

# Origin of the matter-antimatter asymmetry

Michael Dine

*Santa Cruz Institute for Particle Physics, Santa Cruz, California 95064, USA*

Alexander Kusenko

*Department of Physics and Astronomy, University of California, Los Angeles, California 90095-1547, USA*

*and RIKEN BNL Research Center, Brookhaven National Laboratory, Upton, New York 11973, USA*

(Published 16 December 2003)

Although the origin of matter-antimatter asymmetry remains unknown, continuing advances in theory and improved experimental limits have ruled out some scenarios for baryogenesis, for example, sphaleron baryogenesis at the electroweak phase transition in the Standard Model. At the same time, the success of cosmological inflation and the prospects for discovering supersymmetry at the Large Hadron Collider have put some other models in sharper focus. We review the current state of our understanding of baryogenesis with emphasis on those scenarios that we consider most plausible.

## CONTENTS

I. Introduction	1
II. A Baryogenesis Roadmap	4
A. A cosmology overview	4
B. Planck-scale baryogenesis	5
C. GUT baryogenesis	5
D. Electroweak baryon-number violation	7
E. Electroweak baryogenesis	9
1. Electroweak baryogenesis in the Standard Model	9
2. Supersymmetry, a short introduction	11
3. Baryogenesis in the minimal and next-to-minimal supersymmetric Standard Models	13
4. Nonthermal electroweak baryogenesis at preheating	13
F. Leptogenesis	15
G. Baryogenesis through coherent scalar fields	17
III. Affleck-Dine Baryogenesis	17
A. Arguments for coherent production of baryon number	17
B. Baryogenesis through a coherent scalar field	17
C. Flat directions and baryogenesis	19
D. Evolution of the condensate	20
1. Supersymmetry breaking in the early universe	20
2. Appearance of baryon number	20
E. The fate of the condensate	21
F. Inhomogeneities and the condensate	22
1. Stability and fragmentation	22
2. Lumps of scalar condensate: $Q$ balls	23
3. Affleck-Dine $Q$ balls	25
4. Dark matter in the form of stable $B$ balls	25
5. Dark matter from unstable $B$ balls	26
IV. Conclusions	27
Acknowledgments	28
References	28

## I. INTRODUCTION

When we observe the universe, the most obvious and easily studied objects are stars and gas, made up of pro-

tons, neutrons, and electrons. Astrophysicists speak of the density of protons and neutrons, which constitute the bulk of the mass of this matter, as the baryon content of the universe.

But we know that there is much more to the universe than baryons. By indirect means, astronomers have established that approximately 1/3 of the energy density of the universe is in the form of some nonbaryonic matter, referred to as *dark matter*, while roughly 2/3 is in a form with negative pressure, perhaps due to a cosmological constant (Peebles and Ratra, 2003). The baryons make up a mere 5% of the total energy density of the universe.

Another even more striking measure of the smallness of the baryon density is provided by the ratio of baryons to photons in the Cosmic Microwave Background Radiation (CMBR). Big-bang nucleosynthesis gives a good measure of the baryon density; this measurement is well supported by recent measurements of the fluctuations of the cosmic microwave radiation background. As a result, the ratio of baryons to photons is now known to about 5% (Bennett *et al.*, 2003):

$$\frac{n_B}{n_\gamma} = (6.1^{+0.3}_{-0.2}) \times 10^{-10}, \quad (1)$$

where  $n_B$  and  $n_\gamma$  are the densities of baryons and photons, respectively. There is good evidence that there are no large regions of antimatter at any but cosmic distance scales (Cohen, De Rujula, and Glashow, 1998), although some small domains of antimatter in the matter-dominated universe are not ruled out by observations (Dolgov and Silk, 1993; Belotsky *et al.*, 1998; Khlopov *et al.*, 2000).

It was A. Sakharov who first suggested that the baryon density might not represent some sort of initial condition, but might be understandable in terms of microphysical laws (Sakharov, 1967). He listed three ingredients to such an understanding:

- (1)  $B$  violation: Baryon-number violation must occur in the fundamental laws. At very early times, if

baryon-number-violating interactions were in equilibrium, then the universe can be said to have “started” with zero baryon number. Starting with zero baryon number, baryon-number-violating interactions are obviously necessary if the universe is to end up with a nonzero asymmetry. As we will see, apart from the philosophical appeal of these ideas, the success of inflationary theory suggests that, shortly after the big bang, the baryon number was essentially zero.

- (2) *CP* violation: If *CP* (the product of charge conjugation and parity) is conserved, every reaction which produces a particle will be accompanied by a reaction which produces its antiparticle at precisely the same rate, so no baryon number can be generated.
- (3) An arrow of time (departure from thermal equilibrium): The universe, for much of its history, was very nearly in thermal equilibrium. The spectrum of the CMBR is the most perfect blackbody spectrum measured in nature. So the universe was certainly in thermal equilibrium  $10^5$  years after the big bang. The success of the theory of big-bang nucleosynthesis (BBN) provides strong evidence that the universe was in equilibrium two to three minutes after the big bang. But if, through its early history, the universe was in thermal equilibrium, then even *B*- and *CP*-violating interactions could not produce a net asymmetry. One way to understand this is to recall that the *CPT* theorem assures strict equality of particle and antiparticle masses, so at thermal equilibrium, the densities of particles and antiparticles are equal. More precisely, since *B* is odd under *CPT*, its thermal average vanishes in an equilibrium situation. This can be generalized by saying that the universe must have an arrow of time.

One of the great successes of the Standard Model is that it explains why baryon and lepton number are conserved, to a very good approximation. To understand what this means, consider first the modern understanding of Maxwell’s equations. A quantum field theory is specified by its field content and by a Lagrangian density. In the Lagrangian, one distinguishes renormalizable and nonrenormalizable terms. Renormalizable terms have coefficients with mass dimension greater than zero; nonrenormalizable terms have coefficients (couplings) with mass dimension less than zero. For example, in quantum electrodynamics, the electron mass has dimension 1, while the charge of the electron is dimensionless (throughout we use conventions where  $\hbar$  and  $c$  are dimensionless), so these are renormalizable. In fact, requiring Lorentz invariance, gauge invariance, and renormalizability leaves only one possibility for the Lagrangian of electrodynamics: the Maxwell Lagrangian, whose variation yields Maxwell’s equations. One can, consistent with these symmetry principles, write down an infinite number of possible nonrenormalizable terms, which would yield nonlinear modifications of Maxwell’s equations. There is nothing wrong with these, but each is characterized by a mass, or inverse length

scale  $M$ . So the size of nonlinear corrections at wavelength  $\lambda$  is of order  $(\lambda M)^{-n}$  for some integer  $n$ .  $M$  represents some scale at which the laws of electricity and magnetism might be significantly modified. Such corrections actually exist, and are for most purposes quite small.

Similarly, in the Standard Model, at the level of renormalizable terms, there are simply no interactions one can write which violate either baryon number or the conservation of the separate lepton numbers (electron, muon, and tau number). It is possible to add dimension-5 operators (having a scale  $1/M$ ) which violate lepton number, and dimension-6 operators (having a scale  $1/M^2$ ), which violate baryon number. Again, these nonrenormalizable terms must be associated with a mass scale of some new baryon- and lepton-violating physics. The dimension-five lepton-number-violating operators would give rise to a mass for the neutrinos. The recent discovery of neutrino mass probably amounts to a measurement of some of these lepton-number-violating operators. The energy scale of new physics associated with these operators cannot yet be determined, but theoretical arguments suggest a range of possibilities, between about  $10^{11}$  and  $10^{16}$  GeV.

The question, then, is what might be the scale  $M_B$  associated with baryon-number violation. At the very least, one expects quantum effects in gravity to violate all global quantum numbers (e.g., black holes swallow up any quantum numbers not connected with long-range fields like the photon and graviton), so  $M_B \leq M_p$ , where  $M_p = \sqrt{G_N^{-1}} \approx 10^{19}$  GeV is the Planck mass, defined in terms of the Newtonian gravitational constant  $G_N$ . The leading operators of this kind, if they have Planck-mass coefficients, would lead to a proton lifetime of order  $10^{34}$  years or so.

If quantum-gravitational effects were the only source of baryon-number violation, we could imagine that the baryon asymmetry of the universe was produced when the temperature of the universe was of order the Planck energy ( $10^{32}$  K). Some complex processes associated with very energetic configurations would violate baryon number. These need not be in thermal equilibrium (indeed, in a theory of gravity, the notion of equilibrium at such a high temperature almost certainly does not make sense). The expansion of the universe at nearly the moment of the big bang would provide an arrow of time. *CP* is violated already at relatively low energies in the Standard Model through the Kobayashi-Maskawa (KM) mechanism, so there is no reason to believe that it is conserved in very-high-energy processes. So we could answer Sakharov by saying that the magnitude of the baryon number is the result of some very complicated, extremely high-energy process, to which we will never have experimental access. It might be, in effect, an initial condition.

There are good reasons to believe that this pessimistic picture is not the correct one. First, we are trying to understand a small, dimensionless number. But in this Planck-scale baryogenesis picture, it is not clear how such a small dimensionless number might arise. Second,

there is growing evidence that the universe underwent a period of inflation early in its history. During this period, the universe expanded rapidly by an enormous factor (at least  $e^{60}$ ). Inflation is likely to have taken place well below the scale of quantum gravity, and thus any baryon number produced in the Planck era was diluted to a totally negligible level. Third, there are a variety of proposals for new physics—as well as some experimental evidence—which suggests that baryon- and lepton-number-violating interactions might have been important at scales well below the Planck scale. So there is some reason for optimism that we might be able to compute the observed baryon number density from some underlying framework, for which we could provide both direct (i.e., astrophysical or cosmological) and/or indirect (discovery of new particles and interactions) evidence.

Several mechanisms have been proposed to understand the baryon asymmetry:

- (1) Planck-scale baryogenesis: this is the idea, discussed above, that Planck-scale phenomena are responsible for the asymmetry. We have already advanced arguments (essentially cosmological) that this is unlikely; we will elaborate on them in the next section.
- (2) Baryogenesis in grand unified theories (GUT baryogenesis): this, the earliest well-motivated scenario for the origin of the asymmetry, will be discussed more thoroughly in the next section. Grand unified theories unify the gauge interactions of the strong, weak, and electromagnetic interactions in a single gauge group. They invariably violate baryon number, and they have heavy particles, with mass of order  $M_{\text{GUT}} \approx 10^{16}$  GeV, whose decays can provide a departure from equilibrium. The main objections to this possibility come from issues associated with inflation. While there does not exist a compelling microphysical model for inflation, in most models, the temperature of the universe after reheating is well below  $M_{\text{GUT}}$ . But even if it were very large, there would be another problem. Successful unification requires supersymmetry, a hypothetical symmetry between fermions and bosons, which will play an important role in this review. Supersymmetry implies that the graviton has a spin-3/2 partner, called the *gravitino*. In most models for supersymmetry breaking, these particles have masses  $m_{3/2}$  of order TeV, and are very long lived. Even though these particles are weakly interacting, too many gravitinos are produced, unless the reheating temperature is well below the unification scale (Kallosh *et al.*, 2000).
- (3) Electroweak baryogenesis: as we will explain, the Standard Model satisfies all of the conditions for baryogenesis. This is somewhat surprising, since at low temperatures the model seems to preserve baryon number. It turns out that baryon and lepton number are badly violated at very high temperatures. The departure from thermal equilibrium can arise at the electroweak phase transition—a transition between the familiar state in which the  $W$  and

$Z$  bosons are massive and one in which they are massless. This transition can be first order, providing an arrow of time. However, as we will explain below, any baryon asymmetry produced is far too small to account for observations. In certain extensions of the Standard Model, it is possible to obtain an adequate asymmetry, but in most cases the allowed region of parameter space is very small. This is true, for example, of the minimal supersymmetric Standard Model (MSSM). Experiments will soon either discover supersymmetry in this region, or close off this tiny segment of parameter space. This scenario has been reviewed by Cline (2000).

- (4) Leptogenesis: The observation that the weak interactions will convert some lepton number to baryon number means that if one produces a large lepton number at some stage, this will be processed into a net baryon and lepton number. The observation of neutrino masses makes this idea highly plausible. Many but not all of the relevant parameters can be directly measured.
- (5) Production by coherent motion of scalar fields (Affleck and Dine, 1985): This mechanism, which can be highly efficient, might well be operative if nature is supersymmetric. In this case, as we will explain in much greater detail, the ordinary quarks and leptons are accompanied by scalar quarks and leptons. It has been widely conjectured that supersymmetry may be discovered in the next generation of high energy accelerators. So again, one might hope to uncover the basic underlying physics, and measure some (but it will turn out not all) of the relevant parameters. In nonsupersymmetric theories, it is believed that scalar fields with the requisite properties (low mass, very flat potentials) are unnatural. This supersymmetric baryogenesis mechanism will be the main focus of this review. See also the review by Enqvist *et al.* (2002).

In this review we will survey these mechanisms, and explain in more detail why the last two are by far the most plausible. The question then becomes: can we eventually establish that one or the other is correct? In order to establish or rule out particular models for the origin of the matter-antimatter asymmetry, we would hope to bring to bear both astrophysical/cosmological observations and particle-physics experiments, as well as theoretical arguments. Ideally, we would some day be in the position of measuring all of the parameters relevant to the asymmetry, and calculating the asymmetry in much the same way that one presently calculates the light element abundances. One question we will ask is: how close can we come to this ideal situation?

In the next section, after a very brief review of the standard cosmology, we present our survey of these mechanisms, both explaining how they work and discussing their theoretical plausibility. Both electroweak baryogenesis and leptogenesis rely on the existence of processes within the Standard Model which violate

baryon and lepton number at high temperatures, and we include a brief explanation of these phenomena.

We then turn to a more detailed discussion of coherent production of baryons or leptons, the Affleck-Dine mechanism. This mechanism is potentially extremely efficient; it can also operate relatively late in the history of the universe. As a result, it can potentially resolve a number of cosmological puzzles. The Affleck-Dine mechanism presupposes low-energy supersymmetry. Supersymmetry (sometimes called SUSY for short) is a hypothetical extension of Poincaré invariance, a symmetry which would relate bosons to fermions. If correct, it predicts that for every boson of the Standard Model, there is a fermion, and vice versa. It is believed that the masses of the new particles should be about a TeV. As supersymmetry will play an important role in much of our discussion, a brief introduction to supersymmetry will be provided in the next section. The supersymmetry hypothesis will be tested over the next decade by the Tevatron and the Large Hadron Collider at CERN. Interestingly, most other proposals for baryogenesis invoke supersymmetry in some way. These include electroweak baryogenesis and most detailed models for leptogenesis.

## II. A BARYOGENESIS ROADMAP

### A. A cosmology overview

Our knowledge of the big bang rests on a few key observational elements. First, there is the Hubble expansion of the universe. This allows us to follow the evolution of the universe to a few billion years after the big bang. Second, there is the CMBR. This is a relic of the time, about  $10^5$  years after the big bang, when the temperature dropped to a fraction of an electron volt and electrons and nuclei joined to form neutral atoms. Third, there is the abundance of the light elements. This is a relic of the moment of neutrino decoupling, when the temperature was about 1 MeV. As we have noted, theory and observation are now in good agreement, yielding the baryon-to-photon ratio given by Eq. (1). Finally, there are the fluctuations in the temperature of the microwave background, measured recently on angular scales below one degree by BOOMERANG (de Bernardis *et al.*, 2000; Netterfield *et al.*, 2002), MAXIMA (Hanany *et al.*, 2000), DASI (Pryke *et al.*, 2002), and WMAP (Bennett *et al.*, 2003). These fluctuations are probably a relic of the era of inflation (discussed in more detail below). The baryon density can be inferred independently from the CMBR data and from the BBN determination of the baryon density based on the measurements of the primordial deuterium abundance (Burles, Nollett, and Turner, 2001a, 2001b; Kirkman *et al.*, 2003). The agreement is spectacular:  $\Omega_B h^2 = 0.0214 \pm 0.002$  based on BBN (Kirkman *et al.*, 2003), while the CMBR anisotropy measurements yield  $\Omega_B h^2 = 0.0224 \pm 0.0009$  (Bennett *et al.*, 2003). Here  $\Omega_B$  is the fraction of critical density contributed by baryons, and  $h$  is the Hubble constant in units of  $100 \text{ km sec}^{-1} \text{ Mpc}^{-1}$ .

The first and perhaps most striking lesson of the measurements of the CMBR is that the universe, on large scales, is extremely homogeneous and isotropic. As a result, it can be described by a Robertson-Walker metric:

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

$R(t)$  is known as the scale factor; the *Hubble rate* is  $H = \dot{R}/R$ . The Hubble rate is essentially the inverse of the time; in the radiation-dominated era,  $H = 1/2t$ ; in the matter-dominated era,  $H = 2/3t$ . Here the constant  $k$  is the curvature of the universe. It is puzzling that the universe should be homogeneous and isotropic to such a high degree. If one runs the clock backward, one finds that vast regions of the universe which have only recently been in causal contact have essentially the same temperature.

Inflation provides an explanation for this and other puzzles (Kolb and Turner, 1990; Linde, 1990). The basic idea (Guth, 1981; Albrecht and Steinhardt, 1982; Linde, 1982) is that for a brief period,  $R(t)$  grew extremely rapidly, typically exponentially. This has several effects:

- The observed universe grew from a microscopically small region, explaining homogeneity and isotropy.
- $k=0$ , i.e., the universe is spatially flat. This is now well verified by observations.
- Small fluctuations in the metric and the field during inflation explain the observed small (one part in  $10^{-5}$ ) variation in the temperature of the CMBR. Detailed features of this structure, in agreement with the inflationary theory, have now been observed. These fluctuations provide the seeds for formation of the observed structure in the universe.
- Inflation also explains the absence from the universe of objects such as magnetic monopoles expected in many particle physics theories.

While it is probably fair to say that no compelling microscopic theory of inflation yet exists, it is very successful as a phenomenological theory. Most pictures of inflation invoke the dynamics of a scalar field in a crucial way. This scalar field must have very special properties. Typically, for example, the curvature of its potential must be very small. The most plausible theories which achieve this invoke supersymmetry in a significant way. Supersymmetry, a hypothetical symmetry between fermions and bosons, will be discussed at greater length later in this article. It has been widely considered as a possible solution to many puzzles in particle physics. Most importantly for inflation, supersymmetry is a theoretical framework which naturally gives rise to scalars with very flat potentials. It also gives rise to stable particles with just the right properties to constitute the dark matter. There are difficulties as well. One is associated with the fermionic partner of the graviton, the gravitino. In many models, this particle is very long-lived ( $\tau$

$>10^6$  sec), and can spoil conventional big-bang cosmology if too many are produced.

There is not space here to review the subject of inflation (Kolb and Turner, 1990; Linde, 1990). Instead, we will give a “narrative” of a possible history of the universe, which will be useful to orient our discussion:

- Before  $t \approx 10^{-25}$  sec, the universe was very inhomogeneous, with an extremely large energy density. At  $t \approx 10^{-25}$ , inflation began in a small patch. This was associated with a scalar field, called the *inflaton*, which moved slowly toward the minimum of its potential.
- The scale of the inflaton potential was of order  $10^{60}$  GeV<sup>4</sup>, give or take a few orders of magnitude.
- During inflation, the scale factor increased by an enormous factor. Any conserved or approximately conserved quantities, such as monopole number or baryon number, were reduced by at least a factor of  $10^{60}$  in this process.
- Inflation ended as the inflaton approached the minimum of its potential. At this point, decays of the inflaton lead to reheating of the universe to a high temperature. Depending on the detailed microscopic picture, there are constraints on the reheating temperature. If nature is supersymmetric, there is a danger of producing too many gravitinos and other long-lived particles. Typically, this constrains the reheating temperature to be below  $10^9$  GeV. Even without supersymmetry, detailed inflationary models have difficulty producing high reheating temperatures without fine tuning.
- The baryon asymmetry is generated some time after the era of inflation. Any upper limit on the reheating temperature constrains the possible mechanisms for baryogenesis.

## B. Planck-scale baryogenesis

It is generally believed that a quantum theory of gravity cannot preserve any global quantum numbers. For example, in the collapse of a star to form a black hole, the baryon number of the star is lost; black holes are completely characterized by their mass, charge, and angular momentum. Virtual processes involving black holes, then, would also be expected to violate baryon number.

In string theory, the only consistent quantum theory of gravity we know, these prejudices are born out. There are no conserved global symmetries in string theory (Banks *et al.*, 1988). While we cannot reliably extract detailed predictions from quantum gravity for baryon-number violation, we might expect that it will be described at low energies by operators which appear in an effective-field theory. The leading operators permitted by the symmetries of the Standard Model which violate baryon number carry dimension 6. An example of an interaction term in the Lagrangian is

$$\mathcal{L}_B = \frac{1}{M^2} \bar{e} \bar{d}^* \bar{d}^* \bar{d}^*. \quad (3)$$

In this equation and those which follow, the various fermion fields,  $d$ ,  $\bar{d}$ ,  $e$ ,  $\bar{e}$ ,  $\nu$ , etc., are spinors of left-handed chirality.  $\bar{d}$  contains the creation operator for the right-handed  $d$  quark;  $d^\dagger$  for the left-handed anti- $d$  quark. The other two  $d$ -quark states are created by  $d$  and  $d^\dagger$ . We have indicated that its coefficient has dimensions of inverse mass squared because the operator is of dimension 6. This is analogous to the effective interaction in the Fermi theory of weak interactions. If quantum gravity is responsible for this term, we might expect its coefficient to be of order  $1/M_p^2$ .

