9 Baryogenesis

We now turn to a more speculative subject, baryogenesis, providing an explanation for the origin of matter in the universe. Recall that from nucleosynthesis one derives

$$4(3) \times 10^{-10} \lesssim \eta \lesssim 7(10) \times 10^{-10} \ . \tag{169}$$

Assuming that there are no anti-baryons [2], this is equivalent to the bounds on $\eta = n_B/n_\gamma$:

$$0.015(0.011) \lesssim \Omega_b h^2 \lesssim 0.026(0.038) \tag{170}$$

where the figures in parentheses are the most conservative bounds one can imagine. We actually see fewer baryons than that, with $\Omega_b^{luminous} h^2 \lesssim 0.01$. There are several pieces of evidence that the baryons we see are truly baryons and not sequestered anti-baryons. For one thing, cosmic ray anti-protons are only seen at the 10^{-4} level, consistent with their production in the atmosphere from proton collision, and anti-nuclei have not been discovered in cosmic rays. Secondly, we know that in the intergalactic space a lot of gas resides, which means that galaxies are *not* totally sequestered from each other. Thus is there were regions of matter and anti-matter, we would have expected to see annihilation processes at their interfaces, in the form of high energy photons.

Of course, it is possible that the universe has zero net baryon number, but that within our horizon there is an excess of baryons, compensated for by an excess of antimatter in some other horizon. This idea cannot work without invoking inflation however. Suppose there was a mechanism at early time to produce regions where $n_B = \pm n_\gamma$ with random sign. The maximum size for these regions from causality arguments would be the horizon size at that time. Our current horizon would therefore contain $N = (H/H_0)^3$ such domains, and therefore we would find today $\eta \simeq 1/\sqrt{N}$. Even if this epoch which produced the huge baryon number fluctuations was as low as $T = 1$ TeV, we would find $N \simeq 10^{42}$, which would lead to a value for η over ten orders of magnitude too small too explain the baryon asymmetry we see.

Recall from eq. (122) that $n_B/s \simeq \eta/7$ is a constant during the adiabatic expansion of the universe. Therefore n_B/s was about 10^{-10} when the temperature was far above the nucleon mass. At that epoch, there would have been quarks and anti-quarks in thermal equilibrium, with densities $(n_b + n_{\bar{b}}) \simeq s$ (up to some not too large factor), so that the early universe would have to exhibit the tiny asymmetry between matter and anti-matter of $(n_b - n_{\bar{b}})/(n_b + n_{\bar{b}}) \sim 10^{-10}$.

Sakharov was the first to consider the possibility that the matter/anti-matter asymmetry could arise from micro-physical processes during the big bang. He discovered three conditions that would have to be fullfilled for this to be the case:

1. Baryon number violating processes must exist;
2. C and CP violation must exist;
3. There must have been an epoch when baryon violating processes whent out of theremal equilibrium.

[2]Recall that protons and neutrons have baryon number $B = 1$; anti-nucleons have baryon number $B = -1$. Quarks and anti-quarks have baryon number $+\frac{1}{3}$ and $-\frac{1}{3}$ respectively. I will use n_b and $n_{\bar{b}}$ to represent the density of baryons and anti-baryons respectively, while $n_B = (n_b - n_{\bar{b}})$ is the net baryon number density; Ω_b will represent $(\rho_b + \rho_{\bar{b}})/\rho_c$.

The reason for the first rule is obvious: if the universe begins with no net baryon number, and ends up with nonzero baryon number, baryon number must not be conserved. In the $SU(3) \times SU(2) \times U(1)$ standard model (SM), the lowest dimension operators one can write down which violate baryon number are dimension six, taking the form

$$\frac{qq q\ell}{\Lambda^2} \tag{171}$$

where the q's are quark fields, ℓ is a lepton, and Λ has dimension of mass. The lifetime of the proton ($\tau \gtrsim 10^{34}$ years) forces Λ to be a very large scale, since the above operator can lead to the decay $p \to \pi^0 e^+$, or $p \to K^+\nu$, etc. Since $\tau \sim m_p^5/\Lambda^4$, one finds $\Lambda \gtrsim 10^{16}$ GeV, which is the grand unified theory (**GUT**) scale. Operators such as eq. (171) can be generated, in grand unified theories, for example, by the exchange of bosons with mass M and coupling g, with $g^2/M^2 \simeq 1/\Lambda^2$. The large Λ scale implies that if baryogenesis is to exploit this hypothesized source of baryon violation, it would have to occur at very high temperature, and is referred to as "GUT-scale" baryogenesis.

