CP VIOLATION AND THE ORIGINS OF MATTER

Andrew G. Cohen
Department of Physics, Boston University
Boston, MA 02215

ABSTRACT

I present a gentle introduction to baryogenesis, the dynamical production of a baryon asymmetry during the early universe. I review the evidence for a cosmic baryon asymmetry and describe some of the elementary ingredients necessary for models of baryon number production.

1 Introduction and Experiment

Even though the Universe has a size, age and complexity far beyond our everyday experience, the laws of physics determined in the laboratory can be extrapolated to the vast realms of the cosmos. This program, pursued since the earliest developments in the physical sciences, has seen enormous change over the last century. Especially important for particle physics has been the close interaction between the high energy frontier and the very early universe, and cosmological arguments are now routinely used to constrain the rampant imaginings of particle theorists. One area that is closely connected with the principle topic of this years school, CP violation, is baryogenesis, the dynamical production of a net baryon number during the early universe. This asymmetry, which is well established experimentally, is one of the most important features of the cosmos as a whole, and represents an enormous departure from the CP invariant state of equal matter and antimatter densities, with no net baryon number. The subject has been of concern to particle physicists since the discovery of microscopic CP violation, which encouraged the construction of concrete baryogenesis scenarios. The subject became a standard part of modern cosmology with the introduction of grand unified theories (GUTs), introduced in the 1970s, which establish a possible source for baryon number violation, an essential component of baryogenesis. More recent ideas have attempted to link the baron asymmetry with details of models of electroweak symmetry breaking, and offer the possibility of testing models of baryogenesis in future colliders such as the LHC.

There are many good reviews of baryogenesis at all levels*. Here we give only a brief overview of the subject and encourage further consultation of the references.

1.1 Initial Data

One of the fundamental questions concerning the large scale structure of our universe is surprisingly difficult to answer: What is the universe made of? In general terms this question reduces to the value of a single parameter, the total energy density of the universe, which is usually quoted in terms of a "critical" density related to the current Hubble expansion rate:

$$0.01 \lesssim \Omega_0 \equiv \frac{\rho}{\rho_{\text{crit}}} \lesssim 3 \, , \tag{1}$$

*The book[1] by Kolb and Turner is a good (although somewhat dated) starting point. There are many more recent reviews,[2–4] as well as references therein.

where $\rho_{\text{crit}} \equiv 3H_0^2/(8\pi G_N)$, H_0 is the Hubble constant and G_N is Newton's gravitational constant. The lower value comes from the visible content of the universe, the mass-energy associated with stars, galaxies, *etc*. The larger value comes from various measurements of large scale structure, especially measurements of the potential associated with gravitating (but not necessarily visible) mass-energy. The discrepancy between these numbers suggests that the majority of the mass-energy of the universe is dark, possibly a completely new kind of material. But even for the visible mass, we have no direct experience of the stuff out of which distant stars are made, although we believe this stuff to be matter similar to that which makes up our own star. The detailed physics of distant stars, such as stellar evolution, spectral lines, *etc*, is convincing evidence that these objects are made of baryons and leptons much as ourselves, but there remains the possibility that they are constructed from *antimatter*, *i.e.* antiquarks and positrons, rather than quarks and electrons. The transformation CP acting on a state of ordinary matter (by which we mean baryons, objects made of quarks carrying a positive baryon number) produces a state of antimatter (with negative baryon number). Thus if all stars in the universe contain matter (in the form of baryons) rather than antimatter (in the form of antibaryons), then this matter antimatter (or baryon) asymmetry represents a departure from CP symmetry as well.

What evidence is there that distant objects are made of matter rather than antimatter? For that matter, how do we know that the earth itself is matter? Matter and antimatter couple electromagnetically with known strength. Contact between matter and antimatter leads naturally to annihilation into photons with characteristic energy of 100s of MeV. Casual observation easily demonstrates the absence of this radiation when matter (in the form of ourselves, say) comes in contact with another terrestrial object. Thus we easily deduce that the earth (and all its occupants) are made of matter. A similarly pedestrian argument indicates that the moon too is made of matter. Indeed our exploration of nearby space convincingly shows that the solar system is composed of matter.

In fact it is not necessary that a man-made item come into contact with distant objects to establish the nature of such objects. If anything known to be matter is in contact with an unknown object, the absence of gamma radiation from annihilations demonstrates the object is not antimatter. For example micro-meteorites are continuously bombarding the earth without such radiation, and are therefore not antimatter. But these objects also rain upon Mars, which is therefore also not antimatter. This argument can obviously be extended: as long as a sufficiently dense matter trail extends from our solar system, absence of 100 MeV gamma rays demonstrates the absence of antimatter. This trail extends to distances comparable to the size of our local galactic cluster,[5] the Virgo cluster, a distance of 20 Mpc.

Unfortunately this region covers only a tiny fraction of the observable universe, which has a characteristic linear size several orders of magnitude larger than that of the Virgo cluster. Constraining the composition of objects beyond our local neighborhood requires a more complex analysis.

Experiments to search for cosmic antimatter from beyond this 20 Mpc distance have been proposed. The most ambitious of these, the Alpha Magnetic Spectrometer[6–11] (AMS) is scheduled to be deployed aboard the International Space Station sometime in the distant future[†]. This device, essentially a large mass spectrometer, will search for negatively charged nuclei in cosmic rays. The device should place a direct limit on antimatter in cosmic rays coming from a distance of nearly an order of magnitude beyond our local cluster. Although this distance scale remains small compared to the current visible universe, it is a significant step beyond our local cluster.

Lacking further direct experimental evidence against distant regions of antimatter, we must rely on alternative observational and theoretical analyses. Our original argument, the lack of gamma radiation emanating from points of contact between regions of matter and antimatter, fails when the density of both matter and antimatter becomes so small that the expected gamma ray flux falls below a detectable level. However this suggests an improvement on this argument: since the density of matter (and any putative antimatter) is decreasing with the cosmic expansion of the universe, we might expect that the flux of gamma radiation from such points of contact was larger in the early universe than it is today. Thus we might search for radiation from matter antimatter annihilation that occurred not today but sometime in the far past. A search for such radiation would differ from those which already place stringent limits on antimatter in our local neighborhood. Firstly, once produced as gamma rays, radiation would subsequently redshift as the universe expands. Consequently rather than searching for gamma rays with energies of 100s of MeV, we should search for lower energy radiation. Secondly, when we look out to large redshift (the distant past) on the night sky we are integrating over large portions of the universe. Consequently rather then seeking point sources we should search for a diffuse background of radiation coming from many points of intersection of domains of matter with those of antimatter.

[†] A prototype device has flown in the space shuttle. Although the exposure was insufficient to detect antimatter, this brief test has returned interesting cosmic ray physics.[12]

In order to use this technique to place limits on cosmic antimatter we must have some idea of how a diffuse photon spectral flux is related to the properties of domains of antimatter, in particular their size. We already know that such domains should be larger than the 20 Mpc limit we have in hand. The environment of this photon production, the interface between regions of matter and antimatter in the early universe, involves known principles of physics, and upper limits on the photon flux can be deduced. Although rather complicated in detail, the basic strategy is straightforward:

- The observed uniformity of the cosmic microwave background radiation implies that matter and antimatter must have been extremely uniform at the time when radiation and matter decoupled, a redshift of about 1100 or a time of about 10^{13} seconds. Thus at this time domains of matter and antimatter cannot be separated by voids, and must be in contact with each other. Prior to this time it is conceivable that matter and antimatter domains *are* separated by voids, and thus we do not include any annihilation photons prior to this epoch.
- Annihilation proceeds near matter antimatter boundaries through combustion, converting matter into radiation according to standard annihilation cross-sections. This change of phase in the annihilation region leads to a drop in pressure, and matter and antimatter then flow into this region. This leads to a calculable annihilation flux via the flow of matter and antimatter into this combustion zone. The annihilation process also gives rise to high energy leptons which deposit energy in the matter and antimatter fluids, significantly enhancing the annihilation rate.
- At a redshift of about 20 (approximately 10^{16} s after the big bang) inhomogeneities leading to structure formation begin to become significant. Although this likely does not affect the rate of annihilations significantly, rather than analyze this era in detail it is safer (more conservative) to ignore any further annihilation.
- The spectrum of photons produced prior to a redshift of 20 continues to evolve due to the expansion of the universe as well as subsequent scattering.