Because of this very tiny coefficient, these effects could be important only at extremely early times in the universe, when, for example,  $H \sim M_p$ . It is probably very difficult to analyze baryon production in this era. It is certainly unclear in such a picture where the small number in Eq. (1) might come from. But even if the baryon number was produced in this era, it was completely diluted in the subsequent period of inflation. So gravitational baryogenesis seems unlikely to be the source of the observed matter-antimatter asymmetry.

## C. GUT baryogenesis

The earliest well-motivated scenarios for implementing Sakharov’s ideas within a detailed microscopic theory were provided by grand unified theories (GUT’s) (Kolb and Turner, 1990). In the Standard Model, the strong, weak, and electromagnetic interactions are described by non-Abelian gauge theories based on the groups SU(3), SU(2), and U(1). Grand unification posits that the underlying theory is a gauge theory with a simple group, and that this gauge symmetry is broken down to the group of the Standard Model at some very-high-energy scale. This hypothesis immediately provides an explanation of the quantization of electric charge. It predicts that, at very high energies, the strong, weak, and electromagnetic couplings (suitably normalized) should have equal strength. And most important, from the point of view of this article, it predicts violation of baryon and lepton numbers.

If nature is not supersymmetric, the GUT hypothesis fails. One can use the renormalization group to determine the values of the three gauge couplings as a function of energy, starting with their measured values. One finds that they do not meet at a point, i.e., there is no scale where the couplings are equal. Alternatively, one can take the best measured couplings, the SU(2) and U(1) couplings, and use the GUT hypothesis to predict the value of the strong coupling. The resulting prediction is off by 12 standard deviations (Hagiwara, 2002). But if one assumes that nature is supersymmetric, and that the new particles predicted by supersymmetry all have masses equal to 1 TeV, one obtains unification, within  $3\sigma$ . The scale of unification turns out to be  $M_{\text{GUT}} \approx 2 \times 10^{16}$  GeV. The unified coupling  $\alpha_{\text{GUT}} \approx g^2/4\pi$  is approximately 1/25. Relaxing the assumption that the new particles are degenerate, or assuming that there are additional, so-called threshold corrections to

the couplings at the GUT scale (within  $\sim 4\%$  of  $M_{\text{GUT}}$ ), one can attain complete agreement.

This value of  $M_{\text{GUT}}$  is quite interesting. It is sufficiently below the Planck scale that one might hope to analyze these theories without worrying about quantum gravity corrections. Moreover, it leads to proton decay at a rate which may be accessible to current proton decay experiments. In fact, the simplest SUSY GUT based on the gauge group SU(5) is almost completely ruled out by the recent Super-Kamiokande bounds (Murayama and Pierce, 2002). However, there are many other models. For example, nonminimal SU(5) or SO(10) SUSY GUT's may have a proton lifetime about a factor of 5 above the present experimental limit (Altarelli, Feruglio, and Masina, 2000; Babu, Pati, and Wilczek, 2000; Dermisek, Mafi, and Raby, 2001; Bajc, Perez, and Senjanovic, 2002). Witten has recently advocated an approach to GUT model building (Friedmann and Witten, 2002; Witten, 2002) which resolves certain problems with these models, and in which proton decay might be difficult to see even in large detectors which are being considered for the future.

GUT's provide a framework which satisfies all three of Sakharov's conditions. First, baryon-number violation is a hallmark of these theories: they typically contain gauge bosons and other fields which mediate  $B$ -violating interactions such as proton decay. Second,  $CP$  violation is inevitable; any model necessarily contains at least the KM mechanism for violating  $CP$ , and typically there are many new couplings which can violate  $CP$ . Third, departure from equilibrium is associated with the dynamics of the massive,  $B$ -violating fields. Typically one assumes that these fields are in equilibrium at temperatures well above the grand unification scale. As the temperature becomes comparable to their masses, the production rates of these particles fall below their rates of decay. Careful calculations in these models often lead to baryon densities compatible with what we observe.

We can illustrate the basic ideas with the simplest GUT model, due to Georgi and Glashow (1974). Here the unifying gauge group is SU(5). The model we will discuss is not supersymmetric, but it illustrates the important features of GUT baryon number production. A single generation of the Standard Model (e.g., electron, electron neutrino,  $u$  quark, and  $d$  quark) can be embedded in the  $\bar{5}$  and 10 representations of SU(5). It is conventional, and convenient, to treat all quarks and leptons as left-handed fields. So in a single generation of quarks and leptons, one has the quark doublet  $Q$ , the singlet  $\bar{u}$  and  $\bar{d}$  antiquarks (their antiparticles are the right-handed quarks), and the lepton doublet,

$$L = \begin{pmatrix} e \\ \nu \end{pmatrix}.$$

Then it is natural to identify the fields in the  $\bar{5}$  as

$$\bar{5}_i = \begin{pmatrix} \bar{d} \\ \bar{d} \\ \bar{d} \\ e \\ \nu \end{pmatrix}. \quad (4)$$

The generators of SU(3) of color are identified as

$$T^a = \begin{pmatrix} \frac{\lambda^a}{2} & 0 \\ 0 & 0 \end{pmatrix}, \quad (5)$$

where  $\lambda^a$  are the eight Gell-Mann matrices, while those of SU(2) are identified with

$$T^{i'} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^i \end{pmatrix}, \quad (6)$$

where the  $\sigma^i$  are the three Pauli matrices and  $i' = i + 8$ . The U(1) generator is the diagonal matrix

$$Y' = \frac{1}{\sqrt{60}} \begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & -3 & \\ & & & & -3 \end{pmatrix}. \quad (7)$$

Here the coefficient has been chosen so that the normalization is the same as that of the SU(3) and SU(2) matrices [ $\text{Tr}(T^a T^b) = \delta_{ab}$ ]. The corresponding gauge boson couples with the same coupling constant as the gluons and  $W$  and  $Z$  bosons. This statement holds at  $M_{\text{GUT}}$ ; at lower energies, there are significant radiative corrections (which in the supersymmetric case reproduce the observed low-energy gauge couplings).

In the Standard Model, the hypercharge  $Y$  is related to the ordinary electric charge,  $Q$ , and the isospin generator,  $T_3$ , by  $Q = T_3 + Y/2$ . So one sees that electric charge is quantized, and that  $Y = \sqrt{3/40} Y'$ . Since  $Y$  couples with the same strength as the SU(2) generators, this gives a prediction of the U(1) coupling of the Standard Model, and correspondingly of the Weinberg angle  $\theta_W$ ,  $\sin^2(\theta_W) = 3/8$ . This prediction is subject to radiative corrections which, assuming supersymmetry, bring it within experimental errors of the measured value. In a single generation, the remaining fields lie in the 10 representation. The 10 transforms as the antisymmetric product of two  $\bar{5}$ 's. It has the form

$$10 = \begin{pmatrix} 0 & \bar{u}_2 & -\bar{u}_1 & Q_1^1 & Q_1^2 \\ -\bar{u}_2 & 0 & \bar{u}_3 & Q_2^1 & Q_2^2 \\ \bar{u}_1 & -\bar{u}_3 & 0 & Q_3^1 & Q_3^2 \\ -Q_1^1 & -Q_2^1 & -Q_3^1 & 0 & \bar{e} \\ -Q_1^2 & -Q_2^2 & -Q_3^2 & -\bar{e} & 0 \end{pmatrix}, \quad (8)$$

where  $Q_i^1 = u_i$  and  $Q_i^2 = d_i$  are left-handed quark fields, which transform as doublets under SU(2).

SU(5) is not a manifest symmetry of nature. It can be broken by the expectation value of a scalar field  $\Phi$  in the adjoint representation having the same form as  $Y$ :

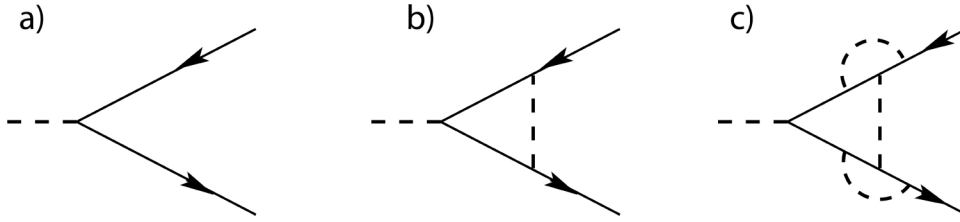


FIG. 1. Interference between the tree-level (a) and one-loop (b) diagrams with complex Yukawa couplings can provide the requisite source of  $CP$  violation for GUT baryogenesis. In viable models, to avoid the unwanted cancellations, one must often assume that the two scalars are different or go to higher loops, as in (c) (Barr, Segrè, and Weldon, 1979; Kolb and Turner, 1990).

$$\langle \Phi \rangle = v \begin{pmatrix} 2 & & & & \\ & 2 & & & \\ & & 2 & & \\ & & & -3 & \\ & & & & -3 \end{pmatrix}, \quad (9)$$

where  $v$  is a constant. The unbroken generators are those which commute with  $\Phi$ , i.e., precisely the generators of  $SU(3) \times SU(2) \times U(1)$  above.

The vector bosons which correspond to the broken generators gain mass of order  $gv$ . We will refer to the corresponding gauge bosons as  $X$ ; they are associated with generators which do not commute with  $\langle \Phi \rangle$ , such as

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (10)$$

They carry color and electroweak quantum numbers and mediate processes which violate baryon number. In this example, one sees from the definition (4) that there is a coupling of the  $X$  bosons to a  $\bar{d}$  quark and an electron. Similarly, there is a coupling of the  $X$  boson to a quark doublet and a positron. Note that there is no way to assign baryon and lepton number to the  $X$  boson so that it is conserved by these couplings.

In the GUT picture of baryogenesis, it is usually assumed that at temperatures well above the GUT scale, the universe is in thermal equilibrium. As the temperature drops below the mass of the  $X$  bosons, the reactions which produce the  $X$  bosons are not sufficiently rapid to maintain equilibrium. The decays of the  $X$  bosons violate baryon number; they also violate  $CP$ . So all three conditions are readily met: baryon-number violation,  $CP$  violation, and departure from equilibrium.

To understand in a bit more detail how the asymmetry can come about, note that  $CPT$  conservation requires that the total decay rate of  $X$  is the same as that of its antiparticle  $\bar{X}$ . But it does not require equal partial widths, i.e., the decays to particular final states. So starting with equal numbers of  $X$  and  $\bar{X}$  particles, there can be a slight asymmetry between the processes

$$X \rightarrow dL; \bar{X} \rightarrow \bar{Q}\bar{u} \quad (11)$$

and

$$\bar{X} \rightarrow \bar{d}\bar{L}; \bar{X} \rightarrow \bar{Q}u. \quad (12)$$

The tree graphs for these processes, shown in Fig. 1, are necessarily equal; any  $CP$ -violating phase simply cancels out when we take the absolute square of the amplitude. This is not true in higher order, where additional phases associated with real intermediate states can appear. Actually computing the baryon asymmetry requires a detailed analysis of a kind we will encounter later when we consider leptogenesis.

There are reasons to believe, however, that GUT baryogenesis is not the origin of the observed baryon asymmetry. Perhaps the most compelling of these has to do with inflation. Assuming that there was a period of inflation, any preexisting baryon number was greatly diluted. So in order that one produce baryons through  $X$  boson decay, it is necessary that the reheating temperature after inflation be at least comparable to the  $X$  boson mass. But as we have explained, a reheating temperature greater than  $10^9$  GeV leads to cosmological difficulties, especially overproduction of gravitinos.

#### D. Electroweak baryon-number violation

Earlier, we stated that the renormalizable interactions of the Standard Model preserve baryon number. This statement is valid classically, but it is not quite true of the quantum theory. There are, as we will see in this section, very tiny effects which violate baryon number ('t Hooft, 1976). These effects are tiny because they are due to quantum-mechanical tunneling, and are suppressed by a barrier penetration factor. At high temperatures, there is no such suppression, so baryon-number violation is a rapid process, which can come to thermal equilibrium. This has at least two possible implications. First, it is conceivable that these *sphaleron* processes can themselves be responsible for generating a baryon asymmetry. This is called electroweak baryogenesis (Kuzmin, Rubakov, and Shaposhnikov, 1985). Second, as we will see, sphaleron processes can process an existing lepton number, producing a net lepton and baryon number. This is the process called leptogenesis (Fukugita and Yanagida, 1986).

In this section, we summarize the main arguments that the electroweak interactions violate baryon number at high temperature. In the next section, we explain why

the electroweak interactions might produce a small baryon excess, and why this excess cannot be large enough to account for the observed asymmetry.

One of the great successes of the Standard Model is that it explains the observed conservation laws. In particular, there are no operators of dimension four or less consistent with the gauge symmetries which violate baryon number or the separate lepton numbers. The leading operators which can violate baryon number are of dimension six, and thus suppressed by  $\mathcal{O}(1/M^2)$ . The leading operators which violate the separate lepton numbers are of dimension 5, and thus suppressed by one power of  $1/M$ . In each case,  $M$  should be thought of as the energy scale associated with some very-high-energy physics which violates baryon or lepton number. This scale cannot be determined except through measurement or by specifying a more microscopic theory.

However, it is not quite true that the Standard Model preserves all of these symmetries. There are tiny effects which violate them, of order

$$e^{-(2\pi/\alpha_W)} \approx 10^{-91}, \quad (13)$$

where  $\alpha_W \approx 0.03$  is the weak coupling constant. These effects are related to the fact that the separate baryon-number and lepton-number currents are anomalous. When one quantizes the theory carefully, one finds that the baryon-number current  $j_B^\mu$  is not exactly conserved but rather satisfies

$$\partial_\mu j_B^\mu = \frac{3}{16\pi^2} F_{\mu\nu}^a \tilde{F}_{\mu\nu}^a = \frac{3}{8\pi^2} \text{Tr} F_{\mu\nu} \tilde{F}_{\mu\nu}. \quad (14)$$

Here  $F_{\mu\nu}$  are the SU(2) field strengths, and we have introduced matrix-valued fields in the last expression,

$$F_{\mu\nu} = \sum_a F_{\mu\nu}^a T^a. \quad (15)$$

The dual of  $F$ ,  $\tilde{F}$ , is defined by

$$\tilde{F}_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} F^{\rho\sigma}. \quad (16)$$

In electromagnetism,  $F_{\mu\nu} \tilde{F}_{\mu\nu} = 2\vec{E} \cdot \vec{B}$ .

The same anomaly (14) appears in the lepton-number current as well, i.e.,

$$\partial_\mu j_L^\mu = \frac{3}{16\pi^2} F_{\mu\nu}^a \tilde{F}_{\mu\nu}^a = \frac{3}{8\pi^2} \text{Tr} F_{\mu\nu} \tilde{F}_{\mu\nu}. \quad (17)$$

However, the difference of the two,  $j_B^\mu - j_L^\mu$ , is anomaly free and is an exactly conserved quantity in the Standard Model as well as in SU(5) and SO(10) grand unified theories.

One might think that such a violation of current conservation would lead to dramatic violations of the symmetry. But the problem is more subtle. The right-hand side of the anomaly equation is itself a total divergence:

$$\text{Tr} F_{\mu\nu} \tilde{F}_{\mu\nu} = \partial_\mu K^\mu, \quad (18)$$

where

$$K^\mu = \epsilon^{\mu\nu\rho\sigma} \text{Tr} \left( F_{\nu\rho} A_\sigma + \frac{2}{3} A_\nu A_\rho A_\sigma \right), \quad (19)$$

where the  $A_i$  are components of the vector potential.<sup>1</sup> In view of this,

$$\tilde{j} = j_B^\mu - \frac{3g^2}{8\pi^2} K^\mu \quad (20)$$

is conserved. In perturbation theory (i.e., calculated by Feynman diagrams),  $K^\mu$  falls to zero rapidly (typically like  $1/r^6$ ) at infinity, and so its integral is zero. This fact ensures that baryon number is conserved.

In Abelian gauge theories, this is the end of the story. In non-Abelian theories, however, there are nonperturbative field configurations which contribute to the right-hand side. These lead to violations of baryon number and the separate lepton numbers proportional to  $e^{-2\pi/\alpha}$ , where  $\alpha$  is the coupling constant of the theory. These configurations are called *instantons*. We will not discuss them in detail here; a pedagogical treatment is given by Coleman (1989). They correspond to the contribution of a tunneling amplitude. To understand what the tunneling process is, one must consider more carefully the ground state of the field theory. Classically, the ground states are field configurations for which the energy vanishes. The trivial solution of this condition is  $\vec{A} = 0$ , where  $\vec{A}$  is the vector potential. More generally, one can consider  $\vec{A}$  which is a ‘‘pure gauge,’’

$$\vec{A} = \frac{1}{i} U_g^{-1} \vec{\nabla} U_g, \quad (21)$$

where  $U_g$  is a gauge transformation matrix. In an Abelian [i.e., U(1)] gauge theory, fixing the gauge eliminates all but the trivial solution,  $\vec{A} = 0$ .<sup>2</sup> This is not the case for non-Abelian gauge theories. There is a class of gauge transformations, labeled by a discrete index  $n$ , which do not tend to unity as  $|\vec{x}| \rightarrow \infty$ , which must be considered to be distinct states. These have the form

$$U_n(\vec{x}) = e^{inf(\vec{x}) \cdot \vec{\tau}/2}, \quad (22)$$

where  $f(x) \rightarrow 2\pi$  as  $\vec{x} \rightarrow \infty$ ,  $f(\vec{x}) \rightarrow 0$  as  $\vec{x} \rightarrow 0$ , and  $\tau$  are the generators of the gauge group.

So the ground states of the gauge theory are labeled by an integer  $n$ . Now if we evaluate the integral of the current  $K^0$ , we obtain a quantity known as the *Chern-Simons number*:

$$\begin{aligned} n_{CS} &= \frac{1}{16\pi^2} \int d^3x K^0 \\ &= \frac{2/3}{16\pi^2} \int d^3x \epsilon_{ijk} \text{Tr} (U_g^{-1} \partial_i U_g U_g^{-1} \partial_j U_g U_g^{-1} \partial_k U_g). \end{aligned} \quad (23)$$

<sup>1</sup>The reader can quickly check this for a U(1) gauge theory like electromagnetism.

<sup>2</sup>More precisely, this is true in axial gauge. In the gauge  $A_0 = 0$ , it is necessary to sum over all time-independent transformations to construct a state which obeys Gauss's law.



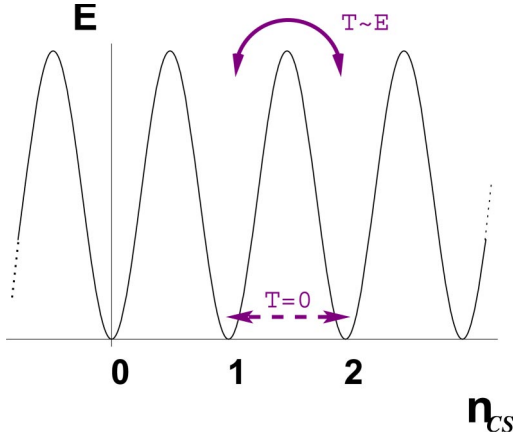


FIG. 2. (Color in online edition) Schematic Yang-Mills vacuum structure. At zero temperature, the instanton transitions between vacua with different Chern-Simons numbers are suppressed. At finite temperature, these transitions can proceed via sphalerons.

For  $U_g = U_n$ ,  $n_{CS} = n$ . The reader can also check that for  $U'_n = U_n(x)h(x)$ , where  $h$  is a gauge transformation which tends to unity at infinity (a so-called “small gauge transformation”), this quantity is unchanged.  $n_{CS}$  is topological in this sense. Note that for  $\vec{A}$ 's which are not pure gauge,  $n_{CS}$  is not quantized.

Schematically, we can thus think of the vacuum structure of a Yang-Mills theory as indicated in Fig. 2. At weak coupling, we have an infinite set of states labeled by integers and separated by barriers from one another. In tunneling processes which change the Chern-Simons number, the baryon and lepton numbers will change because of the anomaly. The exponential suppression found in the instanton calculation is typical of tunneling processes, and in fact the instanton calculation which leads to the result for the amplitude is nothing but a field-theoretic WKB calculation.