Another possible source of baryon number violation is the standard model, surprisingly enough. 't Hooft showed that because of an **anomaly**, neither baryon number B nor lepton number L are conserved by the weak interactions, although the combination $(B-L)$ is. At zero temperature, he found that

$$\Delta B = \Delta L \propto N_f e^{-4\pi/\alpha_w} \ , \tag{172}$$

where N_f is the number of families (3) and α_w is the weak fine structure constant, $\alpha_w \simeq 1/25$, and so the exponential suppression is fantastically small. However, it has been shown that at finite temperature $T > 100\,\text{GeV}$ (the scale of the W and Z boson masses) the electroweak B violation rate is much more rapid, $\Gamma_{\not{B}} \simeq 30 \times \alpha_w^5 T$. This opens the possibility for having baryogenesis occur at the electroweak scale. It also presents a challenge for GUT scale baryogenesis, as electroweak baryon violation can wash out a baryon asymmetry produced earlier. A solution to this quandry is to produce nonzero $(B-L)$ charge at the GUT scale, since this quantum number is conserved by the electroweak interactions. Note that the operator eq. (171) won't do for generating nonzero $(B-L)$, however, as it also consserves $(B-L)$.

The second of Sakharov's rules follow similar reasoning. A baryon charge is odd under both of the discrete symmetries C, which interchanges quarks with anti-quarks, and CP, which interchanges quarks with anti-quarks and followed by a parity flip. Since we wish to have the universe start off initially any net charges, the production of a B asymmetry must necessarily involve both C and CP violation. It is worth commenting that the SM violates C maximally since C would exchange a left-handed neutrino with a left-handed anti-neutrino, but the latter does not exist in the SM. In contrast, CP would exchange a left-handed neutrino with a right-handed anti-neutrino, which does exist. Nevertheless, it is known that CP violation exists in nature, probably in the SM, as is observed in the kaon and B meson systems. Nevertheless, the CP violation in the standard model appears to be much too small to account for the baryon asymmetry we see. The effects are large in the B and K meson systems because we have nearly degenerate states there, which have large mixing effects induced by very small interactions. However, at finite temperature, all states are broad, and there should not be any such enhancement of CP violation. Then a reparametrization invariant estimate of SM CP violation gives the dimensionless measure $\delta_{CP} \simeq 10^{-20}$, much too small to explain $\eta \sim 10^{-10}$. Therefore it is a pretty robust prediction of baryogenesis that there must exiast new sources of CP violation, which ould possibly be measurable in experiments search for electric dipole moments in quarks and leptons [3].

The third of Sakharov's rules mandates departure from thermal equilibrium, and is a consequence of CPT symmetry. The idea is that CPT symmetry implies that for every state with baryon number B and energy E, there exists a state with baryon number $-B$ and identical energy E. (For example, CPT symmetry implies that proton and anti-proton have the same mass). Thus in thermal equilibrium, states with baryon number B and $-B$ will be equally populate4d, and the net baryon number will have to be zero, even if there are B- and C- and CP-violating processes. Formally, if Ω is a CPT transformation, then $\Omega\hat{B}\Omega^{-1}{=}-\hat{B}$, and $\Omega\hat{H}\Omega^{-1} = \hat{H}$, so that

$$\begin{aligned}\langle B\rangle &= \text{Tr }\hat{B}e^{-\beta\hat{H}} \\ &= \text{Tr }\Omega^{-1}\Omega\hat{B}\Omega^{-1}\Omega e^{-\beta\hat{H}} \\ &= \text{Tr }\left(\Omega\hat{B}\Omega^{-1}\right)\left(\Omega e^{-\beta\hat{H}}\Omega^{-1}\right) \\ &= \text{Tr }-\hat{B}e^{-\beta\hat{H}} \\ &= -\langle B\rangle \end{aligned} \tag{173}$$

implying $\langle B\rangle = 0$.