The results of this calculation[13] are shown in Figure 1. The upper curve represents the computed spectral flux of diffuse radiation from domains of antimatter with a characteristic size of 20 Mpc, the lower limit allowed by other analyses. The lower curve represents the spectral flux for a domain size of 1000 Mpc, a large fraction of the visible universe. In both cases this calculated flux is substantially larger then the observed diffuse gamma ray background (by balloon and satellite experiments). In particular such a flux would be in serious conflict with the results of the COMPTEL satellite

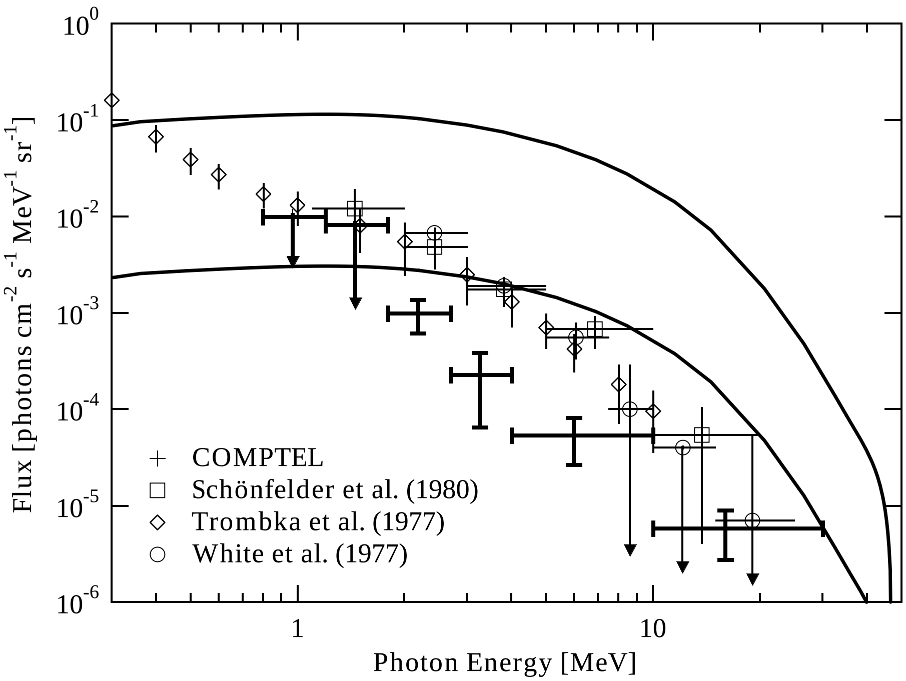

Fig. 1. Data[14] and expectations for the diffuse γ-ray spectrum.

experiments. We conclude that domains of antimatter of size less than 1000 Mpc are excluded.

1.2 A Baryon Asymmetry

The arguments of the preceding section indicate that the universe contains predominately matter and very little antimatter (or that matter and antimatter have been separated into several near universe-sized domains, a possibility[15,16] we will not consider here.) This asymmetry has been a focus of contemporary cosmology and particle physics principally because of its implied CP violation. To decide the significance of this asymmetry we need a quantitative measure of this departure from baryon antibaryon equality. Normally we will use the baryon density to photon number density ratio:

$$\eta \equiv \frac{n_B + n_{\bar{B}}}{n_\gamma} \ . \tag{2}$$

This choice is motivated partly by the dimensionless nature of the ratio, but more importantly, by the way in which this ratio scales with the expansion of the universe. Provided the expansion is isentropic (and ignoring baryon production or destruction) both the numerator and denominator densities dilute with the cosmic expansion in the same way, inversely proportional to the change in volume, and thus the ratio η is time independent.

Since our previous arguments suggest that $n_{\bar{B}}$ is insignificant, we may use the observed (visible) baryon density and the microwave background radiation density to obtain an experimental lower limit on η

$$10^{-10} \lesssim \eta \; . \tag{3}$$

In fact a more constrained value may be obtained by using some additional theoretical information. The synthesis of the light elements in the early universe depends quite sensitively on the baryon density. Using the best observations on the primordial elements this constrains η[17]:

$$4 \cdot 10^{-10} \lesssim \eta \lesssim 7 \cdot 10^{-10} \; . \tag{4}$$

Is this value significant? To get a better idea of how large this number is, we might imagine its value in a baryo-symmetric universe. In this case, as the universe cools from temperatures above 1 GeV where baryons and antibaryons are in thermal equilibrium with a thermal number density proportional to T^3, baryon number is kept in thermal equilibrium by baryon antibaryon annihilation. Once the rate for this process becomes slower than the expansion rate, the probability of a subsequent annihilation becomes negligible. Using a typical hadronic cross-section, this equality of rates occurs at a temperature of about 20 MeV. At this time baryons, in the form of protons and neutrons, have an equilibrium number density proportional to:

$$n_B \propto (T m_N)^{3/2} e^{-m_N/T} \tag{5}$$

and give a value for η

$$\eta \sim 10^{-20} \; . \tag{6}$$

This value, in gross conflict with the experimental number, cannot be avoided with thermal equilibrium between equal number of baryons and antibaryons, reflecting the efficient and near total annihilation of all matter. However there is a simple path to obtain a much larger value. If the number of baryons exceeds that of antibaryons by even a small amount, than the inability of each baryon to “pair up” with an antibaryon prevents total annihilation. In fact this excess need be only a few parts per billion at high temperature (leading to one extra baryon for each several billion photons) to achieve an adequate value for η.

But where would such an excess come from? It might appear as an initial condition, set at the beginning of the universe in some way beyond our ken. Note that such an initial condition is irrelevant in the context of inflation; following the reheating phase at the end of inflation all memory of such an initial condition is erased. Without inflation this is a rather unpleasant possibility that we must acknowledge, but we will favor an explanation that does not rely on a *deus ex machina* of this type. What is preferable is a mechanism by which this peculiar excess arises dynamically during the evolution of the universe, a possibility known as *baryogenesis*.

As was first observed by Andrei Sakharov,[18] there are three conditions that must be met in order for baryogenesis to occur:

- Baryon violation. Obviously if the universe is going to evolve a non-zero baryon number from a time when the baryon number vanishes (at the end of inflation, say) than the laws of physics must allow the baryon number to change.
- C and CP violation. Whatever process changes the baryon number must do so in a way that favors baryon production, rather than antibaryon production. Since both C and CP transformations change the sign of the baryon number, the laws of physics must violate both C and CP in order to obtain a positive value. Fortunately nature has provided us with both of these elements. As an example:

$$\frac{\text{Rate}[K_L^0 \to e^+\pi^-\nu]}{\text{Rate}[K_L^0 \to e^-\pi^+\bar{\nu}]} \simeq 1.006 \tag{7}$$

- Departure from thermal equilibrium. Roughly speaking if we populate all levels according to a Boltzmann distribution, since CPT guarantees that each level with a positive baryon number has a corresponding level with a negative baryon number, the total baryon number must vanish. More formally, since $\hat{B}$ is CPT odd and the Hamiltonian CPT even, in thermal equilibrium

$$\langle \hat{B} \rangle = \text{Tr}\hat{B}e^{-\beta\hat{H}} = \text{Tr}\, \Omega_{CPT}\Omega_{CPT}^{-1}\hat{B}e^{-\beta\hat{H}} = -\langle \hat{B} \rangle = 0\,. \tag{8}$$

Discussions of baryogenesis are often, not surprisingly, focused on the origin of these three ingredients. Beginning in the late 1970s it was realized that all three arise in commonly considered extensions of the standard model:

- Baryon Violation. Grand Unified theories, in which quarks and leptons appear in the same representation of a gauge group, naturally give rise to baryon violation.
- C, CP violation. Kaon physics already implies a source of C and CP violation.