At zero temperature, the decay amplitude is suppressed, not only by  $e^{-2\pi/\alpha}$ , but by factors of Yukawa couplings. The probability that a single proton has decayed through this process in the history of the universe is infinitesimal. But this picture suggests that, at finite temperature, the rate should be larger. One can determine the height of the barrier separating configurations of different  $n_{CS}$  by looking for the field configuration which corresponds to the barrier saddle point. This is the sphaleron solution of the static equations of motion with finite energy (Manton, 1983). When one studies the small fluctuations about this solution, one finds that there is a single negative mode, corresponding to the possibility of rolling downhill into one or the other well. The sphaleron energy  $E_{sp}$  has the form

$$E_{sp} = \frac{c}{\alpha} M_W, \quad (24)$$

where  $c$  is a numerical coefficient and  $M_W$  is the mass of the  $W$  particle. The form can be seen by scaling arguments on the classical equations; determining the coefficient  $c$  requires a more detailed analysis. The rate  $\Gamma_{sp}$

for thermal fluctuations to cross the barrier per unit time per unit volume should be of order of the Boltzmann factor for this configuration, times a suitable prefactor (Kuzmin, Rubakov, and Shaposhnikov, 1985; Arnold and McLerran, 1988; Dine *et al.*, 1990),

$$\Gamma_{sp} = T^4 e^{-E_{sp}/T}. \quad (25)$$

Note that the rate becomes large as the temperature  $T$  approaches the  $W$  boson mass. In fact, at some temperature the weak interactions undergo a phase transition to a phase in which the  $W$  boson mass vanishes. At this point, the computation of the transition rate is a difficult problem—there is no small parameter—but general scaling arguments show that the baryon-violating transition rate is of the form<sup>3</sup>

$$\Gamma_{bv} = \alpha_W^4 T^4. \quad (26)$$

Returning to our original expression for the anomaly, we see that while the separate baryon and lepton numbers are violated in these processes, the combination  $B - L$  is conserved. This result leads to three observations:

- (1) If in the early universe, one creates baryon and lepton number, but no net  $B - L$ ,  $B$  and  $L$  will subsequently be lost through sphaleron processes.
- (2) If one creates a net  $B - L$  (e.g., creates a lepton number) the sphaleron process will leave both baryon and lepton numbers comparable to the original  $B - L$ . This realization is crucial to the idea of leptogenesis, to be discussed in more detail below.
- (3) The Standard Model satisfies, by itself, all of the conditions for baryogenesis.

## E. Electroweak baryogenesis

As we will see, while the Standard Model satisfies all of the conditions for baryogenesis (Kuzmin, Rubakov, and Shaposhnikov, 1985), nothing like the required baryon number can be produced. It is natural to ask whether extensions of the Standard Model, such as theories with complicated Higgs, or the minimal supersymmetric Standard Model (MSSM), can generate an asymmetry, using the sphaleron process discussed in the previous section. We will refer to such a possibility more generally as electroweak baryogenesis.

### 1. Electroweak baryogenesis in the Standard Model

How might baryons be produced in the Standard Model? From our discussion, it is clear that the first and second of Sakharov's conditions, baryon-number violation and  $CP$  violation, are satisfied. What about the need for a departure from equilibrium?

Above we alluded to the fact that in the electroweak theory, there is a phase transition to a phase with massless gauge bosons. It turns out that, for a sufficiently light Higgs, this transition is first order. At zero tempera-

<sup>3</sup>More detailed considerations alter slightly the parametric form of the rate (Arnold, Son, and Yaffe, 1997, 2000).

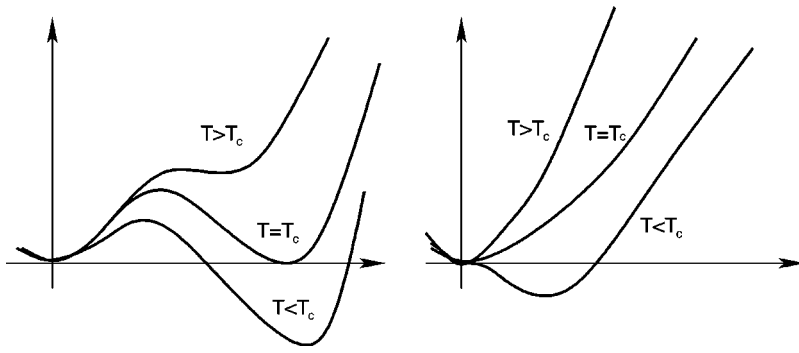


FIG. 3. Generic free-energy functions for first- and second-order phase transitions.

ture, in the simplest version of the Standard Model with a single Higgs field  $\Phi$ , the Higgs potential is given by

$$U(\Phi) = -\mu^2 |\Phi|^2 + \frac{\lambda}{2} |\Phi|^4. \quad (27)$$

The potential has a minimum at  $\Phi = (1/\sqrt{2})v_o = \sqrt{\mu^2/\lambda}$ , breaking the gauge symmetry and giving mass to the gauge bosons by the Higgs mechanism.

What about finite temperatures? By analogy with the phase transition in the Landau-Ginsburg model of superconductivity [see, e.g., de Gennes (1966)], one might expect that the value of  $\langle \Phi \rangle$  would change as the temperature increases. To determine the value of  $\Phi$ , one must compute the free energy as a function of  $\Phi$ . The leading temperature-dependent corrections are obtained simply by noting that the masses of the various fields in the theory—the  $W$  and  $Z$  bosons and the Higgs field, in particular—depend on  $\Phi$ . So the contributions of each species to the free-energy density  $\mathcal{F}$  are  $\Phi$  dependent:

$$\mathcal{F}(\Phi, T) = \pm T \sum_i \int \frac{d^3p}{2\pi^3} \ln(1 \mp e^{-\sqrt{p^2 + m_i^2(\Phi)}/T}), \quad (28)$$

where  $T$  is the temperature, the sum is over all particle species and physical helicity states, and the plus sign is for bosons, the minus for fermions. In the Standard Model, for temperature  $T \sim 100$  GeV, one can treat all the quarks as massless, except for the top quark. The effective potential (28) then depends on the top quark mass  $m_t$ , the vector boson masses  $M_Z$  and  $M_W$ , and on the Higgs mass  $M_H$ . Performing the integral in the equation yields

$$\mathcal{F}(\Phi, T) = D(T^2 - T_o^2)\Phi^2 - E T \Phi^3 + \frac{\lambda}{4} \Phi^4 + \dots \quad (29)$$

The parameters  $T_o$ ,  $D$ , and  $E$  are given in terms of the gauge boson masses and the gauge couplings below. For the moment, though, it is useful to note certain features of this expression.  $E$  turns out to be a rather small dimensionless number, of order  $10^{-2}$ . If we ignore the  $\Phi^3$  term, we have a second-order transition, at temperature  $T_o$ , between a phase with  $\Phi \neq 0$  and a phase with  $\Phi = 0$ . Because the  $W$  and  $Z$  masses are proportional to  $\Phi$ , this is a transition between a state with massive and massless gauge bosons.

Because of the  $\Phi^3$  term in the potential, the phase transition is potentially at least weakly first order. This is indicated in Fig. 3. Here one sees the appearance of a

second distinct minimum at some critical temperature. A first-order transition is not, in general, an adiabatic process. As the temperature decreases towards the transition temperature, the transition proceeds by the formation of bubbles. Inside the bubble the system is in the true equilibrium state, i.e., the state which minimizes the free energy, while outside it is closer to the original state. These bubbles form through thermal fluctuations at different points in the system, and grow until they collide, completing the phase transition. The moving bubble walls are regions where the Higgs fields are changing, and all of Sakharov's conditions are satisfied. It has been shown that various nonequilibrium processes near the wall can produce baryon and lepton numbers (Cohen, Kaplan, and Nelson, 1993; Rubakov and Shaposhnikov, 1996).

Describing these processes would take us far afield. Even without going through these details, however, one point is crucial: after the bubble has passed any given region, the baryon-violating processes should cease. If these processes continue, they wash out the baryon asymmetry produced during the phase transition.

To avoid washing out the asymmetry, the sphaleron rate after the phase transition should be small compared to the expansion rate of the universe. According to Eq. (25), this requires that after the transition the sphaleron energy be large compared to the temperature. This, in turn, means that  $M_W$  [see Eq. (24)] and the Higgs expectation value must be large immediately after the transition. Using Eq. (29) or more refined calculations to higher orders, one can relate the change in the Higgs expectation value to the Higgs mass at zero temperature. It turns out that the current lower limit on the Higgs boson mass<sup>4</sup> rules out any possibility of a large enough Higgs expectation value immediately after the phase transition, at least in the minimal model with a single Higgs doublet.

The shape of the free-energy potential  $\mathcal{F}(\Phi, T)$  near the critical temperature  $T_c$  determines whether the phase transition is first order, a necessary condition for electroweak baryogenesis to work. Equation (29) represents the lowest-order term in perturbation theory; higher-order terms have been computed as well (Arnold

<sup>4</sup>Carena and Haber (2003) report a lower limit of  $M_H \approx 115$  GeV from LEP searches.

and Espinosa, 1993; Bagnasco and Dine, 1993; Farakos *et al.*, 1994; Laine, 1994). However, these calculations are not reliable, because of infrared divergences which arise in the perturbation expansion. These arise because the Higgs field is nearly massless at the transition. Numerical simulations are required, often combined with a clever use of perturbation theory (Farakos *et al.*, 1995). The simulations (Kajantie *et al.*, 1996a, 1996b, 1997, 1999; Gurtler *et al.*, 1997; Karsch *et al.*, 1997; Rummukainen *et al.*, 1998; Csikor *et al.*, 1999) have shown that, for the Higgs mass above 80 GeV (which it must be, to satisfy the present experimental constraints), the sharp phase transition associated with the low mass Higgs turns into a smooth crossover.

However, even for an unrealistically light Higgs, the actual production of baryon asymmetry in the minimal Standard Model would be highly suppressed. The Standard Model  $CP$  violation arises from the Yukawa couplings of the quarks to the Higgs boson and must involve all three generations (Kobayashi and Maskawa, 1973). As a result, the lowest-order diagram that contributes to  $CP$ -violating processes relevant to baryogenesis is suppressed by 12 factors of the Yukawa couplings (Shaposhnikov, 1986, 1987). These couplings are small, leading to a contribution of the order of  $10^{-20}$  to the amount of the baryon asymmetry that could arise in the Standard Model.

Clearly, one must look beyond the Standard Model for the origin of baryon asymmetry of the universe. One of the best motivated candidates for new physics is supersymmetry.

Before closing this section, for completeness, we give the values of the parameters  $T_o$ ,  $B$ ,  $D$ , and  $E$  in Eq. (29) (Dine, Leigh, *et al.*, 1992):

$$T_o^2 = \frac{1}{2D}(\mu^2 - 4Bv_o^2) = \frac{1}{4D}(M_H^2 - 8Bv_o^2), \quad (30)$$

while the parameters  $B$ ,  $D$ , and  $E$  are given by

$$\begin{aligned} B &= \frac{3}{64\pi^2 v_o^4}(2M_W^4 + M_Z^4 - 4m_t^4), \\ D &= \frac{1}{8v_o^2}(2M_W^2 + M_Z^2 + 2m_t^2), \\ E &= \frac{1}{4\pi v_o^3}(2M_W^3 + M_Z^3) \sim 10^{-2}. \end{aligned} \quad (31)$$

As before,  $M_Z$ ,  $M_H$ , and  $M_W$  are the masses of the  $Z$ , the Higgs, and the  $W$  particles, and  $m_t$  is the mass of the top quark.

## 2. Supersymmetry, a short introduction

In this section we provide a brief introduction to supersymmetry. Much more detail can be found in Dine (1996) and in many texts.

There are many hints that supersymmetry, a hypothetical symmetry between fermions and bosons, might play some role in nature. For example, supersymmetry seems to be an essential part of superstring theory, the

only consistent theory of quantum gravity which we know. If supersymmetry is a symmetry of the laws of nature, however, it must be badly broken; otherwise we would have seen, for example, scalar electrons (“selectrons”) and fermionic photons (“photinos”). It has been widely conjectured that supersymmetry might be discovered by accelerators capable of exploring the TeV energy range. There are several reasons for this. The most compelling is the “hierarchy problem.” This is, at its most simple level, the puzzle of the wide disparity of energies between the Planck scale (or perhaps the unification scale) and the weak scale—roughly 17 orders of magnitude. While one might take this as simply a puzzling fact, within quantum theory the question is made sharper by the fact that scalar masses (particularly the Higgs mass) are subject to very divergent quantum corrections. A typical expression for the quantum corrections to a scalar mass is

$$\delta m^2 = \frac{\alpha}{\pi} \int d^4k \frac{1}{k^2}. \quad (32)$$

This integral diverges quadratically for large momentum  $k$ . Presumably, the integral is cut off by some unknown physics. If the energy scale of this physics is  $\Lambda$ , then the corrections to the Higgs mass are much larger than the scale of weak interactions unless  $\Lambda \sim \text{TeV}$ . In this form, the hierarchy problem is often referred to as the *naturalness* problem. While various cutoffs have been proposed, one of the most compelling suggestions is that the cutoff is the scale of supersymmetry breaking. In this case, the scale must be about 1000 GeV. If this hypothesis is correct, the Large Hadron Collider under construction at CERN should discover an array of new particles and interactions.<sup>5</sup>

The supersymmetry generators  $Q_\alpha$  are fermionic operators. Acting on bosons they produce fermions degenerate in energy; similarly, acting on fermions, they produce degenerate bosons. Their algebra involves the total energy and momentum  $P^\mu$ ,

$$\{Q_\alpha, Q_\beta\} = P^\mu \gamma_{\alpha\beta}^\mu, \quad (33)$$

where  $\gamma^\mu$  are the Dirac matrices. Neglecting gravity, supersymmetry is a global symmetry. Because of the structure of the algebra, the symmetry is broken if and only if the energy of the ground state is nonzero. If the symmetry is unbroken, for every boson there is a degenerate fermion, and conversely.

If we neglect gravity, there are two types of supermultiplets which may describe light fields. These are the chiral multiplets  $\Phi_i$ , containing a complex scalar  $\phi_i$  and a Weyl (two-component) fermion  $\psi_i$ ,

$$\Phi_i = (\phi_i, \psi_i), \quad (34)$$

and the vector multiplets  $V^a$ , containing a gauge boson  $A_\mu^a$  and a Weyl fermion (gaugino)  $\lambda^a$ ,

<sup>5</sup>It is possible that supersymmetry could be discovered at the Tevatron beforehand.

$$V^a = (A_\mu^a, \lambda^a). \quad (35)$$

In global supersymmetry, the Lagrangian is specified by the gauge symmetry and an analytic (more precisely holomorphic) function of the scalar fields,  $W(\phi_i)$ , known as the *superpotential*. For renormalizable theories,  $W$  has the form

$$W(\phi_i) = \frac{1}{2} m_{ij} \phi_i \phi_j + y_{ijk} \phi_i \phi_j \phi_k, \quad (36)$$

where  $m$  is a mass matrix and  $y$  is a matrix of couplings. Given the function  $W$ , the supersymmetric Lagrangian includes the following:

- (1) The usual covariant kinetic terms for all of the fields, for example,

$$\lambda^a D \lambda^{a*}, \quad \psi D \psi^*, \quad \frac{-1}{4} F_{\mu\nu}^2, \quad |D_\mu \phi|^2, \quad (37)$$

where  $D$  and  $D_\mu$  are the gauged derivative operators.

- (2) Yukawa couplings with gauge strength:

$$\sqrt{2} g^a \lambda^a \phi^* T^a \psi + \text{c.c.} \quad (38)$$

- (3) Mass terms and Yukawa couplings from  $W$ :

$$\frac{1}{2} \frac{\partial^2 W}{\partial \phi_i \partial \phi_j} \psi_i \psi_j \quad (39)$$

$$= m_{ij} \psi_i \psi_j + \frac{3}{2} y_{ijk} \phi_i \psi_j \psi_k. \quad (40)$$

- (4) A scalar potential:

$$U = \sum_i \left| \frac{\partial W}{\partial \phi_i} \right|^2 + \sum_a \frac{1}{2} (g^a)^2 \left( \sum_i \phi_i^* T^a \phi_i \right)^2. \quad (41)$$

It is convenient to define two types of auxiliary fields, the  $F$  and  $D^a$  fields:

$$F_i = \frac{\partial W}{\partial \phi_i}, \quad D^a = g^a \sum_i \phi_i^* T^a \phi_i. \quad (42)$$

In terms of these, the potential is simply

$$U = |F_i|^2 + \frac{1}{2} |D^a|^2 \quad (43)$$

and, at the classical level, supersymmetry is unbroken if and only if all of the  $D^a$  and  $F$  fields vanish at the minimum of the potential.

It is useful to consider some examples. Take first a model with a single chiral field  $\phi$  and the superpotential

$$W = \frac{1}{2} m \phi^2. \quad (44)$$

In this case, the potential is

$$U = \left| \frac{\partial W}{\partial \phi} \right|^2 = m^2 |\phi|^2. \quad (45)$$

On the other hand, the fermion mass from Eq. (40) comes out to be  $m$ , so the model describes two bosonic and two fermionic degrees of freedom, degenerate in mass.

A more interesting model is the supersymmetric version of the Standard Model, mentioned earlier, known as the minimal supersymmetric Standard Model (MSSM). The gauge group is  $SU(3) \times SU(2) \times U(1)$  as in the Standard Model and there is one vector multiplet for each gauge generator. In addition, for each of the usual fermions of the Standard Model, one has a chiral field with the same quantum numbers:

$$Q_a \in (3, 2)_{1/3}, \quad \bar{u}_a \in (\bar{3}, 1)_{-4/3}, \quad \bar{d}_a \in (\bar{3}, 1)_{2/3}, \\ L_a \in (1, 2)_{-1}, \quad \bar{e} \in (1, 1)_2. \quad (46)$$

Here  $a$  is a generation index,  $a=1,2,3$ . The representation of the gauge group is denoted by  $(n_3, n_2)_Y$  where  $n_3$  and  $n_2$  refer to  $SU(3)$  and  $SU(2)$ , respectively, and  $Y$  denotes the  $U(1)$  hypercharge. In addition, two Higgs fields are needed to cancel anomalies and to give mass to quarks and leptons,

$$H_u \in (1, 2)_1, \quad H_d \in (1, 2)_{-1}. \quad (47)$$

The superpotential of the model is a generalization of the Yukawa couplings of the Standard Model:

$$W = h^{ab} Q_a \bar{u}_b H_u + h^{ad} Q_a \bar{d}_d H_d + h^{ae} L_a \bar{e}_a H_d \\ + m_{ud} H_u H_d, \quad (48)$$

where the  $h$ 's are couplings. We have used our freedom to make field redefinitions to take the  $d$  quark and the lepton Yukawa couplings diagonal. If one supposes that the Higgs fields  $H_u$  and  $H_d$  have expectation values, this gives, through Eq. (39), masses for the quarks and leptons just as in the Standard Model. If supersymmetry is unbroken, their scalar partners have identical masses. The ratio of expectation values

$$\tan \beta \equiv \frac{\langle H_u \rangle}{\langle H_d \rangle} \quad (49)$$

is an important parameter concerning baryogenesis. The last term in Eq. (48) is a supersymmetric mass term for the Higgs fields. Supersymmetry breaking, essential to obtain a realistic model, will be discussed momentarily.

The gauge symmetries actually permit many more couplings than those written in Eq. (48). Couplings such as  $HL$ ,  $\bar{u}\bar{d}\bar{d}$ , and others would violate baryon or lepton number if they appeared. Because these are dimension 4, they are unsuppressed, unless they have extremely tiny dimensionless coefficients. They can be forbidden by a symmetry under which ordinary fields are even (quarks, leptons, and Higgs bosons) while their supersymmetric partners are odd. This symmetry is called *R parity*.

By itself, this model is not realistic, since supersymmetry is unbroken and all ordinary fields (quarks, leptons, gauge bosons, Higgs) are degenerate with their superpartners (squarks, sleptons, gauginos). The simplest solution to this is just to add soft breaking terms which explicitly break the supersymmetry. Because they are soft, they do not spoil the good features of these theories. These soft terms include mass terms for the squarks

and sleptons, Majorana mass terms for the gauginos, and cubic couplings of the scalar fields,

$$\bar{m}_{ij}^2 |\phi^{i*} \phi^j|^2 + m_\lambda \lambda \lambda + m_A A_{ijk} \phi^i \phi^j \phi^k. \quad (50)$$

In the MSSM, the couplings require 105 real parameters. We will think of all of the mass parameters as being of order the gravitino mass  $m_{3/2} \sim M_Z$ . These parameters are highly constrained, both by low-energy physics, particularly by the suppression of flavor-changing processes in weak interactions, and by direct searches at LEP and the Tevatron. Theoretical approaches to understanding these soft breakings can be divided broadly into two classes, gravity and gauge mediation. Both assume that some dynamics gives rise to spontaneous breakdown of supersymmetry. In “gravity mediation,” very high-energy physics is responsible for generating the soft terms; in gauge mediated models, lower-energy gauge interactions communicate supersymmetry breaking to ordinary fields.

$R$  parity, if present, implies that the lightest of the new particles, called the lightest supersymmetric particle (LSP), is stable. Typically this is the partner of a neutral gauge or Higgs boson; in this case the LSP is called a *neutralino*. One can calculate the abundance of the neutralinos as a function of the various supersymmetry-breaking parameters. The assumption that the supersymmetry-breaking masses are hundreds of GeV in magnitude leads automatically to a neutralino density of order the dark matter density of the universe, and this particle is a leading candidate for the dark matter (see Sec. III.F.5).

### 3. Baryogenesis in the minimal and next-to-minimal supersymmetric Standard Models

Supersymmetric extensions of the Standard Model contain new sources of  $CP$  violation (Dine *et al.*, 1991; Dine, Huet, and Singleton, 1992; Huet and Nelson, 1996) and an enlarged set of parameters which allow a greater possibility of a first-order transition (Espinosa, Quiros, and Zwirner, 1993; Espinosa, 1996; Bodeker *et al.*, 1997; Carena, Quiros, and Wagner, 1998; Cline and Moore, 1998). So it would seem possible that electroweak baryogenesis could operate effectively in these theories.