We have seen that departure from equilibrium can occur whenever there are relic particles which cannot annihilate or decay fast enough as the temperature drops. Annihilation processes qunch themselves, as the number density of the relic drops, since the rate is proportional to the relic density. Therefore one will typically depart from thermal equilibrium whenever the lifetime of the relic is long compared to the Hubble time, $1/H$. Note that $1/H = 3M_{Pl}/8\pi\sqrt{g_*}T^2$ which is a very short time scale at the GUT scale $(8\pi/3)\sqrt{g_*}T \sim M_{pl}/5$, since the universe is expanding so quickly then. However, at the electroweak scale, one has $(8\pi/3)\sqrt{g_*}T \sim M_{pl}^{-16}$ and so only incredibly weakly interacting particles can go out of thermal equilibrium. However, one can also go out of thermal equilibrium during a first order phase transition (when ther is typically supercooling, and then explosive formation of bubbles of the new low temperature phase). For this reason, electroweak baryogenesis is associated with a strong first order electroweak phase transition. Other possibilities exist for nonequilibrium epochs, such as during the period of reheating after inflation.

GUT-scale baryogenesis Now I will sketch several scenarios for baryogenesis, starting with a GUT scale model. Suppose there are X and $\bar{X}$ bosons which can either decay into 2 quarks or an anti-quark and a lepton.

particle	final state	B_{final}	branching ratio
X	qq	$2/3$	r
X	$\bar{q}\bar{\ell}$	$-1/3$	$1-r$
$\bar{X}$	$\bar{q}\bar{q}$	$-2/3$	$\bar{r}$
$\bar{X}$	$q\ell$	$1/3$	$1-\bar{r}$

(174)

We can define a measure of the asymmetry in baryon number production

$$\epsilon_X = \sum_{\text{final state f}} B_f\,\frac{\Gamma(X\to f)+\Gamma(\bar{X}\to\bar{f})}{\Gamma_X} \tag{175}$$

[3]Electric dipole moments violate T (time reversal symmetry) since the spin of a particle reverses under T, but its charge distribution does not. However, in a Lorentz-invariant and unitary (probability conserving) theory, CPT is an automatic symmetry, and so T violation implies CP violation, which is harder to measure directly.

where Γ_X is the total width and B_f is the baryon number of the state f. By CPT, the total widths of the X and its antiparticle $\bar{X}$ are equal; their branching ratios into a particular final state f may be different, provided there is C, CP violation. In the present example, $\epsilon_X = (r - \bar{r})$. The scenario then is that X and $\bar{X}$ bosons are massive, and that at the time when $T \sim M_X$, their annihilation is not efficient, and more importantly, their lifetime is much longer than the Hubble time. Suppose their width is given by $\Gamma = \frac{\alpha}{4\pi} M_X$, where $\alpha = g^2/4\pi$ and g is a typical coupling of the X. Then we are requiring that $\alpha/4\pi > \sqrt{g_*} M_X/M_{Pl}$, or $M_X > \left[\alpha/(4\pi\sqrt{g_*})\right] M_{Pl}$. We know that $G_* \simeq 100$ in the SM, and could be ~ 1000 in a GUT. So for $\alpha/4\pi \sim 10^{-3}$ it is reasonable to have M_X be as low as $10^{14} - 10^{15}$ GeV. Ignoring annihilation then, which will shut off as the particle density gets lower anyway,, the X's and $\bar{X}$'s will hang around for a while, both with number density $n_X = n_{\bar{X}} \simeq n_\gamma$ until they decay at a time $t \sim \Gamma_X^{-1}$. They decay with their unequal branching ratios, and produce a baryon asymmetry $n_B/n_\gamma \simeq \epsilon_X$. The reverse processes which remove baryon number, $f \to X$ and $f \to \bar{X}$, are out of thermal equilibrium and do not proceed, as the particles in the final state f now have an energy $T \ll m_X$ and the process will have an $e^{-m_X/T}$ suppression.

It is the baryon to entropy ratio that is conserved, rather than the baryon to photon ratio, so what one finds today is $\eta \simeq \epsilon_X g_*^s(\text{today})/g_*^s(T = M_X)$ which could give another 10^{-2} suppression factor.