- Departure from thermal equilibrium. The universe is known to be expanding and cooling off. This change in the temperature with time *is* a departure from thermal equilibrium.

We will turn to an evaluation each of these items in somewhat more detail.

Baryon violation is severely constrained by its apparent absence in the laboratory: experiments searching for proton decay have already placed a limit on the proton lifetime greater than 10^{32} years. How can baryon violation be significant for baryogenesis yet avoid a disastrous instability of the proton? The key is the notion of an accidental symmetry: a symmetry of all possible local operators of dimension four or less constrained by the particle content and gauge invariance of a theory. The significance of accidental symmetries appears when we consider the effects of new physics at high energies. These effects may be incorporated at low energies by including all possible local operators that respect the symmetries of this new physics. By dimensional analysis all operators of dimension higher than four will be suppressed by powers of the ratio of the low energy scale to the high energy scale. Now imagine that new physics at high energies does not respect some symmetry, like baryon number. At low energies we must include all local operators, including those that violate baryon number, an apparently disastrous result. But if the theory has an accidental symmetry, the only such operators are of dimension greater than four (by the definition of accidental symmetry), and thus new physics at high energies which violates this symmetry is suppressed by the high energy scale. In the standard model baryon number is exactly such an accidental symmetry: no baryon violating operators of dimension four or less can be constructed out of the standard particles consistent with the $SU(3) \times SU(2) \times U(1)$ gauge invariance. In fact the leading baryon violating operator in this construction is dimension six. If we then contemplate new physics which violates baryon number at a high energy scale, such as in grand unified theories, baryon violating effects will be suppressed at low energies by two powers of this high energy scale. Thus if the scale is greater than 10^{16} GeV, proton decay (a low energy process taking place near 1 GeV) is hugely suppressed.

As already indicated, CP violation is present in the kaon system at a level which appears more than adequate to explain a baryon asymmetry of less than one part in one billion. However CP violation in the standard model arising from a phase in the CKM matrix (which may or may not account for the phenomena observed in the kaon system) is unlikely to be responsible for the baryon asymmetry of the universe. As we will see, the effects of this phase in the early universe are quite small.

If the CP violation in the standard model can not account for the observed baryon asymmetry of the universe, what can? In fact almost *any* new source of CP violation beyond that of the phase in the CKM matrix gives rise to significant effects in the early universe. From a particle physics perspective, this is the principal reason for interest in the cosmic baryon asymmetry: it is a strong indication of physics beyond the standard model.

Lastly, the expansion of the universe which characterizes a departure from thermal equilibrium is governed by the Hubble parameter:

$$\frac{\dot{T}}{T} = -H \tag{9}$$

(at least during periods of constant co-moving entropy.) Today the Hubble parameter is quite small; the characteristic time scale for expansion of the universe is 10 billion years. Since most microphysical processes lead to thermal equilibrium on much shorter time scales, baryogenesis must take place either at a time when H is much larger, or at a time when Eq. (9) doesn't hold.

2 Grand Unification

Together the items of the previous section suggest that baryogenesis occurs at relatively early times, when the universe was hot and baryon violation was important. In particular the ingredients on our list all fit quite naturally into many grand unified theories. In such theories, super-heavy gauge bosons associated with the grand unified gauge group, as well as super-heavy Higgs bosons associated with GUT symmetry breaking, can mediate baryon violating processes. Although suppressed at low energies, at the high temperatures prevalent in the early universe baryon violation rates can be large. In addition, the rapid expansion rate

$$H \sim \frac{T^2}{M_P} \tag{10}$$

allows for significant departure from thermal equilibrium. Finally the interactions associated with new scalar fields that all GUT models must have may include CP violating couplings.

To see how this works in more detail, consider a toy model consisting of bosons X (and $\bar{X}$) which couple to quarks and leptons in a baryon violating, and CP violating, way. For example imagine that the X ($\bar{X}$) boson decays into the two final states qq ($\bar{q}\bar{q}$)

and $\bar{q}\bar{l}$ (ql) with branching fractions r ($\bar{r}$) and $1-r$ ($1-\bar{r}$) respectively. The parameters of this toy are constrained by symmetry. For example, CPT insures that the masses of the bosons are equal $m_X = m_{\bar{X}}$, as are the total widths $\Gamma_X = \Gamma_{\bar{X}}$. The baryon number of each final state is conventional: $B(qq) = 2/3$, $B(\bar{q}\bar{l}) = -1/3$, *etc*. Finally C and CP symmetry would imply $r = \bar{r}$. However lacking these symmetries, generically r will differ from $\bar{r}$[‡].

If we now imagine starting with thermal number densities of X and $\bar{X}$ bosons, our CPT constraint insures that these densities are equal $n_{\bar{X}} = n_X$. Using the parameters of introduced in the preceding paragraph we can compute the net baryon number of the quarks and leptons which result from the X and $\bar{X}$ decays:

$$n_B + n_{\bar{B}} = n_X\Big[r\frac{2}{3} + (1-r)(-\frac{1}{3})\Big] + n_{\bar{X}}\Big[\bar{r}(-\frac{2}{3}) + (1-\bar{r})\frac{1}{3}\Big] = n_X(r-\bar{r})\ . \quad (11)$$

Although this formula is correct, it is the answer to the wrong question. If all interactions are in thermal equilibrium, the X and $\bar{X}$ bosons will be replenished at the same time that they decay. That is, the rate for the inverse process, production of X (and $\bar{X}$) bosons through qq or $\bar{q}\bar{l}$ fusion, will have a rate in equilibrium which is precisely the same as the decay rate, when the number densities of all the particles are equal to their thermal equilibrium values. For example, at temperatures small compared to the X boson mass, the production rate of quarks and leptons via X decay is small, since there are very few X bosons in equilibrium, $n_X \propto \exp(-m_X/T)$. Conversely the inverse process, creation of an X boson, is rare since the quarks and leptons are exponentially unlikely to have the energy necessary to produce a real X boson. So in equilibrium the baryon number does not change, and Eq. (11) is not relevant.

This suggests what turns out to be the key to baryon production—we need the number density of X and $\bar{X}$ bosons at $T << m_X$ to be much larger then the exponentially small equilibrium number density. Under these circumstances the X and $\bar{X}$ production processes will be much smaller than the decay processes. If the number density of X and $\bar{X}$ bosons is sufficiently large, we may even ignore the inverse process all together.

How do we arrange this miracle? Clearly we must depart from thermal equilibrium, something we already knew from our discussion of Sakharov's conditions. But as we have also discussed the universal expansion allows such a departure when the rate for an equilibrating process is slow compared to the expansion rate. In this case, we need the processes that keeps the number density of X and $\bar{X}$ bosons in equilibrium to be

[‡]Of course C and CP violation are not sufficient—interference with a scattering phase is also necessary.

slow compared to the expansion. There are two processes which decrease the number of bosons: the decay of the X and $\bar{X}$ bosons; and annihilation of the X and $\bar{X}$ bosons into other species. Both of these processes can be slow if the couplings of the X boson are weak. Of course "slow" means in comparison with the Hubble expansion rate, $H \sim T^2/M_P$. If this is indeed the case, the number density of X bosons will not track the equilibrium value proportional to $\exp(-m_X/T)$, but instead remain larger. Then once the age of the universe is larger than the lifetime of the X boson, decay will occur, leading to a baryon number according to Eq. (11).

There is one important constraint that we have overlooked. Even though the X and $\bar{X}$ bosons are not re-produced around the time that they decay, there are other processes we must not forget. In particular, there are processes which violate baryon number through the mediation of a (virtual) X boson. In our toy example these may be represented by the effective four-fermion operator $qqql$. This dimension six operator has a coefficient proportional to two inverse powers of the m_X mass, and thus at temperatures low compared to this mass the effects of this operator are small. Nevertheless processes of this type will change the baryon number, tending to equilibrate this number to zero. Therefore we must further require that baryon violating processes such as this one must also be out of equilibrium at the time the X and $\bar{X}$ bosons decay.