The new sources of  $CP$  violation may come, for example, from the chargino mass matrix:

$$\bar{\psi}_R M_\chi \psi_L = (\bar{\tilde{w}}^+, \bar{\tilde{h}}_2^+)_R \begin{pmatrix} m_2 & gH_2(x) \\ gH_1(x) & \mu \end{pmatrix} \begin{pmatrix} \tilde{w}^+ \\ \tilde{h}_1^+ \end{pmatrix}_L + \text{H.c.}, \quad (51)$$

where  $\tilde{w}$  and  $\tilde{h}$  are the superpartners of  $W$  boson and the charged Higgs, and  $m_2$  and  $\mu$  are mass parameters. Other possible sources include phases in scalar masses.<sup>6</sup> We will focus, however, on the terms in Eq. (51).

As long as  $m_2$  and  $\mu$  are complex, spatially varying phases in the bubble wall provide a source of (spontaneous)  $CP$  violation (Lee, 1974; Weinberg, 1976; Cohen, Kaplan, and Nelson, 1991). However, in light of the constraints on Higgs and superpartner masses, the present window for electroweak baryogenesis in the MSSM is very narrow if it exists at all (Cline, Joyce, and Kainulainen, 1998, 2000; Carena, Quiros, Seco, and Wagner, 2003). As we discussed above, the lighter the Higgs the easier it is to avoid the wash-out of baryon asymmetry produced in the phase transition. A light right-handed *stop*, the scalar partner of the top quark, allows for a first-order phase transition even for Higgs as heavy as 115 GeV, which is barely consistent with the current bounds. The predictions of the light Higgs and the light *stop* will soon be tested in experiment.

However, even for the most optimistic choice of parameters, it is difficult to obtain a baryon asymmetry as large as the observed value quoted in Eq. (1). Several parameters must be adjusted to maximize the baryon asymmetry. In particular, one must assume that the wall is very thin and choose the “optimal” bubble wall velocity  $v_w \approx 0.02$ . The origin of these difficulties lies, once again, in the strength of the electroweak phase transition. In the MSSM, the phase transition can be enhanced if the right-handed *stop* is assumed to be very light, while the left-handed *stop* is very heavy (Carena, Quiros, and Wagner, 1998). Then two-loop effects (Espinosa, 1996; Bodeker *et al.*, 1997) change the scalar potential sufficiently to allow for a first-order phase transition; lattice simulations support this perturbative result (Laine and Rummukainen, 1998; Csikor *et al.*, 2000). However, severe constraints arise from the experimental bounds on the chargino mass, as well as the chargino contribution to the electric dipole moment of the neutron (Chang, Chang, and Keung, 2002; Pilaftsis, 2002).

Different calculations of the baryon asymmetry in the MSSM yield somewhat different results (Cline, Joyce, and Kainulainen, 1998, 2000; Carena, Quiros, Seco, and Wagner, 2003), as can be seen from Fig. 4. According to Carena, Quiros, Seco, and Wagner (2003), it is possible to produce enough baryons if the Higgs boson and the right-handed *stop* are both very light, near the present experimental limits. In any case, electroweak baryogenesis in the MSSM is on the verge of being confirmed or ruled out by improving experimental constraints (Cline, 2000).

The strength of the phase transition can be further enhanced by adding a singlet Higgs to the model. In the next-to-minimal supersymmetric model (NMSM), the phase transition can be more strongly first-order (Davies, Froggatt, and Moorhouse, 1996; Huber *et al.*, 2001; Kainulainen *et al.*, 2001). The singlet also provides additional sources of  $CP$  violation which increase the baryon asymmetry (Huber and Schmidt, 2001).

### 4. Nonthermal electroweak baryogenesis at preheating

In light of these difficulties, various proposals have been put forth to obtain a viable picture of electroweak

<sup>6</sup>By field redefinitions, some of these can be shifted from fermion to scalar mass terms.

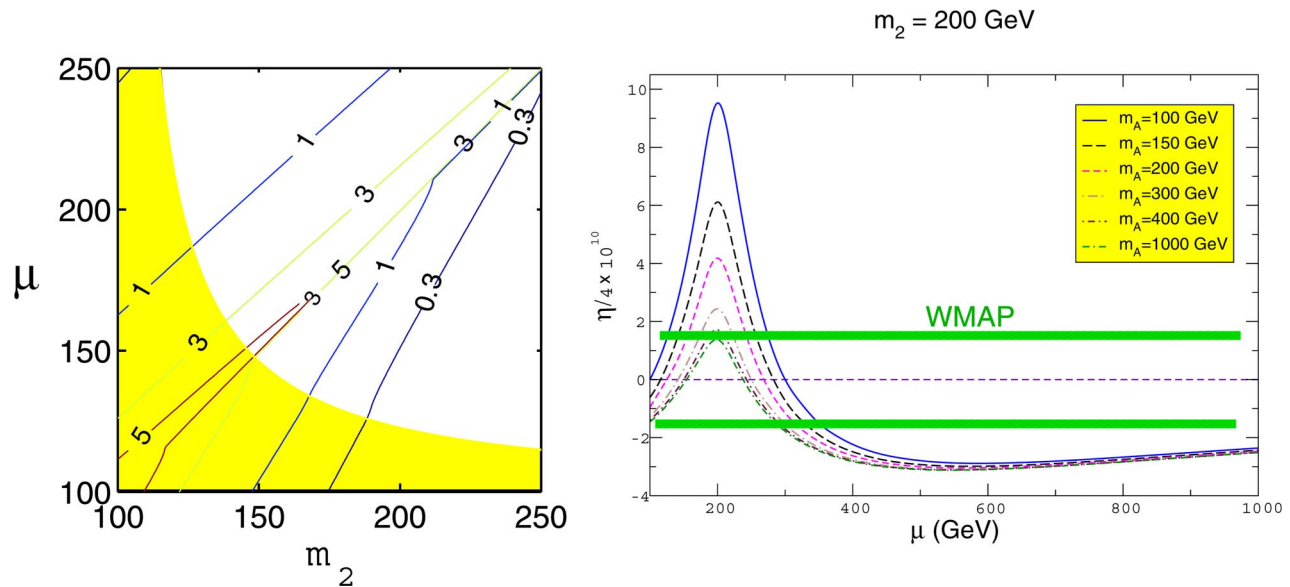


FIG. 4. (Color in online edition) Results of two independent calculations of baryon asymmetry in the MSSM. On the left, contours of constant baryon asymmetry in units  $10^{-10}$  calculated by Cline, Joyce, and Kainulainen (1998, 2000) as a function of the two masses  $\mu$  and  $m_2$  [see Eq. (51)], where mass units are GeV. The bubble wall velocity is taken as  $v_w = 0.03$  and  $\tan \beta \leq 3$  [see Eq. (49)]. To maximize the baryon asymmetry, they assume that the bubble wall is very narrow,  $\ell_w \approx 6/T$ . The shaded region is excluded by the LEP2 limit on the chargino mass,  $m_{\chi^\pm} > 104$  GeV. The plot on the right represents the results of Carena, Quiros, Seco, and Wagner (2003) for  $\tan \beta = 10$  and a maximal  $CP$  violating phase of the  $\mu$  parameter.  $m_A$  is the mass of the pseudoscalar Higgs particle. Also shown is the observed value of the baryon asymmetry reported by WMAP (Bennett *et al.*, 2003),  $\eta = 6.1^{+0.3}_{-0.2} \times 10^{-10}$ .

baryogenesis. These typically involve more drastic departures from thermal equilibrium than the weakly first-order phase transitions described above. The more extreme proposals suppose that inflation occurred at the electroweak scale and kicked the universe out of equilibrium, setting the stage for baryogenesis. It is generally believed that the natural scale for inflation is much higher than 100 GeV. Although models with weak (Randall and Thomas, 1995; German, Ross, and Sarkar, 2001) or intermediate (Randall, Soljagic, and Guth, 1996) scale inflation have been constructed, a lower scale of inflation is generally difficult to reconcile with the observed density perturbations  $(\delta\rho/\rho) \sim 10^{-5}$ . As a rule, the smaller the scale of inflation, the flatter the inflaton potential must be to produce the same density fluctuations. A weak-scale inflation would require the inflaton potential to be extremely flat, perhaps flatter than can plausibly be obtained in any physical theory.

Of course, one does not have to assume that the same inflation is responsible for  $(\delta\rho/\rho)$  and for baryogenesis. One could imagine that the universe has undergone more than one inflationary period. The primary inflation at a high scale could be responsible for the flatness of the universe and for the observed density perturbations. A secondary inflation at the weak scale need not produce an enormous expansion of the universe, and could create fertile soil for baryogenesis. One can debate the plausibility of invoking a second stage of inflation just for this purpose. In favor of such a possibility, it has been argued that a low-scale inflation might ameliorate the cosmological moduli problem common to many supersymmetric theories (Randall and Thomas, 1995; Ger-

man, Ross, and Sarkar, 2001). Nevertheless, inflationary models at the electroweak scale, which would help generate baryon asymmetry, usually suffer from naturalness problems (Lyth, 1999), which may be less severe in some cases (Copeland *et al.*, 2001).

What inhibits electroweak baryogenesis in the Standard Model is too much equilibration and too little  $CP$  violation. Both of these problems might be rectified if inflation is followed by reheating to a temperature just below the electroweak scale (García-Bellido *et al.*, 1999; Krauss and Trodden, 1999). Reheating, especially its variant dubbed *preheating*, involves a radical departure from thermal equilibrium,<sup>7</sup> and proceeds as follows.

During inflation, all matter and radiation are inflated away. When inflation is over, the energy stored in the inflaton is converted to thermal plasma. There are several possibilities for this reheating process. One possibility is that the inflaton may decay perturbatively into light particles, which eventually thermalize. However, in a class of models, a parametric resonance may greatly enhance the production of particles in some specific energy bands (Kofman, Linde, and Starobinsky, 1996). This process, caused by coherent oscillations of the inflaton, is known as *preheating*. Alternatively, the motion of the condensate may become spatially inhomogeneous on scales smaller than the horizon. This kind of transi-

<sup>7</sup>For references on the reheating scenario, see Khlebnikov and Shaposhnikov (1988); Kofman, Linde, and Starobinsky (1994, 1996, 1997); Kolda and March-Russell (1999).

tion in the motion of the inflaton, called a spinodal decomposition, may lead to a very rapid *tachyonic* preheating (Felder *et al.*, 2001).

All of these variants of reheating force the universe into a nonequilibrium state after the end of inflation and before thermalization takes place. This is, obviously, an opportune time for baryogenesis. The usual considerations of sphaleron transitions do not apply to a nonequilibrium system. But it turns out that baryon-number-violating processes similar to sphaleron transitions do take place at ordinary preheating (García-Bellido *et al.*, 1999; Cornwall and Kusenko, 2000), as well as at tachyonic preheating (Smit and Tranberg, 2002; García-Bellido *et al.*, 2003). This has been demonstrated by a combination of numerical and analytical arguments. In addition, preheating allows the coherent motions of some condensates to serve as sources of *CP* violation (Cornwall *et al.*, 2001). Such sources are poorly constrained by experiment and could have significant impact on baryogenesis. It is conceivable therefore that the electroweak-scale inflation could facilitate generation of the baryon asymmetry.

## F. Leptogenesis

Of the five scenarios for baryogenesis which we have listed in the Introduction, we have discussed two which are connected to very-high-energy physics: gravitational and GUT baryogenesis. We have given cosmological arguments why they are not likely. These arguments depend on assumptions which we cannot now reliably establish, so it is still possible that these mechanisms were operative. But if we tentatively accept these arguments we can significantly narrow our focus. Similarly we have seen that electroweak baryogenesis, while a beautiful idea, cannot be implemented in the Standard Model, and probably not in its minimal supersymmetric extension. So again, while we cannot rule out the possibility that electroweak baryogenesis in some extension of the Standard Model is relevant, it is tempting, for the moment, to view this possibility as unlikely. Adopting this point of view leaves leptogenesis and Affleck-Dine baryogenesis as the two most promising possibilities. These mechanisms are exciting because they have consequences for experiments which will be performed at accelerators over the next few years.

While there is no experimental evidence for supersymmetry apart from the unification of couplings, evidence for neutrino masses has become more and more compelling in the last few years (Gonzalez-Garcia and Nir, 2002). This comes from several sources: the fact that the flux of solar neutrinos does not match theoretical expectations, in the absence of masses and mixings (Fukuda *et al.*, 2001; Ahmad *et al.*, 2002); the apparent observation of neutrino oscillations among atmospheric neutrinos (Fukuda *et al.*, 1998); and direct measurements of neutrino mixing (Eguchi, 2003).

We will not review all of these phenomena here, but just mention that the atmospheric neutrino anomaly sug-

gests oscillations between the second and third generation of neutrinos (i.e., between muon and tau neutrinos):

$$\Delta m_\nu^2 = 10^{-2} - 10^{-4} \text{ eV}^2 \quad (52)$$

with mixing of order 1, while the solar neutrino deficit suggests smaller masses ( $\Delta m^2 \sim 10^{-6} \text{ eV}^2$ ). There is other evidence for neutrino oscillation from accelerator experiments. The SNO experiment has recently provided persuasive evidence in support of the hypothesis of mixing, as opposed to modifications of the standard solar model. The results from Super-Kamiokande, SNO, and KamLAND are in good agreement. There is also evidence of mixing from an experiment at Los Alamos (LSND). This result should be confirmed, or not, by the MiniBoone experiment at Fermilab. The mixing suggested by atmospheric neutrinos is currently being sought directly by accelerators. The data so far support the mixing interpretation, but are not yet decisive.

The most economical explanation of these facts is that neutrinos have Majorana masses arising from lepton-number-violating dimension-5 operators.<sup>8</sup> We have stressed that the leading operators permitted by the symmetries of the Standard Model which violate lepton number are nonrenormalizable operators of dimension-5, i.e., suppressed by one power of some large mass. Explicitly, these have the form

$$\mathcal{L}_{lv} = \frac{1}{M} LHLH. \quad (53)$$

Replacing the Higgs field by its expectation value  $v$  gives a mass for the neutrino of order  $v^2/M$ . If  $M = M_p$ , this mass is too small to account for either set of experimental results. So one expects that some lower scale is relevant. The “see-saw” mechanism provides a simple picture of how this scale might arise. One supposes that in addition to the neutrinos of the Standard Model, there are some  $SU(2) \times U(1)$ -singlet neutrinos,  $N$ . Nothing forbids these from obtaining a large mass. This could be of order  $M_{\text{GUT}}$ , for example, or a bit smaller. These neutrinos could also couple to the left-handed doublets  $\nu_L$ , just like right-handed charged leptons. These couplings take the form

$$\mathcal{L} = N^i H L^j h_{ij} + \text{c.c.} \quad (54)$$

Assuming, for the moment, that these couplings are not particularly small, one would obtain a mass matrix, in the  $\{N, \nu_L\}$  basis, of the form

$$M_\nu = \begin{pmatrix} M_N & m_l \\ m_l^T & 0 \end{pmatrix}. \quad (55)$$

Here  $m_l$  is a matrix whose entries are of the order of the lepton masses inside. Then  $M_\nu$  has an eigenvalue of order  $m_l^2/M_N$ . The latter number is of the order of magnitude needed to explain the neutrino anomaly for  $M_N$

<sup>8</sup>A Majorana mass is a mass for a two-component fermion, which is permitted if the fermion carries no conserved charges.

$\sim 10^{13}$  or so, i.e., not wildly different from the GUT scale and other scales which have been proposed for new physics.

The fact that couplings of  $N$  violate lepton number in this model is important for leptogenesis (Fukugita and Yanagida, 1986).  $N$  is a heavy particle; it can decay both to  $H + \nu$  and  $H + \bar{\nu}$ , for example. The partial widths for each of these final states need not be the same.  $CP$  violation can enter through phases in the Yukawa couplings and mass matrices of the  $N$ 's. At tree level, however, these phases will cancel out between decays to the various states and their (would-be)  $CP$  conjugates, as in the case of GUT's we discussed earlier. So it is necessary to consider interference between tree and one loop diagrams with discontinuities, as in Fig. 1. In a model with three  $N$ 's, there are  $CP$ -violating phases in the Yukawa couplings of the  $N$ 's to the light Higgs. The heaviest of the right-handed neutrinos, say  $N_1$ , can decay to a lepton  $\ell$  and a Higgs, or to  $\bar{\ell}$  and a Higgs. At tree level, as in the case of GUT baryogenesis, the rates for production of leptons and antileptons are equal, even though there are  $CP$  violating phases in the couplings. It is necessary, again, to look at quantum corrections, in which dynamical phases can appear in the amplitudes. At one loop, the decay amplitude for  $N$  has a discontinuity associated with the fact that the intermediate  $N_1$  and  $N_2$  can be on shell. So one obtains an asymmetry proportional to the imaginary parts of the Yukawa couplings of the  $N$ 's to the Higgs:

$$\epsilon = \frac{\Gamma(N_1 \rightarrow \ell H_2) - \Gamma(N_1 \rightarrow \bar{\ell} \bar{H}_2)}{\Gamma(N_1 \rightarrow \ell H_2) + \Gamma(N_1 \rightarrow \bar{\ell} \bar{H}_2)} \quad (56)$$

$$= \frac{1}{8\pi} \frac{1}{hh^\dagger} \sum_{i=2,3} \text{Im}[(h_\nu h_\nu^\dagger)_{1i}]^2 f\left(\frac{M_2^2}{M_1^2}\right), \quad (57)$$

where  $\Gamma$  is the decay width and  $h$  is the coupling in Eq. (54). Also,  $f$  is a function that represents radiative corrections. For example, in the Standard Model  $f = \sqrt{x}[(x-2)/(x-1) + (x+1)\ln(1+1/x)]$ , while in the MSSM  $f = \sqrt{x}[2/(x-1) + \ln(1+1/x)]$ . Here we have allowed for the possibility of multiple Higgs fields, with  $H_2$  coupling to the leptons. The rough order of magnitude here is readily understood by simply counting loop factors. It need not be terribly small.

Now, as the universe cools through temperatures of order of the masses of the  $N$ 's, they drop out of equilibrium, and their decays can lead to an excess of neutrinos over antineutrinos. Detailed predictions can be obtained by integrating a suitable set of Boltzmann equations. Alternatively, these particles can be produced out of equilibrium, at preheating following inflation (García-Bellido and Morales, 2002).

These decays produce a net lepton number but not baryon number, and hence a net  $B-L$ . The resulting lepton number will be further modified by sphaleron interactions, yielding a net lepton and baryon number (recall that sphaleron interactions preserve  $B-L$ , but violate  $B$  and  $L$  separately). One can determine the

resulting asymmetry by an elementary thermodynamics exercise as follows (Harvey and Turner, 1990). The calculation is particularly simple because at these high temperatures the masses are negligible. One introduces chemical potentials for each neutrino, quark, and charged lepton species. One then considers the various interactions between the species at equilibrium. For any allowed chemical reaction, the sum of the chemical  $\mu_i$  potentials on each side of the reaction must be equal. For neutrinos, the relations come from the following:

- (1) the sphaleron interactions themselves:

$$\sum_i (3\mu_{q_i} + \mu_{\ell_i}) = 0; \quad (58)$$

- (2) a similar relation for QCD sphalerons:

$$\sum_i (2\mu_{q_i} - \mu_{u_i} - \mu_{d_i}) = 0; \quad (59)$$

- (3) vanishing of the total hypercharge of the universe:

$$\sum_i (\mu_{q_i} - 2\mu_{\bar{u}_i} + \mu_{\bar{d}_i} - \mu_{\ell_i} + \mu_{\bar{\nu}_i}) + \frac{2}{N} \mu_H = 0; \quad (60)$$

- (4) the quark and lepton Yukawa couplings give relations:

$$\begin{aligned} \mu_{q_i} - \mu_\phi - \mu_{d_j} &= 0, & \mu_{q_i} - \mu_\phi - \mu_{u_j} &= 0, \\ \mu_{\ell_i} - \mu_\phi - \mu_{e_j} &= 0. \end{aligned} \quad (61)$$

The number of equations here is the same as the number of unknowns. Combining these, one can solve for the chemical potentials in terms of the lepton chemical potential, and finally in terms of the initial  $(B-L)_0$ . With  $N_g$  generations,

$$B = \frac{8N_g + 4}{22N_g + 13} (B-L)_0. \quad (62)$$

Reasonable values of the neutrino parameters give asymmetries of the order we seek to explain. Note sources of small numbers:

- (1) the phases in the couplings;
- (2) the loop factor;
- (3) the small density of the  $N$  particles when they drop out of equilibrium. Parametrically, the production rate is

$$\Gamma \sim e^{(-M/T)} g^2 T. \quad (63)$$

This is much less than the Hubble rate  $H \sim T^2/M_p$ , once the density is suppressed by  $T/M_p$ . This factor of order  $10^{-6}$  for a  $10^{13}$ -GeV particle.

