sin(\theta_W)=0$ throughout, ignoring the weak hypercharge gauge fields.

[12] J. Goldstone and F. Wilczek, Phys. Rev. Lett. **47**, 986 (1981).

[13] N. Turok, Phys. Rev. Lett. **63**, 2625 (1989).

[14] D. Spergel, N. Turok, W. Press, and B. Ryden, Princeton University Report No. PUP-TH-90-1182, 1990 (to be published).

[15] N. Turok and J. Zadrozny (to be published).

[16] For example, see, S. Weinberg, University of Texas Report No. UTTG-04-90 (to be published).

[17] D. Grigoriev, V. Rubakov, and M. Shaposhnikov, Phys. Lett. B **216**, 172 (1989).

[18] J. Ambjorn, M. Laursen, and M. Shaposhnikov, Phys. Lett. B **197**, 49 (1987).

[19] J. Liu and L. Wolfenstein, Nucl. Phys. **B289**, 1 (1987); V. Barger, J. Hewett, and R. Phillips, Phys. Rev. D **41**, 3421 (1990).