The procedure outlined above is usually called a "late decay", or "out-of-equilibrium decay" scenario. Developed extensively from late 1970s through the present, they have provided a framework in which to discuss baryogenesis, and have led to many concrete models that can explain the non-zero value of η. Although successful in principal, models of GUT baryogenesis often have difficulty obtaining the large baryon asymmetry we observe:

- *Rates:* We have seen that a number of rates must be slow compared to the expansion rate of the universe in order to depart sufficiently from equilibrium. These rates are typically governed by the GUT scale, while the expansion rate is proportional to T^2/M_P. The relevant temperature here is that just prior to the decay of the X bosons. Since we need these bosons to be long lived, this temperature is lower than the GUT scale, and the expansion rate is correspondingly slower. Thus the departure from equilibrium is far from automatic and detailed calculations in a specific GUT are necessary to determine whether these conditions can be satisfied.

- *Relics:* One problematic aspect of many GUTS is the presence of possible stable relics. For example some GUTS have exactly stable magnetic monopoles which would be produced in the early universe at temperatures near the GUT scale. Unfortunately these objects are a cosmological disaster: the energy density in the form of monopoles would over-close the universe, in serious conflict with observation. One of the early great successes of inflation was a means for avoiding this catastrophe. At the end of inflation all matter in the universe has been "inflated away", leaving a cold empty space free from all particles (baryons as well as monopoles!). However following the end of inflation, the vacuum energy density in the inflaton field goes into reheating the universe, producing a thermal distribution of particles. If this reheating is fast, energy conservation tells us that the reheat temperature will be close to the original scale of inflation, near or above the GUT scale. Unfortunately this would reintroduce the monopoles. On the other hand if this reheating is slow (as would be the case if the inflaton is weakly coupled) then the energy density in the inflaton field decreases as the universe expands, leading to a much lower reheat temperature. Thus for inflation to solve the monopole crisis, the reheat temperature must be well below the GUT scale, in which case monopoles are not re-introduced during the reheating process. Unfortunately neither are the X and $\bar{X}$ bosons, and thus baryogenesis does not occur.

Neither of these objections are definitive—there are proposals for circumventing them both. For example much of our discussion has focused on small departures from thermal equilibrium. It may be possible to have huge departures, where particle distributions are not even remotely thermal. In this case the analysis of reaction rates is quite different. There may also be many more couplings which allow a greater range of reaction rates. Perhaps these are associated with Yukawa couplings of neutrinos or other sectors of the GUT. These objections do however make these scenarios less compelling. In addition there is another, more philosophical, problem. Often in these models the details of baryogenesis are pushed into very particular aspects of the GUT, physics at scales which are not accessible in the laboratory. Thus in many instances, whether or not GUT baryogenesis occurs is experimentally unanswerable. For these reasons it is advisable to investigate alternatives.

3 Electroweak Baryogenesis

In 1985 Kuzmin, Rubakov and Shaposhnikov[19] made the remarkable observation that all three of Sakharov's criteria may be met in the standard model. Firstly, and perhaps most surprisingly, the standard model of the weak interactions does not conserve baryon number!

The non-conservation of baron number in the standard model is a rather subtle effect. At the classical level, the conservation of baryon number is practically obvious—each term in the classical action respects a transformation of the baryon number. Nöther's theorem then applies, and we can construct a four-vector, the baryon number current, which satisfies the continuity equation, that is whose four-divergence vanishes. Nonetheless this naïve argument is wrong: this four vector does *not* have vanishing four-divergence in the full quantum theory.

This situation is not totally unfamiliar. In the simple case of quantum electrodynamics a corresponding phenomena occurs, known as the axial anomaly. QED has a symmetry of the classical action corresponding to an axial rotation of the electron field (that is, a rotation which is opposite on the left and right chirality electron fields). Aside from the electron mass term which we will ignore, this transformation leaves the action unchanged, and the Nöther procedure leads to a covariantly conserved four-vector, the axial current. However as is well known this current is *not* divergenceless:

$$\partial_\mu J^\mu_a = \frac{e^2}{32\pi^2} F_{\mu\nu}\tilde{F}^{\mu\nu} \propto \vec{E}\cdot\vec{B}\ , \tag{12}$$

where $\tilde{F}^{\mu\nu} \equiv \epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}/2$. This remarkable equation, which can be derived in a number of different ways, embodies the violation of axial charge due to quantum effects in the theory[§]. Note that ignoring spatial variations this equation implies that the time derivative of the baryon density will be non-zero in the presence of a non-zero $\vec{E}\cdot\vec{B}$. Note that the chiral nature of the current couplings are important for obtaining this result; the current with non-chiral couplings, the electromagnetic current, is strictly conserved.

The situation in the standard model is similar. The baryon current derived via the Nöther procedure is vectorial, and thus would seem an unlikely candidate for an

[§]The axial anomaly in QED has a long and well-known history. Eq. (12) may be obtained for example by evaluating the triangle diagram, by computing the change in the functional integral measure under an axial rotation, or by an exact calculation of the electron propagator in a constant background electric and magnetic field.

anomaly. However the weak interaction couplings *are* chiral, which leads to an equation for the divergence of the baryon current corresponding to Eq. (12):

$$\partial_\mu J^\mu_B = 3\Big[\frac{g^2}{32\pi^2} W^a_{\mu\nu}\tilde{W}^{a\mu\nu} + \frac{g'^2}{32\pi^2} F_{Y\mu\nu}\tilde{F}^{\mu\nu}_Y\Big] = \partial_\mu J^\mu_L \ . \qquad (13)$$

In this equation W and F_Y are the gauge field strengths for the $SU(2)$ and $U(1)$ hypercharge gauge potentials, g and g' are the corresponding gauge couplings and the 3 arises from a sum over families. We have also noted that the lepton number current has the same divergence as the baryon number current. Consequently the current $J_{B-L} \equiv J_B - J_L$ is divergenceless, and the quantum number $B - L$ is absolutely conserved.

What does Eq. (13) really mean? To gain some understanding of this equation, imagine constructing an electroweak solenoid surrounding an electroweak capacitor, so that we have a region in which the quantity $\vec{E}^a \cdot \vec{B}^a$ is non-zero. In practice this is rather difficult, primarily because we live in the superconducting phase of the weak interactions, and therefore the weak Meissner effect prevents the development of a weak magnetic field. But lets ignore this for the moment. Now perform the following gedanken experiment: start with no weak electromagnetic fields, and the region between the capacitor plates empty. If we solve the Dirac equation for the quarks and leptons, we obtain the usual free particle energy levels. In this language, we fill up the Dirac sea, and leave all positive energy levels unoccupied. Now imagine turning on the weak $\vec{E}^a$ and $\vec{B}^a$ fields adiabatically. In the presence of these slowly varying fields, the energy level solutions to the Dirac equation will flow, while the occupation of any given level does not change. But according to Eq. (13) the baryon number will change with time. This corresponds to the energy of some of the occupied levels in the Dirac sea flowing to positive energy, becoming real particles carrying baryon number. Although surprising at first, this is not very different from ordinary pair production in a background field. What is peculiar is the creation of quarks in a way different from antiquarks, so that a net baryon number is produced.

By itself this effect is intriguing but not sufficient. After all what we are really after is a transition which changes baryon number without changing the state of the gauge field, much as the four-fermion operator in our grand unified example did. That is, what we would like to do is begin our gedanken experiment as above, but at the end of the day turn off the electric and magnetic fields. Naïvely this would leave us with zero baryon number: if we turn the fields off as the time reverse of how we turned them on, we produce baryon number at first, and then remove it later on. Indeed this is what happens with axial charge in the quantum electrodynamics example. But the non-abelian example contains another wrinkle: it is possible to turn the electric and magnetic fields on and then off in a way which leaves a non-zero baryon number!