It is interesting to ask, assuming that these processes are the source of the observed asymmetry, how many parameters which enter into the computation can be measured? In other words, can we relate the observed number to microphysics? It is likely that, over time, many of the parameters of the light neutrino mass matrices, including possible  $CP$ -violating effects, will be measured (Gonzalez-Garcia and Nir, 2002). But while these measurements determine some of the  $N_i$  couplings and masses, they are not, in general, enough. In order to



make a precise calculation of the baryon number density analogous to calculations of nucleosynthesis, one needs additional information about the masses of the fields  $N_i$  (Pascoli *et al.*, 2003). One requires either some other experimental access to this higher-scale physics, or a compelling theory of neutrino mass in which the number of parameters is reduced, perhaps by symmetries.

### G. Baryogenesis through coherent scalar fields

We have seen that supersymmetry introduces new possibilities for electroweak baryogenesis. But the most striking feature of supersymmetric models, from the point of view of baryogenesis, is the appearance of scalar fields carrying baryon and lepton number. These scalars offer the possibility of coherent production of baryons. In the limit that supersymmetry is unbroken, many of these scalars have flat or nearly flat potentials. They are thus easily displaced from their minima in the highly energetic environment of the early universe. We will often refer to such configurations as “excited.” Simple processes can produce substantial amounts of baryons. This coherent production of baryons, known as Affleck-Dine baryogenesis, is the focus of the rest of this review.

## III. AFFLECK-DINE BARYOGENESIS

### A. Arguments for coherent production of baryon number

In the previous section, we have reviewed several proposals for baryogenesis. None can be firmly ruled out. However, all but two seem unlikely: leptogenesis and Affleck-Dine baryogenesis. While the discovery of neutrino mass gives support to the possibility of leptogenesis, there are a number of reasons to consider coherent production:

- The Standard Model alone cannot explain the baryon asymmetry of the universe, the main obstacle being the large mass of the Higgs. One needs new physics for baryogenesis. The requisite new physics may reside at a very high scale or at a lower scale. An increasing body of evidence implies that inflation probably took place in the early universe. Hence baryogenesis must have happened at or after reheating. To avoid overproducing weakly interacting light particles, for example, gravitinos and other new states predicted in theory, one would like the reheat temperature not to exceed  $10^9$  GeV. This poses a problem for GUT baryogenesis. This also limits possibilities for leptogenesis. Affleck-Dine baryogenesis, on the other hand, is consistent with the low-energy and temperature scales required by inflation.
- Supersymmetry is widely regarded as a plausible, elegant, and natural candidate for physics beyond the Standard Model. Of the two simple scenarios for baryogenesis in the MSSM, the electroweak scenario is on the verge of being confirmed or ruled out by constraints on supersymmetric particles imposed by accelerator experiments.

- The remaining low-reheat SUSY scenario, Affleck-Dine baryogenesis, can naturally reproduce the observed baryon asymmetry of the universe. The formation of an Affleck-Dine condensate can occur quite generically in cosmological models.
- The Affleck-Dine scenario potentially can give rise simultaneously to the ordinary matter and the dark matter in the universe. This can explain why the amounts of luminous and dark matter are surprisingly close to each other, within one order of magnitude. If the two entities formed in completely unrelated processes (for example, the baryon asymmetry from leptogenesis, while the dark matter from freezeout of neutralinos), the observed relation  $\Omega_{\text{DARK}} \sim \Omega_{\text{matter}}$  is fortuitous.<sup>9</sup>
- Many particle physics models lead to significant production of entropy at relatively late times (Cohen, Kaplan, and Nelson, 1993). This dilutes whatever baryon number existed previously. Coherent production can be extremely efficient, and in many models, it is precisely this late dilution which yields the small baryon density observed today.

In the rest of this section, we discuss Affleck-Dine baryogenesis in some detail.

### B. Baryogenesis through a coherent scalar field

In supersymmetric theories, the ordinary quarks and leptons are accompanied by scalar fields. These scalar fields carry baryon and lepton number. A coherent field, i.e., a field which has a large vacuum expectation value (VEV), can in principle carry a large baryon number. As we will see, it is quite plausible that such fields, also called condensates, were excited in the early universe.

To understand the basics of the mechanism, we consider first a model with a single complex scalar field. We take the Lagrangian to be

$$\mathcal{L} = |\partial_\mu \phi|^2 - m^2 |\phi|^2. \quad (64)$$

This Lagrangian has a phase symmetry,  $\phi \rightarrow e^{i\alpha} \phi$ , and a corresponding conserved current, which we will refer to as baryon number:

$$j_B^\mu = i(\phi^* \partial^\mu \phi - \phi \partial^\mu \phi^*). \quad (65)$$

It also possesses a “CP” symmetry:

$$\phi \leftrightarrow \phi^*. \quad (66)$$

With supersymmetry in mind, we will think of  $m$  as of order  $M_W$ .

If we focus on the behavior of spatially constant fields,  $\phi(\vec{x}, t) = \phi(t)$ , this system is equivalent to an isotropic harmonic oscillator in two dimensions. This remains the case if we add higher-order terms which respect the phase symmetry. In supersymmetric models, however,

<sup>9</sup>An additional *ad hoc* symmetry can also help relate the amounts of ordinary matter and dark matter (Kaplan, 1992).

we expect that higher-order terms will break the symmetry. In the isotropic oscillator analogy, this corresponds to anharmonic terms breaking the rotational invariance. With a general initial condition, the system will develop some nonzero angular momentum. If the motion is damped, so that the amplitude of the oscillations decreases, these rotationally noninvariant terms will become less important with time.

Let us add interactions in the following way, which will closely parallel what happens in the supersymmetric case. We include a set of quartic couplings in the Lagrangian

$$\mathcal{L}_I = \lambda |\phi|^4 + \epsilon \phi^3 \phi^* + \delta \phi^4 + \text{c.c.} \quad (67)$$

These interactions clearly violate baryon number. For general complex  $\epsilon$  and  $\delta$ , they also violate  $CP$ . In supersymmetric theories, as we will shortly see, the couplings  $\lambda, \epsilon, \delta, \dots$  will be extremely small,  $\mathcal{O}(M_W^2/M_p^2)$  or  $\mathcal{O}(M_W^2/M_{\text{GUT}}^2)$ .

In order that these tiny couplings lead to an appreciable baryon number, it is necessary that the fields were very large at some stage. To see how the cosmic evolution of this system can lead to a nonzero baryon number, we first note that at very early times, when the Hubble rate satisfies  $H \gg m$ , the mass of the field is irrelevant. It is thus reasonable to suppose that at this early time  $\phi$  has an expectation value  $\phi_o \gg 0$ ; later we will describe some specific suggestions as to how this might come about. How does the field then evolve? First we ignore the quartic interactions. In a gravitational background, the equation of motion for the field is

$$(-|g|)^{-1/2} \partial_\mu (-|g|)^{1/2} g_{\mu\nu}^{-1} \partial_\nu \phi + \frac{\partial U}{\partial \phi} = 0, \quad (68)$$

where  $g_{\mu\nu}$  is the metric. For a spatially homogeneous field  $\phi(t)$ , in a Robertson-Walker background, this becomes

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial U}{\partial \phi} = 0. \quad (69)$$

At very early times,  $H \gg m$ , and the system is highly overdamped and essentially frozen at  $\phi = \phi_o$ . At this point,  $B=0$ . However, once the universe has aged enough that  $H \ll m$ ,  $\phi$  begins to oscillate. Substituting  $H = 1/2t$  or  $H = 2/3t$  for the radiation- and matter-dominated eras, respectively, one finds that

$$\phi = \begin{cases} \frac{\phi_o}{(mt)^{3/2}} \sin(mt) & \text{(radiation)} \\ \frac{\phi_o}{(mt)} \sin(mt) & \text{(matter)}. \end{cases} \quad (70)$$

In either case, the energy behaves as

$$E \approx m^2 \phi_o^2 \left( \frac{R_o}{R} \right)^3, \quad (71)$$

where  $R(t)$  is the scale factor in the Robertson-Walker metric Eq. (2). This decreases like  $R^3$ , as would the en-

ergy of pressureless dust. One can think of this oscillating field as a coherent state of  $\phi$  particles with momentum  $\vec{p}=0$ .

Now let us consider the effects of the quartic couplings. Since the field amplitude damps with time, their significance will decrease with time. Suppose, initially, that  $\phi_o$  is real. Then at early times the imaginary part  $\phi_i$  satisfies

$$\ddot{\phi}_i + 3H\dot{\phi}_i + m^2 \phi_i \approx \text{Im}(\epsilon + \delta) \phi_o^3, \quad (72)$$

in the approximation that  $\epsilon$  and  $\delta$  are small. Asymptotically, the right hand falls as  $t^{-9/2}$ , whereas  $\phi_i$  falls off only as  $t^{-3/2}$ . As a result, just as in our mechanical analogy, baryon-number (angular momentum) violation becomes negligible. The equation goes over to the free equation, with a solution of the form

$$\begin{aligned} \phi_i &= a_r \frac{\text{Im}(\epsilon + \delta) \phi_o^3}{m^2 (mt)^{3/4}} \sin(mt + \delta_r) \quad \text{(radiation)}, \\ \phi_i &= a_m \frac{\text{Im}(\epsilon + \delta) \phi_o^3}{m^3 t} \sin(mt + \delta_m) \quad \text{(matter)}, \end{aligned} \quad (73)$$

in the radiation- and matter-dominated cases, respectively. The constants  $\delta_m$ ,  $\delta_r$ ,  $a_m$ , and  $a_r$  can easily be obtained numerically, and are of order unity:

$$a_r = 0.85, \quad a_m = 0.85, \quad \delta_r = -0.91, \quad \delta_m = 1.54. \quad (74)$$

But now we have a nonzero baryon number. Substituting Eq. (73) in the expression for the current, Eq. (65), we find

$$\begin{aligned} n_B &= 2a_r \text{Im}(\epsilon + \delta) \frac{\phi_o^2}{m(mt)^2} \sin(\delta_r + \pi/8) \quad \text{(radiation)}, \\ n_B &= 2a_m \text{Im}(\epsilon + \delta) \frac{\phi_o^2}{m(mt)^2} \sin(\delta_m) \quad \text{(matter)}. \end{aligned} \quad (75)$$

Two features of these results should be noted. First, if  $\epsilon$  and  $\delta$  vanish or are real,  $n_B$  vanishes. It is remarkable that the Lagrangian parameters can be real, and yet  $\phi_o$  can be complex, still giving rise to a net baryon number. We will discuss plausible initial values for the fields later, after we have discussed supersymmetry breaking in the early universe. Finally, we should point out that, as expected,  $n_B$  is conserved at late times.

This mechanism for generating baryon number could be considered without supersymmetry. In that case, it begs several questions:

- What are the scalar fields carrying baryon number?
- Why are the  $\phi^4$  terms so small?
- How are the scalars in the condensate converted to more familiar particles?

Supersymmetry provides a natural answer to each of these questions. First, as we have stressed, there are scalar fields carrying baryon and lepton number. As we will see, in the limit that supersymmetry is unbroken, there are typically directions in the field space in which the quartic terms in the potential vanish. Finally, the scalar

quarks and leptons will be able to decay (in a baryon- and lepton-number conserving fashion) to ordinary quarks.

### C. Flat directions and baryogenesis

To discuss the problem of baryon-number generation, we first want to examine the theory in a limit in which we ignore the soft SUSY-breaking terms. After all, at very early times,  $H \gg M_W$ , and these terms are irrelevant. We want to ask whether in a model like the MSSM, some fields can have large VEV's. This requires that there are directions in the field space for which the potential is flat. Before considering the full MSSM, it is again helpful to consider a simpler model, in this case a theory with gauge group  $U(1)$ , and two chiral fields,  $\phi^+$  and  $\phi^-$  with opposite charge. We assume the superpotential simply vanishes. In this case the potential is

$$U = \frac{1}{2}(D^y)^2, \quad D^y = g(\phi^{+*}\phi^+ - \phi^{-*}\phi^-). \quad (76)$$

But  $D^y$ , and the potential, vanish if  $\phi^+ = \phi^- = a$ . It is not difficult to work out the spectrum in a vacuum of nonzero  $a$ . One finds that there is one massless chiral field, and a massive vector field containing a massive gauge boson, a massive Dirac field, and a massive scalar.

We now consider a somewhat more elaborate example. Let us take the MSSM and give expectation values to the Higgs and the slepton fields of Eqs. (46) and (47):

$$H_u = \begin{pmatrix} 0 \\ a \end{pmatrix}, \quad L_1 = \begin{pmatrix} a \\ 0 \end{pmatrix}. \quad (77)$$

The  $F$  field of Eq. (42) vanishes in this direction, since the potentially problematic  $H_u L$  term in the superpotential is absent by  $R$  parity. The other possible contributions vanish because  $Q = H_U = 0$ . It is easy to see that the  $D$  term for hypercharge in Eq. (43) vanishes,

$$D^Y = g'^2(|H_u|^2 - |L|^2) = 0. \quad (78)$$

To see that the  $D$  term for  $SU(2)$  vanishes, one can work directly with the Pauli matrices, or use instead the following device which works for a general  $SU(N)$  group. Just as one defines a matrix-valued gauge field,

$$(A_\mu)^i_j = A_\mu^a (T^a)^i_j, \quad (79)$$

one defines

$$(D)^i_j = D^a (T^a)^i_j. \quad (80)$$

Then, using the  $SU(N)$  identity,

$$(T^a)^i_j (T^a)^k_l = \delta_l^k \delta_j^i - \frac{1}{N} \delta_j^i \delta_l^k, \quad (81)$$

the contribution to  $(D)^i_j$  from a field  $\phi$  in the fundamental representation is simply

$$(D)^i_j = \phi^{i*} \phi_j - \frac{1}{N} |\phi|^2 \delta_j^i. \quad (82)$$

In the present case, this becomes

$$(D)^i_j = \begin{pmatrix} |a|^2 & 0 \\ 0 & |a|^2 \end{pmatrix} - \frac{1}{2} |a|^2 \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 0. \quad (83)$$

What is particularly interesting about this direction is that the field carries a lepton number. As we have seen, producing a lepton number is for all intents and purposes like producing a baryon number.

Nonrenormalizable, higher-dimension terms with more fields can lift the flat direction. For example, a quartic term in the superpotential,

$$W = \frac{1}{M} (H_u L)^2, \quad (84)$$

respects all of the gauge symmetries and is invariant under  $R$  parity. It gives rise to a potential

$$U_{\text{lift}} = \frac{\Phi^6}{M^2}, \quad (85)$$

where  $\Phi$  is the superfield whose VEV parametrizes the flat direction.

There are many more flat directions, and many of these carry baryon or lepton number.<sup>10</sup> An example of a flat direction with both baryon and lepton number excited is the following:

$$\text{First generation: } Q_1^1 = b, \quad \bar{u}_2 = a, \quad L_2 = b;$$

$$\text{second: } \bar{d}_1 = \sqrt{|b|^2 + |a|^2}; \quad \text{third: } \bar{d}_3 = a. \quad (86)$$

Here  $a$  and  $b$  are parameters,  $Q$  carries an upper color index and a lower  $SU(2)$  index, and we have suppressed the generation indices. It is straightforward to check that this is indeed a flat direction. That the  $D$  terms vanish for  $SU(2)$ ,  $SU(3)$ , and  $U(1)$  follows from a calculation similar to that in Eq. (83). The  $F$  terms also vanish:

$$\frac{\partial W}{\partial H_u} = \Gamma^{ab} Q^a \bar{u}^b = 0 \quad \text{and} \quad \frac{\partial W}{\partial H_d} = \gamma^a Q^a \bar{d}^a = 0. \quad (87)$$

The first follows since the  $\bar{u}$  fields have their expectation values in different ‘‘color slots’’ than the  $Q$  fields. The second is automatically satisfied since the  $\bar{d}$  and  $Q$  fields have expectation values in different generations, and these Yukawa couplings do not mix generations.

Higher-dimension operators again can lift this flat direction. In this case the leading term is

$$W = \frac{1}{M^3} [Q^1 \bar{d}^2 L^1] [\bar{u}^1 \bar{d}^2 \bar{d}^3]. \quad (88)$$

Here the superscripts denote flavor. We have suppressed color and  $SU(2)$  indices, but the braces indicate sets of fields which are contracted in  $SU(3)$ - and  $SU(2)$ -invariant ways. In addition to being completely gauge invariant, this operator is invariant under ordinary  $R$  parity. Lower-dimension operators including operators of dimension 4, can lift the flat direction but violate  $R$  parity. It gives rise to a term in the potential

<sup>10</sup>The flat directions in the MSSM have been cataloged by Gherghetta, Kolda, and Martin (1996).

$$U_{\text{lift}} = \frac{\Phi^{10}}{M^6}. \quad (89)$$

Here  $\Phi$  refers in a generic way to the fields whose VEV's parametrize the flat directions  $(a, b)$ .

#### D. Evolution of the condensate

For the cosmologies we wish to consider, higher-dimension operators are quite important despite the powers of  $1/M$ . During inflation, for example, such operators can determine the initial value of the field,  $\Phi_0$ . Here again  $\Phi$  denotes in a generic way the fields which parametrize the flat directions.

##### 1. Supersymmetry breaking in the early universe

We have indicated that higher-dimension supersymmetric operators give rise to potentials in the flat directions. To fully understand the behavior of the fields in the early universe, we need to consider supersymmetry breaking, which gives rise to additional potential terms.

We have indicated in Eq. (50) the sorts of supersymmetry-breaking terms which we expect in supersymmetric theories. In the early universe, we expect supersymmetry is much more badly broken. For example, during inflation, the nonzero energy density associated with the cosmological constant breaks supersymmetry. Suppose that  $I$  is the inflaton field, and that the inflaton potential arises because of a nonzero value of the auxiliary field for  $I$ ,  $F_I = \partial W / \partial I$  [see Eq. (42)].  $F_I$  is an order parameter for supersymmetry breaking as are the auxiliary fields for any field; this quantity is roughly constant during inflation. So, during inflation, supersymmetry is broken by a large amount (Dine, Randall, and Thomas, 1995). As a result, not surprisingly, there can be an appreciable supersymmetry-breaking potential of  $\Phi$ . These contributions to the potential have the form<sup>11</sup>

$$U_H = H^2 \Phi^2 f(\Phi^2 / M_p^2), \quad (90)$$

where  $H$  here is the Hubble rate. It is perfectly possible for the second derivative of the potential near the origin to be negative. In this case, we write our higher-dimension term as

$$W_n = \frac{1}{M^n} \Phi^{n+3}, \quad (91)$$

where  $n$  is some integer and  $M$  is the scale of new physics. The potential takes the form

$$U = -H^2 |\Phi|^2 + \frac{1}{M^{2n}} |\Phi|^{2n+4}. \quad (92)$$

<sup>11</sup>When supersymmetric theories are coupled to gravity, there are corrections to Eq. (41). It also makes no sense to restrict the Lagrangian to be renormalizable. The assumption that nonrenormalizable couplings scale with  $M_p$  leads to Eq. (90), as explained in Dine, Randall, and Thomas (1995).

The minimum of the potential then lies at

$$\Phi_0 \approx M \left( \frac{H}{M} \right)^{1/(n+1)}. \quad (93)$$

More generally, one can see that the higher the dimension of the operator which raises the flat direction, the larger the starting value of the field—and the larger the ultimate value of the baryon number. Typically, there is plenty of time for the field to find its minimum during inflation. After inflation,  $H$  decreases, and the field  $\Phi$  evolves adiabatically, oscillating slowly about the local minimum for some time.

Our examples illustrate that in models with  $R$  parity, the value of  $n$ , and hence the size of the initial field, can vary appreciably. With further symmetries, it is possible that  $n$  is larger, and even that all operators which might lift the flat direction are forbidden (Dine, Randall, and Thomas, 1996). For the rest of this section we will continue to assume that the flat directions are lifted by terms in the superpotential. If they are not, the required analysis is different, since the lifting of the flat direction is entirely associated with supersymmetry breaking.

##### 2. Appearance of baryon number

The  $|\partial W / \partial \Phi|^2$  term in the potential does not break either baryon number or  $CP$  symmetry. In most models, it turns out that the leading sources of  $B$  and  $CP$  violation come from supersymmetry-breaking terms associated with  $F_I$ . These have the form<sup>12</sup>

$$am_{3/2}W + bHW. \quad (94)$$

Here  $a$  and  $b$  are complex, dimensionless constants. The relative phase in these two terms,  $\delta = \tan^{-1}(ab^*/|ab|)$ , violates  $CP$ . This is crucial; if the two terms carry the same phase, then the phase can be eliminated by a field redefinition, and we have to look elsewhere for possible  $CP$ -violating effects. Examining Eqs. (84) and (88), one sees that these terms violate baryon and/or lepton number. In following the evolution of the field  $\Phi$ , the important era occurs when  $H \sim m_{3/2}$ . At this point, the phase misalignment of the two terms, along with the baryon-number-violating coupling, leads to the appearance of a baryon number. From the equations of motion, the time rate of change of the baryon number is

$$\frac{dn_B}{dt} = \frac{\sin(\delta)m_{3/2}}{M^n} \Phi^{n+3}. \quad (95)$$

Assuming that the relevant time is  $H^{-1}$ , one is led to the estimate (Dine, Randall, and Thomas, 1996)

$$n_B = \frac{1}{M^n} \sin(\delta) \Phi_o^{n+3}, \quad (96)$$

which is also supported by numerical studies. Here,  $\Phi_o$  is determined by  $H \approx m_{3/2}$ , i.e.,  $\Phi_o^{2n+2} = m_{3/2}^2 M^{2n}$ .