This difference $(r-\bar{r})$ is naturally small, since it cannot arise in leading order in perturbation theory, but must arise as an interference term between a tree-level and a one-loop amplitude. At tree level, for example, the amplitude for $X \to qq$ could equal g, but then the amplitude for $\bar{X} \to \bar{q}\bar{q}$ would equal g^*. If $g \neq g*$, that is a sign of CP violation; however, even if there is CP violation present, one finds at tree level $\Gamma_{X\to qq}/\Gamma_{\bar{X}\to\bar{q}\bar{q}} = |g|^2/|g^*|^2 = 1$. In fact, this feature would appear to persist to higher order perturbation theory: for every graph $X \to f$, there is an analogous graph for $\bar{X} \to \bar{f}$ with all the arrows on the propagators reversed and all couplings replaced by their complex conjugates. Since the widths depend only on the absolute value of the amplitude, they can differ only if there is some source of complex phase in the amplitudes which is not due to complex coupling constants. There is: the $i\epsilon$ in the propagators of internal particles in the Feynman diagrams. So to get $r \neq \bar{r}$ it is necessary to have (i) loops involving internal propagation of particles, and (ii) the kinematics must be such that the internal particles in the diagram could be on shell, so that the amplitude is sensitive to the $i\epsilon$ pole prescription. We see that at the very least, a nonzero value for ϵ_X requires an interference between a tree level and one-loop amplitude, and so there will be typically a factor of $\alpha/4\pi$ suppression, where α is a coupling constant of the theory, on top of suppression factors from the measure of CP violation (often quite small) and from kinematic factors. Note that α has to be small, otherwise the X's could decay quickly and would not go out of thermal equilibrium in the first place.

Various problems to overcome with GUT-scale baryogenesis: (i) If there is inflation, the baryon asymmetry one has produced will typically be inflated away; (ii) If one does not produce a $(B - L)$ asymmetry (and the above example does not), then the anomalous electroweak processes will equilibrate the asymmetry to zero. (This problem is turned into a feature in models of **leptogenesis** where a lepton asymmetry is produced at the GUT scale, which is then equilibrated into nonzero baryon asymmetry by the electroweak anomaly).

Electroweak baryogenesis As mentioned above, there is a phase transition at a temperature of several hundred GeV where the Higgs field H gets an expectation value and breaks the $SU(2) \times U(1)$ gauge symmetry, Phase transitions are typically first order, or second order.

In second order phase transitions, the order parameter ($\langle H \rangle$ in this case) changes continuously at the phase transition, although its derivative with respect to temperature is discontinuous. In the case of a first order phase transition, however, the order parameter has to jump discontinuously, tunnelling through a barrier in the free energy. Typically this process will involve super-cooling, where the universe gets trapped in the symmetric minimum $H = 0$ below the temperature where it ceased to be the point that minimizes the free energy. Eventually then the vacuum tunnels to the correct ground state, y nucleating a bubble of the true vacuum, which then expands and takes over, releasing latent heat in the process. This is a nonequilibrium proces.

Whether or not the electroweak phase transition is first order depends on details of th Higgs interaction about which we have know direct knowledge. If the phase transition was first order, then a rather complicated scenario has been developed for how the baryon asymmetry can be generated. As the bubble walls expand, the "snow-plow" particles in front of them — particularly the top quark, which is massless outside the bubble (symmetric phase) and heavy inside the bubble (broken phase). With CP violating interactions between the top quark and the Higgs field, it is possible to have more anti-top quarks than top quarks pushed in front of the expanding wall. In this symmetric phase, baryon violation is rapid, and tries to equilibrate the anti-top excess by producing net baryon number. As soon as the baryons produced get engulfed by the expanding bubble, they aree in the broken symmetry phase where the baryon violating processes are exponentially suppressed, and so the baryon asymmetry is preserved.

Electroweak baryogenesis has the advantage of being late, so that inflation can be well over with before the baryon asymmetry is produced. It is also appealing because extra CP violation at the weak scale may be testable, and because having the trnasition be first order places constraints on the Higg sector, that can also be experimentally relevant.

Other scenarios There are several other popular scenarios for baryogenesis, which I will just mention. **Affleck-0ine** baryogenesis exploits the fact that in supersymmetri theories, the potentials for scalar fields often have flat directions. In this scenario a large coherent field carrying baryon number develops, and which preferentially decays into quarks. In **leptogenesis**, as mentioned above, a lepton number asymmetry is produced, and subsequently turned into a baryon asymmetry by anomalous weak interactions. There are also theories where baryogenesis is associated with the reheating phase after inflation.