The trick as realized by 't Hooft[20,21] follows from noticing that, unlike the abelian case, there are a large number of non-trivial gauge potentials which have vanishing electric and magnetic fields. It is possible in our gedanken experiment to begin with one of these potentials, and finish with another, thus tying a "knot" in the gauge field¶. The result is a transition from a state with no weak electric and magnetic fields and no baryon number (a "vacuum"), and ending with no weak electric and magnetic fields but non-zero baryon number. Making such a transition requires a "large" gauge field, one in which the field strength is of order $1/g$. In addition, the total change in baryon number is quantized in units of the number of families, presumably 3.

If we accept this fancy formalism, we have an obvious question: why is the proton stable? If the weak interactions violate baryon number, shouldn't the proton lifetime be a characteristic weak time scale? In fact, the proton is absolutely stable even in the presence of this baryon violation, because each process changes the baryon number by 3. Since the proton is the lightest particle carrying baryon number, its decay would require changing the baryon number by 1, which cannot occur if all baryon violating process change the baryon number by multiples of 3. Thus there is a selection which accounts for the stability of the proton.

What about other baryon violating processes? In fact these too are unimportant. In our gedanken experiment above we ignored the fact that the weak interactions are broken, that we live in a superconducting phase of the weak interactions. But this means that there is a large potential energy cost in creating a weak $\vec{E}^a$ and $\vec{B}^a$ field which interpolates between our states with different baryon number. That is, there is a potential barrier that we must overcome in order to change the baryon number by a weak interaction. Since the gauge field must change by order $1/g$, the height of this barrier (the cost of overcoming the Meissner effect) is

$$E_s \sim \frac{M_Z}{\alpha_{wk}} \sim \text{a few TeV} \tag{14}$$

where α_{wk} is the weak analog of the electromagnetic fine structure constant. The gauge field configuration at the peak of this barrier is called the "sphaleron", and hence this

¶This argument is a bit tricky. In order to discuss the physics of gauge potentials it is necessary to gauge fix. Even after gauge fixing there are gauge potentials which begin in the far past with one "vacuum" potential, and end with a different one.

energy is known as the sphaleron energy.

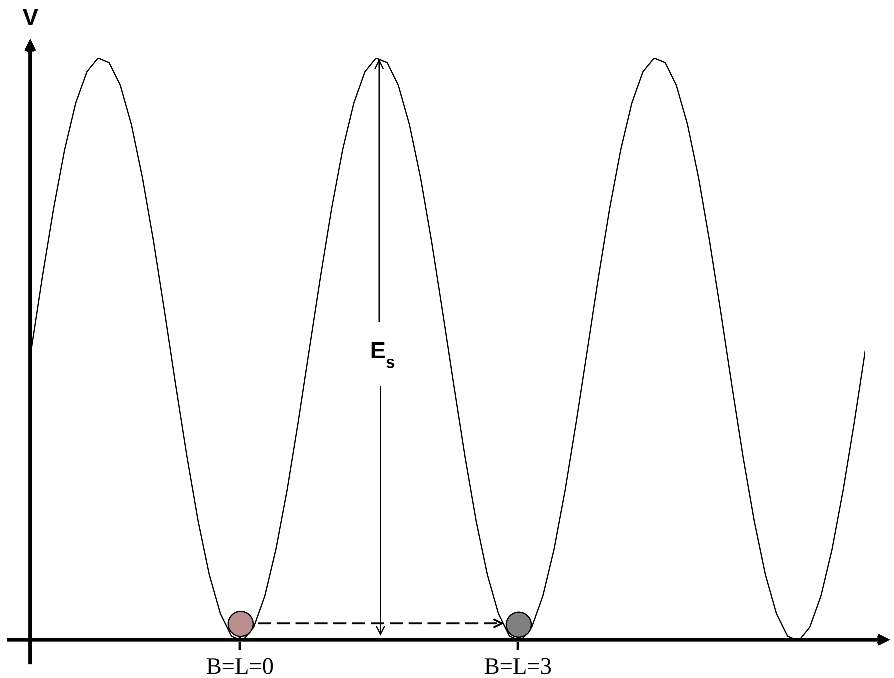

Fig. 2. The potential energy in one direction in gauge field space. This direction has been chosen to go from one zero energy gauge field configuration to another through the pass of lowest energy.

The presence of this barrier means that processes with energies below the barrier height are highly suppressed; they are strictly forbidden classically, but can occur through quantum tunneling. Like all tunneling processes, the probability of such a transition will be proportional to a semi-classical barrier penetration factor:

$$\text{Prob} \propto e^{-4\pi/\alpha_{wk}} \sim 10^{-40} \ , \tag{15}$$

an utterly negligible effect. In contrast to the grand unified case where baryon violation was suppressed at low energies by powers of the ratio of the energy to the grand unification scale, here the baryon violation is exponentially suppressed by the presence of a barrier.

If our interest were only sensitive tests of baryon number conservation in the laboratory, we would safely move on to another area of research. But since our interest is in baryogenesis in the early universe, we must take this picture of baryon violation in the weak interaction by transiting this barrier more seriously. At temperatures comparable or larger than the barrier height we would expect a significant population of states with energies above the barrier. These states could make a transition without the quantum tunneling suppression by simply evolving classically over the top of the barrier. The rate for such a baryon violating process will be controlled by the probability of finding a state with energy at least as large as the sphaleron energy:

$$\Gamma \propto e^{-E_s/T} \; . \tag{16}$$

When the temperature is larger than E_s this exponential is no longer a suppression at all. Hence we expect that at temperatures above a few TeV baryon violation in the weak interactions will occur at a characteristic weak interaction rate. Note that at temperatures of a few TeV weak interactions are extremely rapid compared to the Hubble expansion rate, and thus baryon violating interactions would be in thermal equilibrium.

We come to the first important consequence of baryon violation in the weak interactions: grand unified baryogenesis does not necessarily produce a baryon asymmetry! Even if a late decaying X boson would produce a baryon asymmetry at temperatures near the GUT scale, this asymmetry will be equilibrated away by baryon violating weak interactions. Our discussion of grand unified baryogenesis concluded that baryon violation from virtual X boson exchange must be slow for baryogenesis to succeed, but the real requirement is that *all* baryon violation must be slow; we must take into account *all* sources of baryon violation, including that of the weak interactions.

There is a simple way of avoiding this effect. As indicated in Eq. (13) the baryon and lepton number currents have exactly the same divergence. Hence their difference, the $B - L$ current, is strictly conserved. Therefore if the X boson decay produces a net $B - L$, weak interactions cannot equilibrate this quantum number to zero. The result will be both a net baryon number and a net lepton number. However baryon and lepton number violating weak interactions must be taken into account when calculating the baryon asymmetry produced.

Rather surprisingly we have concluded that baryon violation is present in the standard model, at least at temperatures above a few TeV. In principle this opens the possibility of baryogenesis taking place at temperatures well below the GUT scale. Unfortunately we face another obstacle: departure from thermal equilibrium. As discussed earlier, the expansion rate of the universe at temperatures near a TeV is quite slow: $H \sim T^2/M_P \sim 10^{-16}$ TeV. All standard model interactions lead to reaction rates much larger than this expansion rate, typically of order $\Gamma \sim \alpha_{wk} T \sim 10^{-3}$ TeV. Thus departure from thermal equilibrium is impossible with such a leisurely expansion. Fortunately there are occasions during the early universe in which the smooth variation of the temperature with the expansion, Eq. (9), is invalid. This typically occurs when the equation of state for the content of the universe undergoes an abrupt change, such as during a change in phase structure. For example when the temperature falls below the mass of the electron, electrons and positrons annihilate into photons, converting their energy from a non-relativistic form (the mass-energy of the leptons) into a relativistic form (radiation). But there may be other phase changes in the early universe. With a phase transition there exists the possibility of significant departure from thermal equilibrium, at least if the transition is discontinuous, or first order.