<sup>12</sup>Again, these arise from nonrenormalizable terms in the effective action (Dine, Randall, and Thomas, 1995).

### E. The fate of the condensate

Of course, we do not live in a universe dominated by a coherent scalar field. In this section, we consider the fate of a homogeneous condensate. The following sections will deal with inhomogeneities, and the interesting array of phenomena to which they might give rise.

We have seen that a coherent field can be thought of as a collection of zero-momentum particles. These particles are long lived, since the particles to which they couple gain large mass in the flat direction. If there were no ambient plasma or other fields, the condensate would eventually decay. However, there are a number of effects which cause the condensate to disappear more rapidly, or to produce stable remnants. Precisely which is most important depends on a number of factors. Among the most important are the expansion rate, the dominant form of energy during this epoch, and the amplitude of oscillations.

It is impossible to survey all possibilities; indeed, it is likely that all of the possibilities have not yet been imagined. Instead, we will adopt the picture for inflation described in the previous section. The features of this picture are true of many models of inflation, but by no means all. We will suppose that the energy scale of inflation is  $E \sim 10^{15}$  GeV. We assume that inflation is due to a field, the inflaton  $I$ . The amplitude of the inflaton, just after inflation, we will take to be of order  $M'_p \approx 10^{18}$  GeV, the so-called reduced Planck mass. Correspondingly, we will take the mass of the inflaton to be  $m_I = 10^{12}$  GeV [so that  $m_I^2(M'_p)^2 \approx E^4$ ]. Correspondingly, the Hubble rate during inflation is of order  $H_I \approx E^2/M'_p \approx 10^{12}$  GeV.

After inflation ends, the inflaton oscillates about the minimum of its potential, much like the field  $\Phi$ , until it decays. We will suppose that the inflaton couples to ordinary particles with a rate suppressed by a single power of the Planck mass. Dimensional analysis then gives for the rough value of the inflaton lifetime:

$$\Gamma_I = \frac{m_I^3}{(M'_p)^2} \sim 1 \text{ GeV}. \quad (97)$$

The reheating temperature  $T_R$  can then be obtained by equating the energy density at the time when  $H \approx \Gamma_I$  to the energy density of the final plasma (Kolb and Turner, 1990):

$$T_R = T(t = \Gamma_I^{-1}) \sim (\Gamma_I M'_p)^{1/2} \sim 10^9 \text{ GeV}. \quad (98)$$

The decay of the inflaton is actually not sudden, but leads to a gradual reheating of the universe, as described, for example, by Kolb and Turner (1990). The temperature varies as a function of time as

$$T \approx [T_R^2 H(t) M'_p]^1/4. \quad (99)$$

As for the field  $\Phi$ , our basic assumption is that it attains a large value during inflation, in accord with Eq. (93). When inflation ends, we assume that the inflaton still dominates the energy density for a time, oscillating about its minimum; the universe is matter dominated

during this period. The field  $\Phi$  now oscillates about a time-dependent minimum given by Eq. (93). The minimum decreases in value with time. During this evolution, a baryon number develops classically. At  $H \sim m_{3/2}$ , the minimum of  $\Phi$  drops to zero and the baryon number freezes.

Eventually the condensate will decay, through a variety of processes. As we have indicated, the condensate can be thought of as a coherent state of  $\Phi$  particles. These particles—linear combinations of the squark and slepton fields—are unstable and will decay. However, for  $H \leq m_{3/2}$ , the lifetimes of these particles are much longer than in the absence of the condensate. The reason is that the fields to which  $\Phi$  couples have mass of order  $\Phi$ , and  $\Phi$  is large. In most cases, the most important process which destroys the condensate is what we might call evaporation: particles in the ambient thermal bath can scatter off of the particles in the condensate, leaving final states with only ordinary particles.

We can make a crude estimate for the reaction rate  $\Gamma_p$  as follows. Because the particles which couple directly to  $\Phi$  are heavy, interactions of  $\Phi$  with light particles must involve loops. So we include a loop factor in the amplitude, of order  $\alpha_2^2$ , the square of the running weak coupling, equal to  $\alpha_w^2$  at low mass scales. Because of the large masses, the amplitude is suppressed by  $\Phi$ . Finally, we need to square and multiply by the thermal density of scattered particles. This gives

$$\Gamma_p \sim \alpha_2^2 \pi \frac{1}{\Phi^2} (T_R^2 H M)^{3/4}. \quad (100)$$

The condensate will evaporate when this quantity is of order  $H$ . Since we know the time dependence of  $\Phi$ , we can solve for this time. One finds that equality occurs for  $H_I \sim 10^2 - 10^3$  GeV, if  $n = 1$ . In other words, for the case  $n = 1$ , the condensate evaporates shortly after the baryon number is created. For  $n$  in the range  $4 > n > 1$ , the evaporation occurs significantly later but before the decay of the inflaton. For  $n \geq 4$ , a slightly different analysis is required than that which follows.

The expansion of the universe is unaffected by the condensate as long as the energy density in the condensate,  $\rho_\Phi \sim m_\Phi^2 \Phi^2$ , is much smaller than that of the inflaton,  $\rho_I \sim H^2 M^2$ . Assuming that the  $\Phi$  mass satisfies  $m_\Phi \sim m_{3/2} \sim 0.1 - 1$  TeV, a typical supersymmetry breaking scale, one can estimate the ratio of the two densities at the time when  $H \sim m_{3/2}$  as

$$\frac{\rho_\Phi}{\rho_I} \sim \left( \frac{m_{3/2}}{M_p} \right)^{2/(n+1)}. \quad (101)$$

We are now in a position to calculate the baryon-to-photon ratio in this model. Given our estimate of the inflaton lifetime, the coherent motion of the inflaton still dominates the energy density when the condensate evaporates. The baryon number is the  $\Phi$  energy density just before evaporation divided by  $m_\Phi$  (assumed to be of order  $m_{3/2}$ ), while the inflaton number is  $\rho_I/M_I$ . So the baryon-to-inflaton ratio follows from Eq. (101). With the

assumption that the inflaton energy density is converted to radiation at the reheating temperature  $T_R$ , we obtain

$$\begin{aligned} \frac{n_B}{n_\gamma} &\sim \frac{n_B}{(\rho_I/T_R)} \sim \frac{n_B}{n_\Phi} \frac{T_R}{m_\Phi} \frac{\rho_\Phi}{\rho_I} \\ &\sim 10^{-10} \left( \frac{T_R}{10^9 \text{ GeV}} \right) \left( \frac{M'_p}{m_{3/2}} \right)^{(n-1)/(n+1)}. \end{aligned} \quad (102)$$

Clearly the precise result depends on factors beyond those indicated here explicitly, such as the precise mass of the  $\Phi$  particle(s). But as a rough estimate, it is rather robust. For  $n=1$ , it is in precisely the right range to explain the observed baryon asymmetry. For larger  $n$ , it can be significantly larger. While this may seem disturbing, it is potentially a significant virtue. Many supersymmetric models lead to creation of entropy at late times. For example, in string theory one expects the existence of other light fields ( $m \approx m_{3/2}$ ), known as moduli. These fields cause cosmological difficulties (Coughlan *et al.*, 1983), unless they reheat the universe to temperatures of order 10 MeV when they decay. Afterward nucleosynthesis can occur. These decays produce a huge amount of entropy, typically increasing the thermal energy of the universe by a factor of  $10^7$ . The baryon density is diluted by a corresponding factor. So in such a picture, it is necessary that the baryon number, prior to the moduli decay, should be of order  $10^{-3}$ . This is not the only cosmological model which requires such a large baryon number density.

There are many issues in the evolution of the condensate which we have not touched upon. One of the most serious is related to interactions with the thermal bath (Allahverdi, Campbell, and Ellis, 2000; Anisimov and Dine, 2001). In the case  $n=1$ , the potential minimum  $\Phi_0$  is not so large. The particles which  $\Phi$  couples acquire a mass of order  $\Phi_0$ , and they may be in chemical equilibrium. In this case, the  $\Phi$  particles decay much earlier. This typically leads to significant suppression of the asymmetry, and the viability of the Affleck-Dine mechanism depends on the precise values of the parameters.

Overall, then, there is a broad range of parameters for which the Affleck-Dine mechanism can generate a value for  $n_B/n_\gamma$  equal to or larger than that observed. This baryon number is generated long after inflation, so inflationary reheating does not provide any significant constraint. It can be large, allowing for processes which might generate entropy rather late.

## F. Inhomogeneities and the condensate

We have so far assumed that the condensate is homogeneous. But, as we will now show, under some circumstances the condensate is unstable to fragmentation. This appears to be related to another feature of theories with scalars: the possible existence of nontopological solitons. These can alter the picture of baryon-number generation, and could conceivably be dark matter candidates.

### 1. Stability and fragmentation

To analyze the stability of the condensate (Kusenko and Shaposhnikov, 1998), we write its field  $\varphi = \rho e^{i\Omega}$  in terms of its radial component  $\rho$  and a phase  $\Omega$ , both real functions of space-time. We are interested in the evolution of the scalar field in the small-VEV domain, where the baryon-number-violating processes are suppressed. We will also assume that the scalar potential preserves  $U(1)$  symmetry, depending only on the modulus of  $\varphi$ ,  $U(\varphi) = U(\rho)$ . We determine the classical equations of motion in the time-dependent spherically symmetric metric  $ds^2 = dt^2 - a^2(t) dr^2$ , where  $a(t)$  is the scale factor. From Eq. (68) we obtain the following dynamic equations for  $\Omega$  and  $\rho$ :

$$\ddot{\Omega} + 3H\dot{\Omega} - \frac{1}{a^2} \nabla^2 \Omega + \frac{2\dot{\rho}}{\rho} \dot{\Omega} - \frac{2}{a^2 \rho} \nabla \Omega \cdot \nabla \rho = 0, \quad (103)$$

$$\ddot{\rho} + 3H\dot{\rho} - \frac{1}{a^2} \nabla^2 \rho - \dot{\Omega}^2 \rho + \frac{1}{a^2} |\nabla \Omega|^2 \rho + \frac{\partial U}{\partial \rho} = 0. \quad (104)$$

The Hubble rate, again, is  $H = \dot{a}/a$ ; it is equal to  $t^{-2/3}$  or  $t^{-1/2}$  for a matter- or radiation-dominated universe, respectively.

From the equations of motion (103) and (104), one can derive the equations for small perturbations  $\delta\Omega$  and  $\delta\rho$ :

$$\begin{aligned} \delta\ddot{\Omega} + 3H\delta\dot{\Omega} - \frac{1}{a^2(t)} \nabla^2 \delta\Omega + \frac{2\dot{\rho}}{\rho} \delta\dot{\Omega} \\ + \frac{2\dot{\Omega}}{\rho} \delta\rho - \frac{2\dot{\rho}\dot{\Omega}}{\rho^2} \delta\rho = 0, \end{aligned} \quad (105)$$

$$\delta\ddot{\rho} + 3H\delta\dot{\rho} - \frac{1}{a^2(t)} \nabla^2 \delta\rho - 2\dot{\rho}\dot{\Omega} \delta\Omega + U'' \delta\rho - \dot{\Omega}^2 \delta\rho = 0. \quad (106)$$

To examine the stability of a homogeneous solution  $\varphi(x, t) = \varphi(t) \equiv \rho(t) e^{i\Omega(t)}$ , let us consider a perturbation  $\delta\rho$ ,  $\delta\Omega \propto e^{S(t) - ik\vec{x}}$  and look for growing modes,  $\text{Re } \alpha > 0$ , where  $\alpha = dS/dt$ . The value of  $k$  is the spectral index in the comoving frame and is redshifted with respect to the physical wave number  $\tilde{k} = k/a(t)$  in the expanding background. Of course, if an instability develops, the linear approximation is no longer valid. However, we assume that the wavelength of the fastest-growing mode sets the scale for the high and low density domains that eventually evolve into  $Q$  balls (defined below). This assumption can be verified *a posteriori* by comparison with a numerical solution of the corresponding partial differential equations (103) and (104), where both large and small perturbations are taken into account.

The dispersion relation follows from the equations of motion:

$$\begin{aligned} \left[ \alpha^2 + 3H\alpha + \frac{k^2}{a^2} + \frac{2\dot{\rho}}{\rho} \alpha \right] \left[ \alpha^2 + 3H\alpha + \frac{k^2}{a^2} - \dot{\Omega}^2 + U''(\rho) \right] \\ + 4\dot{\Omega}^2 \left[ \alpha - \frac{\dot{\rho}}{\rho} \right] \alpha = 0. \end{aligned} \quad (107)$$

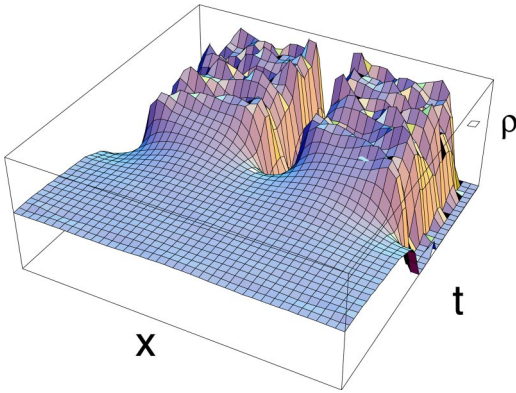


FIG. 5. (Color in online edition) The charge density per comoving unit volume in (1+1) dimensions for a sample potential analyzed numerically during the period when the spatially homogeneous condensate breaks up into high- and low-density domains. Two domains with high charge density are expected to form  $Q$  balls.

If  $[\dot{\Omega}^2 - U''(\rho)] > 0$ , there is a band of growing modes that lies between the two zeros of  $\alpha(k)$ ,  $0 < k < k_{\max}$ , where

$$k_{\max}(t) = a(t) \sqrt{\dot{\Omega}^2 - U''(\rho)}. \quad (108)$$

This simple linear analysis shows that when the condensate is “overloaded” with charge, that is, when  $\omega(t) = \dot{\Omega}$  is larger than the second derivative of the potential, an instability develops. Depending on how  $k_{\max}(t)$  varies with time, the modes in the bands of instability may or may not have time to develop fully.

Numerical analyses (Kasuya and Kawasaki, 2000a, 2000b, 2001; Enqvist *et al.*, 2001), which can trace the evolution of unstable modes beyond the linear regime, have shown that fragmentation of the condensate is a generic phenomenon (see Fig. 5). Numerically one can also study the stability of rapidly changing solutions, hence relaxing the adiabatic assumption made above. This aspect is relevant to the cases where the baryon number density is small and the radial component of the condensate  $\rho(t)$  exhibits an oscillatory behavior, changing significantly on small time scales. An interesting feature of this nonadiabatic regime is that both baryon and antibaryon lumps may form as a result of fragmentation (Enqvist *et al.*, 2001).

## 2. Lumps of scalar condensate: $Q$ balls

The most familiar soliton solutions of nonlinear field theories, such as magnetic monopoles and vortices, can be uncovered by topological arguments. However, field theories with scalar fields often admit nontopological

solitons,<sup>13</sup> called  $Q$  balls, which may be stable or may decay into fermions (Cohen, Coleman, Georgi, and Manohar, 1986).  $Q$  balls appear when a complex scalar field  $\varphi$  carries a conserved charge with respect to some global U(1) symmetry. In supersymmetric extensions of the Standard Model, squarks and sleptons carry the conserved baryon and lepton numbers, and can form  $Q$  balls.

Let us consider a field theory with a scalar potential  $U(\varphi)$  which has a global minimum  $U(0) = 0$  at  $\varphi = 0$ . Let  $U(\varphi)$  have an unbroken global U(1) symmetry at the global minimum, depending only on the modulus of  $\varphi$ . We will look for solutions of the classical equations by minimizing the energy

$$E = \int d^3x \left[ \frac{1}{2} |\dot{\varphi}|^2 + \frac{1}{2} |\nabla\varphi|^2 + U(\varphi) \right] \quad (109)$$

subject to the constraint that the configuration has a definite charge  $Q$ ,

$$Q = \frac{1}{2i} \int \varphi^* (\partial_t - \bar{\partial}_t) \varphi d^3x. \quad (110)$$

To describe the essential features of  $Q$  balls in a simple way, we will follow Coleman (1985) and use a thin-wall ansatz for the  $Q$  ball,

$$\varphi(x, t) = e^{i\omega t} \bar{\phi}(x), \quad (111)$$

where  $\bar{\phi}$  is the step function of height  $\phi_o$  extending out to a radius  $R$ ,

$$\bar{\phi}(\vec{x}) = \begin{cases} 0, & |\vec{x}| > R \\ \phi_o, & |\vec{x}| \leq R. \end{cases} \quad (112)$$

For the real solution, the field varies rapidly between the two regions, changing on a scale of order the Compton wavelength of the  $\varphi$  particle.

Assuming that  $Q$  is large, let us neglect the gradient terms which are relevant only for the wall energy. Then the global charge and the energy of the field configuration (111) and (112) are given by

$$Q = \omega \phi_o^2 V, \quad (113)$$

where  $V$  is the volume  $V = (4/3)\pi R^3$ , and

$$E = \frac{1}{2} \omega^2 \phi_o^2 V + U(\phi_o) V = \frac{1}{2} \frac{Q^2}{V \phi_o^2} + V U(\phi_o). \quad (114)$$

We now minimize  $E$  with respect to  $V$ , obtaining

$$E = \sqrt{\frac{2U(\phi_o)}{\phi_o^2}}. \quad (115)$$

<sup>13</sup>For general references, see Rosen (1968a, 1968b); Friedberg, Lee, and Sirlin (1976); Coleman (1985); Lee and Pang (1992). For additional references on  $Q$  balls, see Kusenko (1997b); Dvali, Kusenko, and Shaposhnikov (1998); Multamaki and Vilja (2000, 2002); Correia and Schmidt (2001); Hisano *et al.* (2001); Multamaki (2001); Allahverdi *et al.* (2002).

It remains to minimize the energy with respect to variations of  $\phi_o$ . A nontrivial minimum exists as long as

$$U(\varphi)/\varphi^2 = \min, \quad \text{for } \varphi = \phi_o > 0. \quad (116)$$

If this condition is satisfied, a  $Q$ -ball solution exists.

We made a number of assumptions above. These assumptions can be avoided in a slightly more involved derivation using the method of Lagrange multipliers (Kusenko, 1997a). We want to minimize

$$\mathcal{E}_\omega = E + \omega \left[ Q - \frac{1}{2i} \int \varphi^* (\bar{\partial}_t - \partial_t) \varphi d^3x \right], \quad (117)$$

where  $\omega$  is a Lagrange multiplier. We will see below that  $\omega = \dot{\Omega}$ . Variations of  $\varphi(x, t)$  and those of  $\omega$  can now be treated independently, the usual advantage of the Lagrange method.

One can rewrite Eq. (117) as

$$\begin{aligned} \mathcal{E}_\omega = & \int d^3x \frac{1}{2} \left| \frac{\partial}{\partial t} \varphi - i\omega \varphi \right|^2 + \int d^3x \left[ \frac{1}{2} |\nabla \varphi|^2 + \hat{U}_\omega(\varphi) \right] \\ & + \omega Q, \end{aligned} \quad (118)$$

where

$$\hat{U}_\omega(\varphi) = U(\varphi) - \frac{1}{2} \omega^2 \varphi^2. \quad (119)$$

We are looking for a solution that extremizes  $\mathcal{E}_\omega$ , while all the physical quantities, including the energy  $E$  are time independent. Only the first term in Eq. (118) appears to depend on time explicitly, but it is positive definite and hence it should vanish at the minimum. To minimize this contribution to the energy, one must therefore require

$$\varphi(x, t) = e^{i\omega t} \varphi(x), \quad (120)$$

where  $\varphi(x)$  is real and independent of time. We have thus derived Eq. (111). For this solution, Eq. (110) yields

$$Q = \omega \int \varphi^2(x) d^3x. \quad (121)$$

It remains to find an extremum of the functional

$$\mathcal{E}_\omega = \int d^3x \left( \frac{1}{2} |\nabla \varphi(x)|^2 + \hat{U}_\omega[\varphi(x)] \right) + \omega Q, \quad (122)$$

with respect to  $\omega$  and the variations of  $\varphi(x)$  independently. We can first minimize  $\mathcal{E}_\omega$  for a fixed  $\omega$ , while varying the shape of  $\varphi(x)$ . If this were an actual potential for a scalar field in three dimensions, one would have the possibility of tunneling between the zero-energy configuration at the origin and possible lower energy configurations at nonzero  $\phi$ . Tunneling, in the semiclassical approximation, is described by the bounce  $\bar{\varphi}_\omega(x)$ , the solution of the classical equations which asymptotically approaches the “false vacuum” at the origin (Callan and Coleman, 1977; Coleman, 1977). The first term in Eq. (122) is nothing but the three-dimensional Euclidean action  $S_3[\bar{\varphi}_\omega(x)]$  of this bounce solution. This is a very useful correspondence. In particular, the condition for the existence of a solution is

simply a corollary: as long as  $\hat{U}_\omega(\varphi)$  has a minimum below zero, the bounce exists, and so does the  $Q$  ball,  $\varphi(x, t) = e^{i\omega t} \bar{\varphi}(x)$ .