Is there any reason to expect a phase transition in the early universe? At temperatures much higher than a few TeV we have very little idea of the state of the universe; until we probe physics at these high energies in the laboratory we cannot say whether or not phase transitions occur. Of course we are permitted to speculate, and indeed there are many proposals for new physics beyond the standard model which lead to interesting dynamics in the early universe. But beyond speculation, we already expect that there is at least one phase transition in the context of the standard model: the electroweak phase transition.

As we have already mentioned we currently live in a superconducting phase of the electroweak interactions. The W and Z boson masses arise from the interaction of the gauge fields with a non-zero order parameter, an object that carries electroweak quantum numbers and has a non-zero expectation value in the vacuum. The short range nature of the weak force is a consequence of this interaction, just as the electromagnetic interaction is short range in ordinary superconductors. In fact it is this property of the weak interactions which leads us to deduce the existence of a non-zero order parameter. We know the value of the order parameter, the weak vev, is approximately 250 GeV; we also know that the order parameter is a weak doublet, from the relation between the W and Z masses and the weak mixing angle. However unlike electromagnetic superconductivity where the order parameter is known to be a composite of two electrons, a so-called “Cooper pair”, the weak order parameter remains mysterious. One possibility is that the order parameter is simply some new field with its own physical excitations, the Higgs field. Another is that it is a composite of two fermions, like the Cooper pair. But until we have probed the details of electroweak symmetry breaking in detail, as we hope to do in future collider experiments, we can not say with any confidence what form the detailed physics of this order parameter takes.

One thing we do expect, in analogy with ordinary superconductivity, is the change in phase of the weak interactions at high temperatures. Just as an electromagnetic superconductor becomes non-superconducting as the temperature is increased, so too the weak interactions should revert to an unbroken phase at high temperature. When the temperature is on the order of 100 GeV, the order parameter should vanish, the weak gauge symmetry will be unbroken and the W and Z (and the quarks and leptons) will become massless. In our discussion of baryon violation in the weak interactions we suggested that at temperatures larger than the sphaleron energy baryon violation would be unsuppressed, as transitions could take place above the barrier. But the barrier itself was a consequence of the Meissner effect, a sign of superconductivity. Indeed Eq. (14) clearly shows the relationship with symmetry breaking: the sphaleron energy is proportional to M_Z which in turn is proportional to the order parameter. At temperatures of a few hundred GeV, well below the sphaleron energy, when the weak symmetry is restored and the order parameter goes to zero, the barrier disappears. Consequently baryon violation will occur rapidly just on the unbroken side of the phase transition.

In order for any of this to play a role in baryogenesis, we require significant non-equilibrium effects at the phase transition. According to the usual classification of phase transitions, such non-equilibrium effects will arise if the phase transition is first order. Under these circumstances the transition itself may proceed in a classic first order form, through the nucleation of bubbles of broken phase[‖]. Indeed as the universe cools from high temperature, we begin with a homogeneous medium in the unbroken phase of the weak interactions. Quarks and leptons are massless, weak interactions are long range (aside from thermal screening effects) and, most importantly, baryon violation is rapid. Calculating the rate for baryon violation requires understanding the details of the classical thermodynamics of the gauge fields, a difficult subject. The result however is relatively simple:

$$\Gamma_{\Delta B} \sim \alpha_{wk}^5 T \tag{17}$$

This is a rather crude approximation; for example there are logarithmic corrections to this relation that may be significant, as well as a potentially large dimensionless coefficient. Nevertheless the exact formula may in principle be obtained numerically in terms of α_{wk} and the temperature.

As the universe cools we eventually reach a moment in which the free energy of the unbroken phase is equal to that of the broken phase, as indicated by the free energy

[‖]This is not the only possibility; for example it may proceed through spinodal decomposition, or some more complicated mechanism. In all these circumstance non-equilibrium phenomena are likely.

curve labeled by T_c in Fig. (3). However if the transition is first order, these two phases are separated by a free energy barrier and the universe, unable to reach the broken phase, remains in the unbroken phase. As the universe continues to expand, the system *supercools*, remaining in the unbroken phase even though the broken phase has a lower free energy. Finally we reach a point where bubbles of the preferred, broken, phase nucleate and begin to grow. Eventually these bubbles percolate, completing the transition.

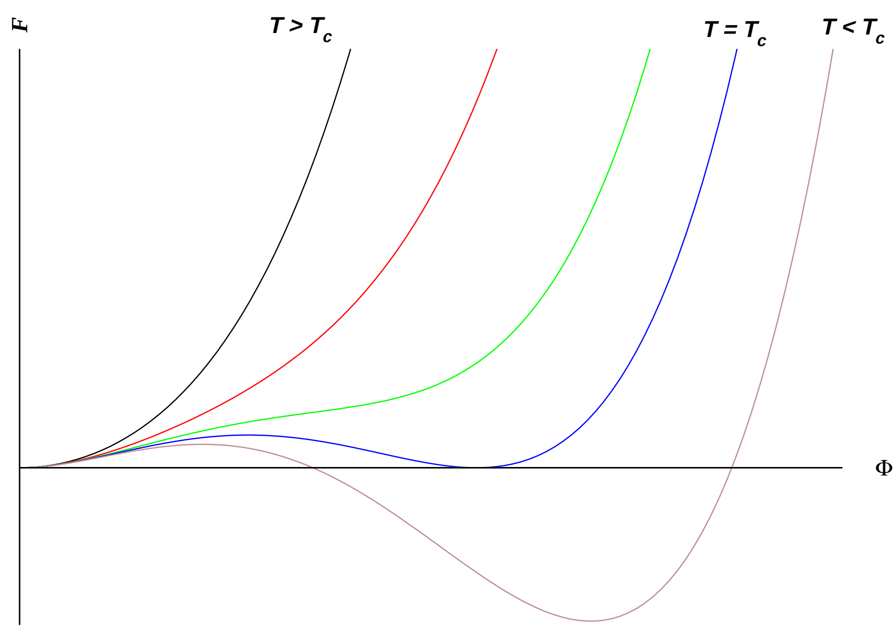

Fig. 3. The free energy versus the order parameter for a classic first order phase transition.

Clearly these expanding bubbles represent a departure from thermal equilibrium. From the point of view of Sakharov's condition the most relevant fact is the discontinuity in the order parameter, the weak vev. In the region outside the bubbles the universe remains in the unbroken phase where the weak order parameter is zero. As discussed previously there is no barrier between the states of different baryon number, and baryon violation is rampant. In the bubble interior the weak vev is non-zero, the W and Z bosons are massive, and *the barrier between states of different baryon number is in place.* In this case the rate of baryon violation is exponentially suppressed according to a Boltzmann factor $\exp(-E_b/T)$ where E_b is the barrier height. Naïvely we might expect E_b to be the sphaleron energy. However the sphaleron energy represented the barrier height at zero temperature; at finite temperature the barrier is generically different, evolving to the zero temperature shape as the universe cools. But E_b is still controlled by the order parameter, the weak vev. If this vev is large, near its vacuum value of 250 GeV, baryon violation will be essentially shut off in the bubble interior. On the other hand if the vev is too small, baryon violation will proceed rapidly inside the bubble as well as out.

The difference in the weak vev in the bubble interior and the bubble exterior, the discontinuity in the weak order parameter, is a measure of the strength of the transition. If these two values are nearly equal, the phase transition is nearly continuous, a second order transition. If on the other hand the discontinuity is large, the phase transition is said to be strongly first order. For electroweak baryogenesis to occur, baryon violation must be out of thermal equilibrium in the bubble interior, a situation that will transpire only if the vev is sufficiently larger. Thus we need a strongly first order electroweak phase transition.

What do we know about the electroweak phase transition? Unfortunately almost nothing. This is due in small measure to our inability to understand the complex thermal environment in a relativistic quantum field theory. Over the past decade there has been a great deal of progress in simulating field theories at finite temperature, deducing details of phase transitions and reaction rates. However these advances are of little use if we don't know what theory to simulate. The main reason we can't say definitely whether the electroweak phase transition is first or second order, whether it is strongly or weakly first order, or practically anything else about it is simple: we have no idea what physics is responsible for electroweak symmetry breaking.