The bounce, and hence the  $Q$  ball, exist if there exists a value of  $\omega$  for which the potential  $\hat{U}_\omega(\varphi)$  has both a local minimum at  $\varphi=0$  and a global minimum at some other value of  $\varphi$ . This condition can be rephrased without reference to  $\omega$  and is identical to the condition (116) found in the thin-wall approximation. The corresponding effective potential is  $\hat{U}_{\omega_0}(\varphi)$ , where  $\omega_0 = \sqrt{2U(\varphi_0)/\varphi_0^2}$ . The potential has two degenerate minima, at  $\varphi=0$  and  $\varphi=\varphi_0$ . The existence of the bounce solution  $\bar{\varphi}_\omega(x)$  for  $\omega_0 < \omega < U''(0)$  follows from the fact that  $\hat{U}_\omega(\varphi)$  has a negative global minimum in addition to the local minimum at the origin (Coleman, 1977; Coleman *et al.*, 1978). Coleman *et al.* (1978) also showed that the solution is spherically symmetric:  $\bar{\varphi}(\vec{x}) = \bar{\varphi}(r)$ ,  $r = |\vec{x}|$ .

The soliton we want to construct is precisely this bounce at the value of  $\omega$  that minimizes  $\mathcal{E}_\omega$ . The last step is to find an extremum of

$$\mathcal{E}_\omega = S_3[\bar{\varphi}_\omega(x)] + \omega Q \quad (123)$$

with respect to  $\omega$ , which can be proven to exist (Kusenko, 1997a). Finally, the soliton is of the form (120), with  $\omega$  that minimizes  $\mathcal{E}_\omega$  in Eq. (123).

Having obtained the solution, one can compute its energy or mass. For a finite  $\varphi_0$  in Eq. (116), in the limit of large  $Q$ , the  $Q$  ball has a thin wall, and its mass is given by

$$M(Q) = \omega_0 Q. \quad (124)$$

As discussed earlier, supersymmetric extensions of the Standard Model have scalar potentials with flat directions lifted only by supersymmetry breaking terms.  $Q$  balls may form with a light scalar field  $\varphi$  that corresponds to that flat direction. If the potential is constant for large  $\varphi$ ,  $U(\varphi) = U_0 = \text{const}$ , the minimum in Eq. (116) is achieved for  $\varphi_0 = \infty$ . In this case,

$$M(Q) = U_0^{1/4} Q^{3/4}. \quad (125)$$

More generally, if the potential grows slower than  $\varphi^2$ , i.e.,  $U(\varphi) \propto \varphi^p$ ,  $p < 2$ , condition (116) is not satisfied at any finite value of  $\phi_o$ , and

$$M(Q) \sim U_0^{1/4} Q^{(3-p/2)/(4-p)}. \quad (126)$$

It is important in what follows that the mass per unit charge is not a constant, but is a decreasing function of the total global charge  $Q$ . There is a simple reason why the soliton mass is not proportional to  $Q$ . Since  $U(\varphi)/\varphi^2$  has no minimum in the supersymmetric theories, the scalar VEV can grow as far as the derivative terms allow it. When the next unit of charge is added, the  $Q$  ball increases in size, which allows the scalar VEV to increase as well. Hence the larger the charge, the greater is the VEV, and the smaller is energy per unit charge.



### 3. Affleck-Dine $Q$ balls

The condition (116) suggests that slowly growing potentials, of the sort that arise in the flat directions of the MSSM, are likely sources of  $Q$  balls.  $Q$  balls can develop along flat directions that carry nonzero baryon number, lepton number, or both. Each flat direction can be parametrized by a gauge-invariant field carrying these global quantum numbers. So the discussion of gauge singlet fields of the previous section also applies to baryonic and leptonic  $Q$  balls in the MSSM. This statement may seem surprising, since all scalar baryons in the MSSM transform nontrivially under the gauge group. Although scalars with gauge interactions can also make  $Q$  balls (Lee *et al.*, 1989), in the case of the MSSM the color structure of large  $Q$  balls is rather simple (Kusenko, Shaposhnikov, and Tinyakov, 1998). If a  $Q$ -ball VEV points along a flat direction, its scalar constituents form a colorless combination (otherwise, that direction would not be flat because of nonvanishing  $D$  terms).

In one proposal for the origin of supersymmetry breaking, “gauge-mediated” supersymmetry breaking, the flat directions are lifted by potentials which grow quadratically for small values of the fields, and then level off to a logarithmic plateau at larger  $\phi$  (Guidice and Rattazzi, 1999).  $Q$  balls in such a potential have masses given by Eq. (125). In another proposal, “gravity-mediated” SUSY breaking, the potentials which arise from supersymmetry breaking grow roughly quadratically even for very large VEV (Nilles, 1984). See the review by Dine, Randall, and Thomas (1996) for further details. Whether  $Q$  balls exist is thus a detailed, model-dependent question.  $Q$  balls in these potentials have masses proportional to the first power of  $Q$ .

By construction,  $Q$  balls are stable with respect to decay into scalars. However, they can decay by emitting fermions (Cohen, Coleman, Georgi, and Manohar, 1986). If the  $Q$  ball has zero baryon number, it can decay by emitting light neutrinos (Cohen, Coleman, Georgi, and Manohar, 1986). However, if a baryonic  $Q$  ball, called a  $B$  ball, develops along a flat direction, it can also be stable with respect to decay into fermions. Stability requires that the baryon number be large enough that the mass of the  $Q$  ball  $M_B$  is below the mass of  $B$  separated baryons. For a  $Q$  ball in a flat potential of height  $M_S$ , the mass per unit baryon number has the order of magnitude

$$\frac{M(Q_B)}{Q_B} \sim M_S Q^{-1/4}. \quad (127)$$

Models of gauge-mediated supersymmetry breaking produce flat potentials with  $M_S \sim 1 - 10$  TeV. If the mass per baryon number is less than the proton mass  $m_p$ , then the  $Q$  ball is entirely stable because it does not have enough energy to decay into a collection of nucleons with the same baryon number. This condition translates into a lower bound on  $Q_B$ :

$$\frac{M(Q_B)}{Q_B} < 1 \text{ GeV} \Rightarrow Q_B \gg \left( \frac{M_S}{1 \text{ GeV}} \right)^4 \gtrsim 10^{16}. \quad (128)$$

To make contact with observational bounds, we will also need the radius of the  $Q$  ball (Arafune *et al.*, 2000),

$$R = \frac{1}{M_S} Q^{-1/4}. \quad (129)$$

### 4. Dark matter in the form of stable $B$ balls

Stable  $Q$  balls that form from an Affleck-Dine condensate are a viable candidate for dark matter. Even if they are unstable, their decay can produce neutralinos at late times, when these neutralinos are out of equilibrium. One way or another, some dark matter can arise from Affleck-Dine baryogenesis. Moreover, since both the ordinary matter and the dark matter have the same origin in the Affleck-Dine scenario, one can try to explain why their amounts in the universe are comparable (Laine and Shaposhnikov, 1998; Enqvist and McDonald, 1999; Banerjee and Jedamzik, 2000; Fujii and Yanagida, 2002). Of the other dark-matter candidates that have been considered, most were weakly interacting particles, and for good reason. If dark-matter particles have strong (relative to their mass) interactions with matter, these interactions might facilitate their loss of momentum and angular momentum, forcing them into the galactic disks, along with ordinary matter. But astronomical observations show that the dark matter forms spherical halos about galaxies, not disks.

These  $Q$  balls are made of squarks and interact strongly with ordinary matter via QCD. However, if they are as heavy as the calculations show they are, the strong interactions are not enough to force dark-matter  $Q$  balls to settle into the galactic disks. Analyses of  $Q$ -ball formation and partial evaporation allow one to relate the amounts of ordinary matter and dark matter. The observed ratio corresponds to  $Q$  balls with baryon number of about  $10^{26 \pm 2}$ , which is in agreement with the expected  $Q$ -ball size from numerical simulations (Kasuya and Kawasaki, 2000a, 2001), as well as with the current experimental bounds summarized by Arafune *et al.* (2000) (see Fig. 6). A  $B$  ball with baryon number  $10^{26}$  is so heavy that it could pass through ordinary stars with only a small change in its velocity [ $(\delta v/v) \sim 10^{-5}$ ]. Hence despite the strong interactions,  $B$  balls make a good dark-matter candidate.

Since  $B$  balls have lower mass-to-baryon ratio than ordinary nuclear matter, interactions of  $B$  balls with ordinary matter result in numerous events with energy releases similar to proton decays (Kusenko, Kuzmin, Shaposhnikov, and Tinyakov, 1998). Hence a  $Q$  ball passing through a detector would produce a spectacular signature. However, the flux is very low, and the strongest limits come from the largest detectors, e.g., Super-Kamiokande (see Fig. 6). Some astrophysical bounds have been considered (Kusenko, Shaposhnikov, Tinyakov, and Tkachev, 1998) but they do not yield very strong constraints. In addition to the existing limits discussed in Arafune *et al.* (2000), future experiments, such as ANTARES, Ice Cube, etc., may be able to detect

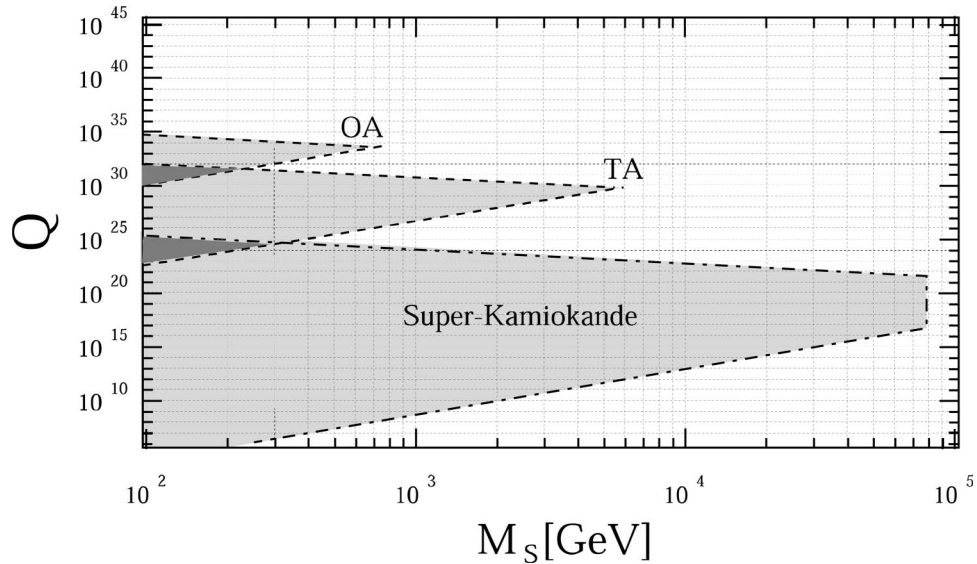


FIG. 6. Limits on neutral dark-matter  $Q$  balls from large detector searches (Arafune *et al.*, 2000). The excluded parameter space is shown in dark as a function of the  $Q$ -ball charge and the supersymmetry-breaking scale  $M_S$ . The authors assume an interaction cross section  $\sigma \sim R^2$  [see Eq. (129)], and a flux consistent with the observed dark-matter density. The experiments shown are OA (Niessen *et al.*, 1999), TA (Scari, 1999), and Super-Kamiokande. The older experiments BAKSAN (Alexeyev *et al.*, 1982) and Kamiokande (Kajita *et al.*, 1985) give limits that are inside the Super-Kamiokande boundaries.

dark-matter  $B$  balls, or rule out the values of  $Q$  that correspond to the observed amount of dark matter.

We note that, although  $Q$  balls are always present in the spectrum of any SUSY extension of the Standard Model, their production in the early universe requires the formation of an Affleck-Dine condensate followed by its fragmentation. Stable  $Q$  balls are too large to form in thermal plasma by accretion (Griest and Kolb, 1989; Postma, 2002). In this sense, an observation of stable dark-matter  $Q$  balls would be evidence of the Affleck-Dine process.

##### 5. Dark matter from unstable $B$ balls

If supersymmetry breaking is mediated by gravity,  $Q$  balls are not stable since they can decay into fermions. However,  $Q$ -ball decay into fermions is a slow process because the fermions quickly fill up the Fermi sea inside the  $Q$  ball, and further decays are limited by the rate of fermion evaporation through the surface. The rate of  $Q$ -ball decay is therefore suppressed compared to that of free scalar particles by the surface-to-volume ratio (Cohen, Coleman, Georgi, and Manohar, 1986). In a typical model, unstable baryonic  $Q$  balls from the Affleck-Dine condensate decay when the temperature is as low as a GeV.

The lightest supersymmetric particles (LSP) are among the decay products of  $Q$  balls.  $B$  balls can decay and produce dark matter in the form of neutralinos at a time when they are out of equilibrium (Enqvist and McDonald, 1998a, 1998b, 1999; Fujii and Hamaguchi, 2002a). This presents another possibility for producing dark matter from the Affleck-Dine condensate and relating its abundance to that of ordinary matter. The re-

quirement that neutralinos not overly close the universe constrains the parameter space of the MSSM (Fujii and Hamaguchi, 2002a, 2002b).

If the LSP's do not annihilate, the ratio of ordinary matter to dark matter is simply (Enqvist and McDonald, 1999)

$$\Omega_{\text{matter}}/\Omega_{\text{LSP}} \sim f^{-1} \left( \frac{m_p}{m_\chi} \right) \left( \frac{n_B}{n_\chi} \right), \quad (130)$$

where  $m_\chi$  and  $n_\chi$  are the mass and number density of the LSP, and  $f$  is the fraction of the condensate trapped in  $Q$  balls. If  $f \sim 10^{-3}$ , this ratio is acceptable.

However, numerical simulations (Kasuya and Kawasaki, 2000a) and some analytical calculations (McDonald, 2001) indicate that, in a wide class of Affleck-Dine models, practically all the baryon number would be trapped in  $Q$  balls, that is  $f \sim 1$ . If that is the case, the LSP would overly close the universe, according to Eq. (130). A solution, proposed by Fujii and Hamaguchi (2002a, 2002b), is to eliminate the unwanted excess of neutralinos by using an LSP with a higher annihilation cross section. The LSP in the MSSM is an admixture of several neutral fermions. Depending on the parameters in the mass matrix, determined largely by the soft SUSY breaking terms, the LSP can be closely aligned with the *binos* [one of the SUSY partners of the  $SU(2) \times U(1)$  gauge bosons], the *Higgsinos* (the fermion counterpart of the Higgs boson), or with one of the other weak eigenstates. The traditional freezeout scenario for LSP production favors the binolike LSP (Jungman, Kamionkowski, and Griest, 1996). However, according to Fujii and Hamaguchi (2002a, 2002b), SUSY dark matter produced from the Affleck-Dine process has to be in the form of a Higgsino-like LSP (see Fig. 7). In this case, the ratio of matter densities is (Fujii and Yanagida, 2002)

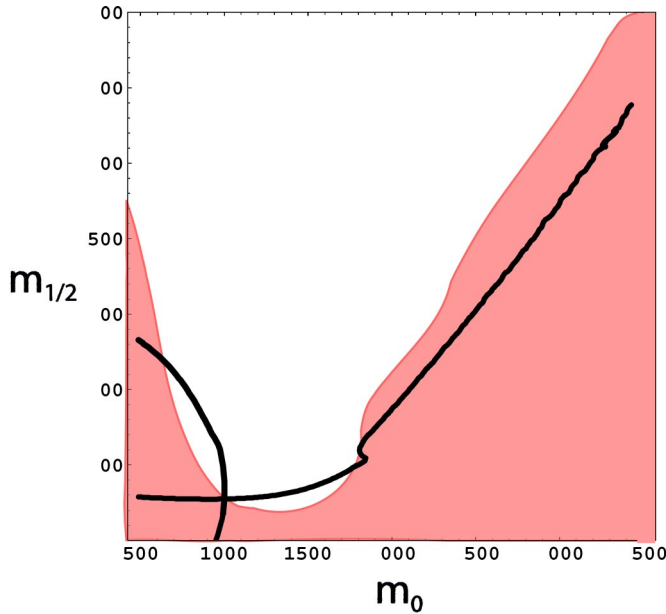


FIG. 7. (Color in online edition) The allowed range of parameters for nonthermal LSP dark matter from the Affleck-Dine process is very different from that in the standard freezeout case. The shaded region above the solid lines represents allowed parameters for nonthermal LSP dark matter in the minimal supergravity model with  $\tan\beta=40$ ;  $m_{1/2}$  is the gaugino mass (Fujii and Hamaguchi, 2002a, 2002b).

$$\Omega_{\text{matter}}/\Omega_{\text{LSP}}=10^{3-4}\left(\frac{m_{\Phi}^2}{\langle\sigma v\rangle_{\chi}}\right)^{-1}\left(\frac{m_p}{m_{\chi}}\right)\delta_{\text{CP}}, \quad (131)$$

where  $\delta_{\text{CP}}\sim 0.1$  is the effective  $CP$  violating phase of the Affleck-Dine condensate. For a Higgsino-like LSP, which has an annihilation rate constant  $\langle\sigma v\rangle_{\chi}\sim 10^{-(7-8)}\text{ GeV}^{-2}$ , this yields an acceptable result.

There are important implications for both direct dark matter searches and the collider searches for SUSY. First, the parameter space of the MSSM consistent with LSP dark matter is very different, depending on whether the LSP's froze out of equilibrium (Arnowitz and Dutta, 2002) or were produced from the evaporation of Affleck-Dine  $B$  balls (Fujii and Yanagida, 2002). Second, Higgsino and bino LSP's interact differently with matter, so the sensitivity of direct dark-matter searches also depends on the type of the LSP.

If supersymmetry is discovered, one will be able to determine the properties of the LSP experimentally. This will in turn provide some information on the how the dark-matter SUSY particles could be produced. The discovery of a Higgsino-like LSP would be evidence in favor of Affleck-Dine baryogenesis.

#### IV. CONCLUSIONS

The origin of the matter-antimatter asymmetry is one of the great questions in cosmology. Yet we can obtain only limited information about the events which gave rise to the baryon asymmetry by looking at the sky. Filling out the picture requires a deeper understanding of

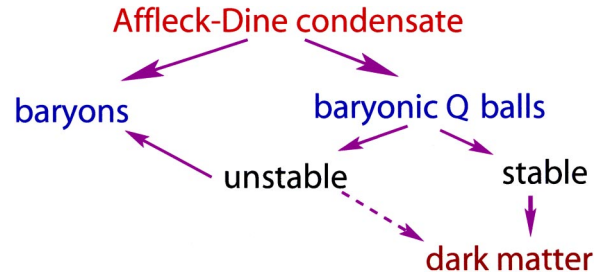


FIG. 8. (Color in online edition) Fate of the Affleck-Dine condensate. The Affleck-Dine condensate may decay perturbatively and produce baryons. Depending on the mode of supersymmetry breaking, the condensate may fragment into  $Q$  balls. If the  $Q$  balls are stable, they contribute to the present dark matter. If they are unstable, they produce additional baryons, as well as dark matter in the form of LSP's.

fundamental physical law. One elegant possibility, that the Standard Model produced the baryon number near the electroweak scale, is ruled out decisively by the LEP bounds on the Higgs mass. This is a bittersweet conclusion: while one has to give up an elegant scenario, it would be perhaps the strongest evidence yet for physics beyond the Standard Model—a precursor of future discoveries.

Supersymmetry is widely regarded as a prime candidate for such new physics. Theoretical arguments in favor of supersymmetry are based on the naturalness of the scale hierarchy, the success of coupling unification in supersymmetric theories, and the nearly ubiquitous role of supersymmetry in string theory. The upcoming LHC experiments will put this hypothesis to a definitive test. If low-energy supersymmetry exists, there are several ways in which it might play the crucial role in baryogenesis. It could conceivably revive the electroweak baryogenesis scenario. However, the phase transition in the MSSM is only slightly stronger than that in the Standard Model; a noticeable improvement forces one into a narrow corner of the MSSM parameter space, which may soon be ruled out.

But supersymmetry opens a completely new and natural avenue for baryogenesis (see Fig. 8). If inflation took place in the early universe, for which we have an increasing body of evidence, then formation of an Affleck-Dine condensate and subsequent generation of some baryon asymmetry is natural. In a wide class of models this process produces the observed baryon asymmetry. Perhaps more striking is that the process can lead to very large baryon asymmetries. This may be important in many cosmological proposals where one produces substantial entropy at late times.

Finally, the Affleck-Dine process can produce dark matter, either in the form of stable SUSY  $Q$  balls, or in the form of a thermally or nonthermally produced LSP's. There are hints that the comparable magnitudes of matter and dark-matter densities may find its explanation in the same process as well. If supersymmetry is discovered, given the success of inflation theory, the Affleck-Dine scenario will appear quite plausible.