We do have some theories of electroweak symmetry breaking, and huge effort has been invested in determining the details of the phase transition in these cases. The original theory of electroweak symmetry breaking relied on the introduction of a fundamental weak doublet scalar field, the Higgs field. In this rather simple case, the electroweak phase transition is first order only if the physical Higgs scalar is very light, with a mass well below the current experimental bound. But this theory is not the most popular alternative for electroweak symmetry breaking due to its theoretical shortcomings. Of somewhat greater appeal is the minimal supersymmetric standard model, the MSSM. In this case there are a host of new particles: supersymmetric partners of the quarks, leptons and gauge bosons, as well as two Higgs multiplets. In fact this theory also requires some of these new states to be relatively light in order to obtain a sufficiently strongly first order phase transition. As the LEP bound on the MSSM Higgs mass improves, the region of parameter space for which the phase transition is appropriate is rapidly disappearing.

Should we take this to mean the weak phase transition is probably inappropriate for electroweak baryogenesis to take place? That depends a bit on our philosophy. Given that these are but 2 ideas out of a nearly infinite variety we should not necessarily become disheartened. More importantly there have been analyses of modest alternatives of the above theories: non-supersymmetric theories with multiple Higgs fields, extensions of the MSSM including singlets, and even strongly interacting theories of electroweak symmetry breaking. In most of these cases a sufficiently strong first order phase transition is easy to arrange, if not generic. In fact this is perhaps one of the more positive aspects of electroweak baryogenesis. The physics responsible for electroweak symmetry breaking is intimately related with the possibility of electroweak baryogenesis: some models of electroweak symmetry breaking do not produce a baryon asymmetry (or not one of sufficient size) while others do. This is one of the few places that the forefront of electroweak physics, electroweak symmetry breaking, may have a profound effect on cosmology (or vice versa).

3.1 Baryon Production

We now have all of Sakharov's ingredients in place, all in the weak interactions: baryon violation, C and CP violation and a departure from thermal equilibrium. But we still have not explored how these ingredients combine to produce a baryon asymmetry.

Clearly we require all three ingredients to work together—the absence of any one implies the absence of baryogenesis. The non-equilibrium requirement, satisfied by the nucleation and subsequent expansion of bubbles of broken phase, is most importantly realized as a spatial separation of baryon violation: baryon violation is rapid outside the bubble, and non-existent in the bubble interior. C and CP violation, at least in the standard model, take place through the Yukawa couplings in the Lagrangian. That is, C and CP violation appear in the form of non-trivial phases in the couplings of quarks (and possibly leptons in extensions of the standard model) to the Higgs field, the order parameter for electroweak symmetry breaking. But it is precisely this field which represents the electroweak bubbles which appear at the phase transition.

The details of how the baryon asymmetry may be calculated in the context of these expanding bubbles is complicated, and we will not discuss it at any length. The ingredients are clear: the CP violating interaction of quarks and leptons with the expanding bubbles can in principle bias the production of various quantum numbers (including but not limited to baryon and lepton number); all that is required is an interaction that allows the creation or destruction of a net value for such a quantum number. For example, the interaction with the expanding bubble may bias the production of left-chirality top quarks over right-chirality top quarks (to pick a random example). Provided CP violation (either directly or in the form of one of these quantum number asymmetries) biases baryon number in a region outside the bubble where baryon violation is rapid, a net baryon number will be produced. Following our example, an excess of left-chirality top quarks (which have a weak interaction) over right-chirality top quarks (which do not) biases the weak interactions in the direction of increasing baryon number. An important element which complicates the discussion is the transportation of quark and lepton charges from one region of space to another. The transport properties of the plasma are crucial in understanding how the baryon violating interactions, which take place outside the bubble, are biased by CP violation, which is dominant where the Higgs field is changing inside the bubble. Depending on the details of the bubble profile the analysis looks a bit different, although the results are qualitatively similar.

3.2 CP Violation

We finally must come to grips with CP violation; now that we understand how it is relevant to electroweak baryogenesis, we can ask what the characteristic size of CP violating effects of the sort described in the last paragraph will be. In fact this question is not as difficult as might be supposed. CP violation in the standard model arises from a non-trivial phase in the Yukawa couplings of the quarks. The only tricky issue is that this phase has no unique location: we may move it from one coupling to another by making field redefinitions. More physically this means that an interaction will only violate CP when the interaction involves enough couplings such that we cannot remove this phase from all these couplings simultaneously. For example, if a process involves only two families of quarks, the CP violating phase may be put in the third family, and this process will be CP conserving.

Since the Yukawa couplings are relatively small (even the top quark coupling), perturbation theory should be an adequate guide to the size of CP violating effects. To estimate this size we must construct an object perturbatively out of the various coupling constants of the standard model in a way which involves an (irremovable) CP violating phase. Clearly there must be a large number (8) of Yukawa couplings from all three families as well as a large number (4) of weak interactions in order to get an irremovable phase. This product of small dimensionless coupling constants is an invariant measure of CP violation in any perturbative process. One such example, involving the largest Yukawa couplings, is

$$\delta_{CP} \sim \alpha_{wk}^2 \lambda_t^4 \lambda_b^2 \lambda_s \lambda_d \sin^2\theta_1 \sin\theta_2 \sin\theta_3 \sin\delta \sim 10^{-16} \ . \tag{18}$$

This remarkably small number, many orders of magnitude smaller than the observed baryon asymmetry, is a consequence of the detailed symmetries of the standard model, where CP violation is intimately connected with flavor violation. As long as the flavor physics of baryogenesis is perturbative, the standard model has no hope of producing a baryon asymmetry large enough. Although we have consistently maintained that the standard model has CP violation, and that this is one of the most interesting reasons to investigate baryogenesis, it now seems that we have been misled, that this CP violation is far too small to be relevant for baryon production in the early universe.

Why did we argue earlier that CP violation in the kaon system, Eq. (7), was so much larger than this perturbative estimate? In fact we have been careful to argue that the estimate of CP violation, Eq. (18), only applies when the standard model Yukawa interactions can be used perturbatively. This is not the case for CP violation in the kaon system. If we wish to compute CP violating effects at kaon energies, $E << 250$ GeV, we must first construct the effective theory appropriate to these energy scales by integrating out modes with energies larger than E. This includes for example the W and Z, the top and bottom quarks, *etc*. As usual this process introduces inverse powers of these heavy masses, such as $1/M_W^2$ and $1/m_t^2$. Since these masses are proportional to the weak couplings g and λ_t appearing above, this effective theory has interactions which can *not* be represented as a power series in couplings (although it is easy enough to construct this effective theory and keep track of the Yukawa couplings), and the estimate Eq. (18) does not apply**.

But we have now come to the crux of the matter, and if it were not for the interesting physics associated with baryon violation, cosmic expansion, *etc.* that we wished to discuss we could have started (and ended) our discussion of baryogenesis here. The most important message from this analysis is that it is highly unlikely that CP violation from the phase in the CKM matrix has anything at all to do with the cosmic baryon asymmetry. Although we have chosen to mention this in the context of electroweak

**A more old-fashioned language for the same phenomena would note the enhancement of perturbative matrix elements by small energy denominators in perturbation theory.

baryogenesis, there is nothing special about this scenario in our analysis of the size of CP violating effects. Everything we have said applies to standard model CP violation in any theory of baryogenesis that takes place at high energies where our perturbative argument applies. This is certainly the case in grand unified baryogenesis as well as electroweak baryogenesis.

Once more, with feeling: standard model CP violation in the form of a phase in the CKM matrix is not likely to produce a significant baryon asymmetry of the universe. Why is this so important? As we have argued there *is* a cosmic baryon asymmetry, and if it didn't come from CP violation in the standard model, where did it come from? The obvious conclusion is that there is CP violation (and hence new physics) beyond the standard model. This is one of the strongest pieces of evidence we have that the standard model is incomplete.

One comment is in order. We have now repeatedly said that standard model CP violation is inadequate for baryogenesis. This is sometimes confused with the (incorrect) statement that the CP violation observed in the kaon system is too small to produce the observed baryon asymmetry. At the moment our knowledge of CP violation is not extensive enough to say definitively that the observed CP violation is associated with a phase in the CKM matrix. It is perfectly possible that CP violation in the kaon system is dominated by physics beyond the standard model. This would likely show up as a discrepancy between CP violation measured in the B system relative to the expectations from the K system.

If the standard model must be augmented with new CP violation to create the baryon asymmetry, what form is this new CP violation likely to take? We don't know. However it is worth noting that CP violation in the standard model, with its intimate connection to flavor symmetries, is rather special. In almost any extension of the standard model, new interactions and new particles allow for new sources of CP violation. Under these circumstances this new CP violation is not constrained by the standard model flavor symmetries and will typically give large effects. Indeed the apparent smallness of CP violation at low energies is a strong constraint on physics beyond the standard model, since most extensions of the standard model lead to large, even unacceptable, CP violating effects.

Most investigations of baryogenesis have focused on models proposed for reasons other than CP violation and the baryon asymmetry. For example, a natural extension of the original fundamental Higgs standard model includes multiple Higgs fields. With one or more new Higgs fields there are new CP violating couplings, the flavor structure of the model is different, and baryogenesis is certainly possible. A particularly popular extension of the standard model, the MSSM, has a number of new CP violating phases, and can easily have large CP violation at the electroweak scale. As we have discussed, the phase transition in this model may be too weak (depending on the latest bounds on the parameters of the Higgs potential) to allow electroweak baryogenesis, but most non-minimal extensions of this model (for example the inclusion of a new singlet superfield), allow a strongly first order phase transition consistent with current supersymmetry bounds. In grand unified models new CP violation may be associated with the scalar fields necessary to break the grand unified symmetry. Many examples of this type have been proposed.

This is in fact the best news from baryogenesis, especially electroweak baryogenesis. By bringing the physics of baryon production down to energies that we are currently probing in the laboratory, we have an opportunity to verify or falsify these ideas in detail. For example CP violation in the extensions of the standard model mentioned above, particularly supersymmetry, lead to observable effects at low energies, both CP conserving and CP violating. If the next round of collider experiments determine the nature of electroweak symmetry breaking, then the nature of the phase transition and its suitability for electroweak baryogenesis may be determined. If new CP violation is observed in experiments like the B factory, or in electric dipole moment experiments, it will be especially interesting to determine the flavor structure of this CP violation and its possible connection with the baryon asymmetry of the universe.

Although we have only touched on two broad areas of baryogenesis, electroweak and grand unified, there are a variety of other interesting ideas, including spontaneous baryogenesis, topological defects, *etc*. One of the more interesting variants, leptogenesis, involves the production of an asymmetry in lepton rather than baryon number. Subsequent production of baryon number then relies upon further processing of the lepton number asymmetry by interactions, like the electroweak interaction we have already discussed. These models are especially timely since the lepton asymmetry may be connected with the physics of neutrinos, an area where we are now beginning to obtain a great deal of experimental information.

The only bad news here, is the rather vague connection between baryogenesis and *specific* laboratory experiments. There is no single smoking gun; new CP violation large enough to produce the observed baryon asymmetry will almost certainly have low energy effects, but not decisively so. And where these effects show up, be it in EDMs, B or D mixing, or top quark physics, is highly model dependent. Without more experimental information constraining our current theoretical ideas, baryogenesis does not suggest that any one experiment is more likely than another to see new CP violation. But these are minor quibbles. Baryogenesis is already a strong indication of new physics to come, and even tells us that this new physics should emerge in one of the most fascinating areas of current research, CP violation.

Baryogenesis has been a fruitful cross-roads between particle physics and cosmology. Uniting ideas of early universe phase transitions, electroweak symmetry breaking and CP violation, it is an area that touches on many of the most exciting experiments that we look forward to in the coming decade. The B factory, the LHC, the Tevatron and even tabletop atomic physics experiments, may provide provide the clues that help explain the presence of matter in the universe. Unraveling the mystery of the cosmic baryon asymmetry remains one of the most exciting tasks for particle physicists and cosmologists alike.

References

[1] E. W. Kolb and M. S. Turner. *The Early universe*. Addison-Wesley, 1990.

[2] A. G. Cohen, D. B. Kaplan, and A. E. Nelson. Progress in electroweak baryogenesis. *Ann. Rev. Nucl. Part. Sci.*, 43:27–70, 1993.

[3] V. A. Rubakov and M. E. Shaposhnikov. Electroweak baryon number non-conservation in the early universe and in high-energy collisions. *Usp. Fiz. Nauk*, 166:493–537, 1996.

[4] Antonio Riotto and Mark Trodden. Recent progress in baryogenesis. 1999. hep-ph/9901362 submitted to Ann. Rev. Nucl. Part. Sci.

[5] G. Steigman. Observational tests of antimatter cosmologies. *Ann. Rev. Astron. Astrophys.*, 14:339–372, 1976.

[6] S. P. Ahlen. The ams experiment to search for antimatter and dark matter. In *Proceedings of the 5th Annual LeCroy Conference on Electronics for Particle Physics, Chestnut Ridge, NY*, 1995.

[7] S. P. Ahlen. Ams: A magnetic spectrometer for the international space station. In S. J. Ball and Y. A. Kamyshkov, editors, *Proceedings of the International Workshop on Future Prospects of Baryon Instability Search in p decay and n—anti-n Oscillation Experiments, Oak Ridge, TN*, 1996.

[8] R. Battiston. The alpha magnetic spectrometer (ams): Search for antimatter and dark matter on the international space station. *Nucl. Phys. Proc. Suppl.*, 65:19, 1998.

[9] R. Battiston. The alpha magnetic spectrometer (ams). *Nucl. Instrum. Meth.*, A409:458, 1998.

[10] V. Plyaskin. Antimatter and dark matter search with the alpha magnetic spectrometer (ams). *Surveys High Energ. Phys.*, 13:177, 1998.

[11] J. Casaus. The alpha magnetic spectrometer (ams). *Acta Phys. Polon.*, B30:2445, 1999.

[12] U. Bekcer. Alpha magnetic spectrometer ams report on the first flight in space june 2-12. In *Proceedings of the 29th International Conference on High-Energy Physics (ICHEP 98), Vancouver, British Columbia, Canada*, 1998.

[13] A. G. Cohen, A. De Rujula, and S. L. Glashow. A matter-antimatter universe. *Astrophys. J.*, 495:539, 1998.

[14] S.C. Kappadath et al. In *Proc. 12th Int. Cosmic Ray Conf.*, page 25, 1995.

[15] William H. Kinney, Edward W. Kolb, and Michael S. Turner. Ribbons on the cbr sky: A powerful test of a baryon symmetric universe. *Phys. Rev. Lett.*, 79:2620–2623, 1997.

[16] A. G. Cohen and A. De Rujula. Scars on the cbr? *Astrophys. J.*, 496L:63, 1998.

[17] Keith A. Olive, Gary Steigman, and Terry P. Walker. Primordial nucleosynthesis: Theory and observations. 1999. astro-ph/9905320 submitted to Phys. Rep.

[18] A. D. Sakharov. Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe. *JETP Letters*, 5:24, 1967.

[19] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov. On the anomalous electroweak baryon number nonconservation in the early universe. *Phys. Lett.*, B155:36, 1985.

[20] G. 't Hooft. Symmetry breaking through bell-jackiw anomalies. *Phys. Rev. Lett.*, 37:8–11, 1976.

[21] G. 't Hooft. Computation of the quantum effects due to a four- dimensional pseudoparticle. *Phys. Rev.*, D14:3432–3450, 1976.