Other indications that the Affleck-Dine process took place in the early universe may come from detection of dark matter. One of the great attractions of supersymmetry is that it can naturally account for the dark matter. The lightest supersymmetric particle (LSP) is typically stable, and is produced with an abundance in the needed range if the supersymmetry-breaking scale is of order of hundreds of GeV. Its precise contribution to the energy density of the universe depends on its annihilation cross section and mass. A combination of accelerator limits and cosmology presently permits LSP dark matter for a range of parameters. In this range, the LSP, which is an admixture of several states, must be principally bino if it is produced in the standard freezeout scenario. However, if experiments determine that the LSP is primarily a Higgsino, this kind of dark matter could only arise from nonthermal production of the LSP from a fragmented Affleck-Dine condensate (Fujii and Hamaguchi, 2002a, 2002b; Fujii and Yanagida, 2002). Therefore, although a standard binolike LSP is not inconsistent with the Affleck-Dine scenario, a Higgsino-like LSP would provide a strong evidence in its favor. Likewise, a detection of stable baryonic  $Q$  balls would be a definitive confirmation that an Affleck-Dine condensate formed in the early universe and fragmented into  $B$  balls. Since stable SUSY  $Q$  balls must be large, we know of no other cosmological scenario that could lead to their formation.

Among other possibilities for baryogenesis, leptogenesis is also quite plausible. The discovery of neutrino mass, perhaps associated with a rather low scale of new physics, certainly gives strong support to this possibility. The questions of what scales for this physics might be compatible with inflation, and what implications this might have for the origin of neutrino mass are extremely important. Some pieces of the picture will be accessible to experiment, but many of the relevant parameters, including the relevant  $CP$  violation, reside at a very high scale. Perhaps, in a compelling theory of neutrino flavor, some of these questions can be answered.

Future experimental searches for supersymmetry, combined with the improving cosmological data on CMBR and dark matter, will undoubtedly shed further light on the origin of baryon asymmetry and will provide insight into both particle physics and cosmology. The study of the baryon asymmetry has already provided a compelling argument for new physics, and holds great promise of new and exciting discoveries in the future.

## ACKNOWLEDGMENTS

This work was supported in part by the U.S. Department of Energy. M.D. thanks for hospitality the Weizmann Institute, where he presented lectures on which parts of this article are based.

## REFERENCES

- Affleck, I., and M. Dine, 1985, Nucl. Phys. B **249**, 361.  
 Ahmad, Q. R., *et al.*, (SNO Collaboration), 2002, Phys. Rev. Lett. **89**, 011301.  
 Albrecht, A., and P. J. Steinhardt, 1982, Phys. Rev. Lett. **48**, 1220.  
 Alexeyev, E. N., *et al.*, 1982, Lett. Nuovo Cimento Soc. Ital. Fis. **35**, 413.  
 Allahverdi, R., B. A. Campbell, and J. R. Ellis, 2000, Nucl. Phys. B **579**, 355.  
 Allahverdi, R., A. Mazumdar, and A. Ozipineci, 2002, Phys. Rev. D **65**, 125003.  
 Altarelli, G., F. Feruglio, and I. Masina, 2000, J. High Energy Phys. **11**, 40.  
 Anisimov, A., and M. Dine, 2001, Nucl. Phys. B **619**, 729.  
 Arafune, J., T. Yoshida, S. Nakamura, and K. Ogure, 2000, Phys. Rev. D **62**, 105013.  
 Arnold, P., and O. Espinosa, 1993, Phys. Rev. D **47**, 3546; **50**, 6662(E) (1994).  
 Arnold, P., and L. McLerran, 1988, Phys. Rev. D **37**, 1020.  
 Arnold, P., D. Son, and L. G. Yaffe, 1997, Phys. Rev. D **55**, 6264.  
 Arnold, P., D. Son, and L. G. Yaffe, 2000, Phys. Rev. D **62**, 023512.  
 Arnowitt, R., and B. Dutta, 2002, e-print hep-ph/0211042.  
 Babu, K. S., J. C. Pati, and F. Wilczek, 2000, Nucl. Phys. B **566**, 33.  
 Bajc, B., P. F. Perez, and G. Senjanovic, 2002, Phys. Rev. D **66**, 075005.  
 Bagnasco, J. E., and M. Dine, 1993, Phys. Lett. B **303**, 308.  
 Banerjee, R., and K. Jedamzik, 2000, Phys. Lett. B **484**, 278.  
 Banks, T., L. J. Dixon, D. Friedan, and E. Martinec, 1988, Nucl. Phys. B **299**, 613.  
 Barr, S. M., G. Segrè, and H. A. Weldon, 1979, Phys. Rev. D **20**, 2494.  
 Belotsky, K. M., Y. A. Golubkov, M. Y. Khlopov, R. V. Konoplich, and A. S. Sakharov, 1998, e-print astro-ph/9807027.  
 Bennett, C. L., *et al.*, (WMAP Collaboration), 2003, e-print astro-ph/0302207.  
 Bodeker, D. D., P. John, M. Laine, and M. G. Schmidt, 1997, Nucl. Phys. B **497**, 387.  
 Burles, S., K. M. Nollett, and M. S. Turner, 2001a, Phys. Rev. D **63**, 063512.  
 Burles, S., K. M. Nollett, and M. S. Turner, 2001b, Astrophys. J. Lett. **552**, L1.  
 Callan, C. G., and S. Coleman, 1977, Phys. Rev. D **16**, 1762.  
 Carena, M., and H. M. Haber, 2003, Prog. Part. Nucl. Phys. **50**, 63.  
 Carena, M., M. Quiros, M. Seco, and C. E. Wagner, 2003, Nucl. Phys. B **650**, 24.  
 Carena, M., M. Quiros, and C. E. Wagner, 1998, Nucl. Phys. B **524**, 3.  
 Chang, D., W. F. Chang, and W. Y. Keung, 2002, e-print hep-ph/0205084.  
 Cline, J. M., 2000, Pramana, J. Phys. **54**, 1.  
 Cline, J. M., 2002, e-print hep-ph/0201286.  
 Cline, J. M., M. Joyce, and K. Kainulainen, 1998, Phys. Lett. B **417**, 79; **448**, 321(E) (1999).  
 Cline, J. M., M. Joyce, and K. Kainulainen, 2000, J. High Energy Phys. **07**, 18.  
 Cline, J. M., and G. D. Moore, 1998, Phys. Rev. Lett. **81**, 3315.  
 Cohen, A. G., S. R. Coleman, H. Georgi, and A. Manohar, 1986, Nucl. Phys. B **272**, 301.  
 Cohen, A. G., A. De Rujula, and S. L. Glashow, 1998, Astrophys. J. **495**, 539.  
 Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1991, Phys. Lett. B **263**, 86.

- Cohen, A. G., D. B. Kaplan, and A. E. Nelson, 1993, *Annu. Rev. Nucl. Part. Sci.* **43**, 27.
- Coleman, S., 1977, *Phys. Rev. D* **15**, 2929.
- Coleman, S., 1985, *Nucl. Phys. B* **262**, 263.
- Coleman, S., 1989, *Aspects of Symmetry* (Cambridge University Press, Cambridge).
- Coleman, S., V. Glaser, and A. Martin, 1978, *Commun. Math. Phys.* **58**, 211.
- Copeland, E. J., D. Lyth, A. Rajantie, and M. Trodden, 2001, *Phys. Rev. D* **64**, 043506.
- Cornwall, J. M., D. Grigoriev, and A. Kusenko, 2001, *Phys. Rev. D* **64**, 123518.
- Cornwall, J. M., and A. Kusenko, 2000, *Phys. Rev. D* **61**, 103510.
- Correia, F. Paccetti, and M. G. Schmidt, 2001, *Eur. Phys. J. C* **21**, 181.
- Coughlan, G. D., W. Fischler, E. W. Kolb, S. Raby, and G. G. Ross, 1983, *Phys. Lett.* **131B**, 59.
- Csikor, F., Z. Fodor, P. Hegedus, A. Jakovac, S. D. Katz, and A. Piroth, 2000, *Phys. Rev. Lett.* **85**, 932.
- Csikor, F., Z. Fodor, and J. Heitger, 1999, *Phys. Rev. Lett.* **82**, 21.
- Davies, A. T., C. D. Froggatt, and R. G. Moorhouse, 1996, *Phys. Lett. B* **372**, 88.
- de Bernardis, P., *et al.*, (Boomerang Collaboration), 2000, *Nature (London)* **404**, 955.
- de Gennes, P. G., 1966, *Superconductivity of Metals and Alloys* (Addison-Wesley, New York).
- Dermisek, R., A. Mafi, and S. Raby, 2001, *Phys. Rev. D* **63**, 035001.
- Dine, M., 1996, e-print hep-ph/9612389.
- Dine, M., P. Huet, and R. J. Singleton, 1992, *Nucl. Phys. B* **375**, 625.
- Dine, M., P. Huet, R. J. Singleton, 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. Y. Huet, A. D. Linde, and D. A. Linde, 1992, *Phys. Rev. D* **46**, 550.
- Dine, M., L. Randall, and S. Thomas, 1995, *Phys. Rev. Lett.* **75**, 398.
- Dine, M., L. Randall, and S. Thomas, 1996, *Nucl. Phys. B* **458**, 291.
- Dolgov, A., and J. Silk, 1993, *Phys. Rev. D* **47**, 4244.
- Dvali, G., A. Kusenko, and M. Shaposhnikov, 1998, *Phys. Lett. B* **417**, 99.
- Eguchi, K., *et al.*, (KamLAND Collaboration), 2003, *Phys. Rev. Lett.* **90**, 021802.
- Enqvist, K., A. Jokinen, T. Multamaki, and I. Vilja, 2001, *Phys. Rev. D* **63**, 083501.
- Enqvist, K., and A. Mazumdar, 2002, e-print hep-ph/0209244.
- Enqvist, K., and J. McDonald, 1998a, *Phys. Lett. B* **425**, 309.
- Enqvist, K., and J. McDonald, 1998b, *Phys. Lett. B* **440**, 59.
- Enqvist, K., and J. McDonald, 1999, *Nucl. Phys. B* **538**, 321.
- Espinosa, J. R., 1996, *Nucl. Phys. B* **475**, 273.
- Espinosa, J. R., M. Quirós, and F. Zwirner, 1993, *Phys. Lett. B* **307**, 106.
- Farakos, K., K. Kajantie, K. Rummukainen, and M. E. Shaposhnikov, 1994, *Nucl. Phys. B* **425**, 67.
- Farakos, K., K. Kajantie, K. Rummukainen, and M. E. Shaposhnikov, 1995, *Nucl. Phys. B* **442**, 317.
- Felder, G. N., J. García-Bellido, P. B. Greene, L. Kofman, A. D. Linde, and I. Tkachev, 2001, *Phys. Rev. Lett.* **87**, 011601.
- Friedberg, R., T. D. Lee, and A. Sirlin, 1976, *Phys. Rev. D* **13**, 2739.
- Friedmann, T., and E. Witten, 2002, e-print hep-th/0211269.
- Fukuda, S., *et al.*, (Super-Kamiokande Collaboration), 2001, *Phys. Rev. Lett.* **86**, 5656.
- Fukuda, Y., *et al.*, (Super-Kamiokande Collaboration), 1998, *Phys. Rev. Lett.* **81**, 1562.
- Fujii, M., and K. Hamaguchi, 2002a, *Phys. Lett. B* **525**, 143.
- Fujii, M., and K. Hamaguchi, 2002b, *Phys. Rev. D* **66**, 083501.
- Fujii, M., and T. Yanagida, 2002, *Phys. Lett. B* **542**, 80.
- Fukugita, M., and T. Yanagida, 1986, *Phys. Lett. B* **174**, 45.
- García-Bellido, J., M. Garcia-Perez, and A. Gonzalez-Arroyo, 2003, e-print hep-ph/0304285.
- García-Bellido, J., D. Grigoriev, A. Kusenko, and M. Shaposhnikov, 1999, *Phys. Rev. D* **60**, 123504.
- García-Bellido, J., and E. Ruiz Morales, 2002, *Phys. Lett. B* **536**, 193.
- German, G., G. Ross, and S. Sarkar, 2001, *Nucl. Phys. B* **608**, 423.
- Georgi, H., and S. L. Glashow, 1974, *Phys. Rev. Lett.* **32**, 438.
- Gherghetta, T., C. F. Kolda, and S. P. Martin, 1996, *Nucl. Phys. B* **468**, 37.
- Gonzalez-Garcia, M. C., and Y. Nir, 2002, e-print hep-ph/0202058.
- Griest, K., and E. W. Kolb, 1989, *Phys. Rev. D* **40**, 3231.
- Guidice, G. F., and R. Rattazzi, 1999, *Phys. Rep.* **322**, 419.
- Gurtler, M., E. M. Ilgenfritz, and A. Schiller, 1997, *Phys. Rev. D* **56**, 3888.
- Guth, A. H., 1981, *Phys. Rev. D* **23**, 347.
- Hagiwara, K., *et al.*, (Particle Data Group), 2002, *Phys. Rev. D* **66**, 010001.
- Hanany, S., *et al.*, 2000, *Astrophys. J. Lett.* **545**, L5.
- Harvey, J. A., and M. S. Turner, 1990, *Phys. Rev. D* **42**, 3344.
- Hisano, J., M. M. Nojiri, and N. Okada, 2001, *Phys. Rev. D* **64**, 023511.
- Huber, S. J., P. John, and M. G. Schmidt, 2001, *Eur. Phys. J. C* **20**, 695.
- Huber, S. J., and M. G. Schmidt, 2001, *Nucl. Phys. B* **606**, 183.
- Huet, P., and A. E. Nelson, 1996, *Phys. Rev. D* **53**, 4578.
- Jungman, G., M. Kamionkowski, and K. Griest, 1996, *Phys. Rep.* **267**, 195.
- Kainulainen, K., T. Prokopec, M. G. Schmidt, and S. Weinstock, 2001, *J. High Energy Phys.* **06**, 31.
- Kajantie, K., M. Laine, J. Peisa, K. Rummukainen, and M. E. Shaposhnikov, 1999, *Nucl. Phys. B* **544**, 357.
- Kajantie, K., M. Laine, K. Rummukainen, and M. E. Shaposhnikov, 1996a, *Nucl. Phys. B* **466**, 189.
- Kajantie, K., M. Laine, K. Rummukainen, and M. E. Shaposhnikov, 1996b, *Phys. Rev. Lett.* **77**, 2887.
- Kajantie, K., M. Laine, K. Rummukainen, and M. E. Shaposhnikov, 1997, *Nucl. Phys. B* **493**, 413.
- Kajita, T., *et al.*, 1985, *J. Phys. Soc. Jpn.* **54**, 4065.
- Kallos, R., L. Kofman, A. D. Linde, and A. Van Proeyen, 2000, *Phys. Rev. D* **61**, 103503.
- Kaplan, D. B., 1992, *Phys. Rev. Lett.* **68**, 741.
- Karsch, F., T. Neuhaus, A. Patkos, and J. Rank, 1997, *Nucl. Phys. B, Proc. Suppl.* **53**, 623.
- Kasuya, S., and M. Kawasaki, 2000a, *Phys. Rev. D* **61**, 041301.
- Kasuya, S., and M. Kawasaki, 2000b, *Phys. Rev. D* **62**, 023512.
- Kasuya, S., and M. Kawasaki, 2001, *Phys. Rev. D* **64**, 123515.
- Khlebnikov, S. Y., and M. E. Shaposhnikov, 1988, *Nucl. Phys. B* **308**, 885.

- Khlopov, M. Y., S. G. Rubin, and A. S. Sakharov, 2000, *Phys. Rev. D* **62**, 083505.
- Kirkman, D., D. Tytler, N. Suzuki, J. M. O'Meara, and D. Lubin, 2003, e-print astro-ph/0302006.
- Kobayashi, M., and T. Maskawa, 1973, *Prog. Theor. Phys.* **49**, 652.
- Kofman, L., A. Linde, and A. A. Starobinsky, 1994, *Phys. Rev. Lett.* **73**, 3195.
- Kofman, L., A. Linde, and A. A. Starobinsky, 1996, *Phys. Rev. Lett.* **76**, 1011.
- Kofman, L., A. Linde, and A. A. Starobinsky, 1997, *Phys. Rev. D* **56**, 3258.
- Kolb, E. W., and M. S. Turner, 1990, *The Early Universe* (Addison-Wesley, Reading, MA).
- Kolda, C. F., and J. March-Russell, 1999, *Phys. Rev. D* **60**, 023504.
- Krauss, L. M., and M. Trodden, 1999, *Phys. Rev. Lett.* **83**, 1502.
- Kusenko, A., 1997a, *Phys. Lett. B* **404**, 285.
- Kusenko, A., 1997b, *Phys. Lett. B* **405**, 108.
- Kusenko, A., V. Kuzmin, M. Shaposhnikov, and P. G. Tinyakov, 1998, *Phys. Rev. Lett.* **80**, 3185.
- Kusenko, A., and M. Shaposhnikov, 1998, *Phys. Lett. B* **418**, 46.
- Kusenko, A., M. Shaposhnikov, and P. G. Tinyakov, 1998, *Pis'ma Zh. Eksp. Teor. Fiz.* **67**, 229.
- Kusenko, A., M. E. Shaposhnikov, P. G. Tinyakov, and I. I. Tkachev, 1998, *Phys. Lett. B* **423**, 104.
- Kuzmin, V. A., V. A. Rubakov, and M. E. Shaposhnikov, 1985, *Phys. Lett.* **155B**, 36.
- Laine, M., 1994, *Phys. Lett. B* **335**, 173.
- Laine, M., and K. Rummukainen, 1998, *Nucl. Phys. B* **535**, 423.
- Laine, M., and M. E. Shaposhnikov, 1998, *Nucl. Phys. B* **532**, 376.
- Lee, K. M., J. A. Stein-Schabes, R. Watkins, and L. M. Widrow, 1989, *Phys. Rev. D* **39**, 1665.
- Lee, T. D., 1974, *Phys. Rep.* **9**, 143.
- Lee, T. D., and Y. Pang, 1992, *Phys. Rep.* **221**, 251.
- Linde, A. D., 1982, *Phys. Lett. B* **108**, 389.
- Linde, A., 1990, *Particle Physics and Inflationary Cosmology* (Harwood, Chur, Switzerland).
- Lyth, D. H., 1999, *Phys. Lett. B* **466**, 85.
- Manton, N., 1983, *Phys. Rev. D* **28**, 2019.
- McDonald, J., 2001, *J. High Energy Phys.* **103**, 22.
- Multamaki, T., 2001, *Phys. Lett. B* **511**, 92.
- Multamaki, T., and I. Vilja, 2000, *Nucl. Phys. B* **574**, 130.
- Multamaki, T., and I. Vilja, 2002, *Phys. Lett. B* **535**, 170.
- Murayama, H., and A. Pierce, 2002, *Phys. Rev. D* **65**, 055009.
- Netterfield, C. B., *et al.*, (Boomerang Collaboration), 2002, *Astrophys. J.* **571**, 604.
- Niessen, P., (AMANDA Collaboration), 1999, in *Proceedings of the 26th International Cosmic Ray Conference*, Salt Lake City, edited by D. Kieda *et al.*, Vol. 2, p. 344.
- Nilles, H. P., 1984, *Phys. Rep.* **110**, 1.
- Pascoli, S., S. T. Petcov, and W. Rodejohann, 2003, e-print hep-ph/0302054.
- Peebles, P. J. E., and B. Ratra, 2003, *Rev. Mod. Phys.* **75**, 559.
- Pilaftsis, A., 2002, *Nucl. Phys. B* **644**, 263.
- Postma, M., 2002, *Phys. Rev. D* **65**, 085035.
- Pryke, C., N. W. Halverson, E. M. Leitch, J. Kovac, J. E. Carlstrom, W. L. Holzappel, and M. Dragovan, 2002, *Astrophys. J.* **568**, 46.
- Randall, L., M. Soljatic, and A. H. Guth, 1996, *Nucl. Phys. B* **472**, 377.
- Randall, L., and S. Thomas, 1995, *Nucl. Phys. B* **449**, 229.
- Rosen, G., 1968a, *J. Math. Phys.* **9**, 996.
- Rosen, G., 1968b, *J. Math. Phys.* **9**, 999.
- Rubakov, V. A., and M. E. Shaposhnikov, 1996, *Usp. Fiz. Nauk*, **166**, 493 [*Phys. Usp.* **39**, 461 (1996)].
- Rummukainen, K., M. Tsypin, K. Kajantie, M. Laine, and M. E. Shaposhnikov, 1998, *Nucl. Phys. B* **532**, 283.
- Sakharov, A. D., 1967, *JETP Lett.* **6**, 24.
- Scari, L., *et al.*, 1999, in *Proceedings of the 26th International Cosmic Ray Conference*, Salt Lake City, edited by D. Kieda *et al.*, Vol. 2, p. 384.
- Shaposhnikov, M. E., 1986, *JETP Lett.* **44**, 465 [*Pis'ma Zh. Eksp. Teor. Fiz.* **44**, 364 (1986)].
- Shaposhnikov, M. E., 1987, *Nucl. Phys. B* **287**, 757.
- Smit, J., and A. Tranberg, 2002, *J. High Energy Phys.* **212**, 20.
- 't Hooft, G., 1976, *Phys. Rev. Lett.* **37**, 8.
- Weinberg, S., 1976, *Phys. Rev. Lett.* **37**, 657.
- Witten, E., 2002, e-print hep-ph/0207124.