How Do We Know Antimatter Is Absent?

Paolo S. Coppi
*Max-Planck-Institut für Kernphysik, Postfach 10 39 80, 69029 Heidelberg, Germany**

1. INTRODUCTION

One of the great mysteries of modern particle physics is the relative lack of antimatter in our vicinity, a fact that even first year students may easily verify by touching their laboratory equipment and noting the lack of any explosive after effects. (As *Star Trek* teaches us, bad things happen when matter and antimatter come into contact and annihilate...) This is somewhat puzzling given that the relativistic field theories which underlie modern particle physics have built into them a fundamental symmetry: for every particle there is an antiparticle with quantum numbers and charges of the opposite sign. In the old language of Dirac, for example, the relativistic wave equation for the electron has both positive and negative energy solutions, which led Dirac[1] to predict the necessary existence of the positron. In more modern field theory language, we think of creation or annihilation operators acting on a field that respectively create particles and destroy antiparticles or destroy particles and create antiparticles associated with the field. The properties and dynamics of the particles and antiparticles are fundamentally related. If we consider local, Lorentz-invariant field theory equations such as currently used for the "Standard Model" of particle physics, and we flip the signs of all charges that appear in them, effectively turning particles into antiparticles, and then perform a space reversal ($\vec{x} \rightarrow -\vec{x}$) followed by a time reversal ($t \rightarrow -t$), we recover the same equations. This symmetry of the equations, the so-called "CPT" (Charge-Parity-Time reversal) symmetry, implies, for example, that particle and antiparticles should have exactly the same masses. Similarly, if hydrogen atoms exist, then anti-hydrogen atoms with the same characteristic energy levels as hydrogen atoms could in principle exist too (and in fact, are being made now at CERN[2]). Stretching our imagination further, the Universe could then be filled with antimatter stars and galaxies that are indistinguishable from ordinary stars and galaxies if one studies them solely via their light emission or their gravitational attraction on neighboring bodies.

The absence of antimatter equipment in student laboratories is not in and of itself a problem since the fact that such equipment *could* exist does not prove that it *should*. What is a problem, though, is coming up with a way to create a region of space, like a laboratory, that does not contain equal amounts of matter and antimatter. The reason is that the equations of the well-tested Standard Model also imply that total charge as well as quantum numbers like baryon number and lepton number should be conserved in particle interactions. Let us therefore consider a region that initially has no baryons (i.e., protons and neutrons) and no anti-baryons, i.e., a total baryon number of zero. Now let us try to put some baryons there. If we are not going to simply transport baryons over from another region of space, we must create them, e.g., through the collision of energetic particles of other types (as presumably happens during the hot, early phase of the universe). Unfortunately, since such interactions must conserve baryon number, they always result in *pairs* of baryons and antibaryons being produced, i.e., we create equal amounts of matter and antimatter and the net baryon number (baryons minus antibaryons) remains zero. Another alternative might be to start with equal amounts of matter and antimatter and selectively remove the antimatter. If we rule out the possibility of spatial transport, then this does not work either since the Standard Model's baryon number conservation means antibaryons (or baryons) cannot spontaneously decay into particles with less net baryonic content. This leaves baryon-antibaryon annihilation as the only means to decrease the number of antibaryons, which of course decreases

---

*On sabbatical leave from Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 0652-8101, USA the number of baryons in the same way, leading again to no net change in baryon number.

The existence of just a single region of space with a significantly non-zero baryon number therefore leads to the following two important implications. If our local environment is representative of the Universe as a whole, for example, then there is essentially no antimatter in the Universe and the Universe has a net baryon number, $B > 0$. Since standard particle physics can only add or remove baryons and anti-baryons in pairs, i.e., it cannot change B, the inferred $B > 0$ must then represent an initial condition. To many, this is an ugly or unnatural possibility given that matter and antimatter otherwise appear so symmetric. If our environment is instead not typical, we can partially rescue the so-called "symmetric" Universe hypothesis, where the universe has $B = 0$ and thus equal amounts of matter and antimatter overall, by postulating a spatial transport mechanism that rearranges and separates the matter and antimatter into distinct spatial "domains." While such a Universe may appear more elegant than an intrinsically "baryon asymmetric" one, there is a price to be paid here too. Although attempts were made to find one, e.g., [90], we currently know of no satisfactory mechanism to produce such a separation, and even if such domains can be created, one must be careful. If they come into contact, for example, matter and antimatter domains start to annihilate, either vanishing completely or doing possibly bad things such as lighting up the Universe with the annihilation gamma-rays.

Because neither possibility is that attractive and they are the only ones allowed by Standard Model physics, Sakharov[3] proposed a radical alternative: our physics is wrong! More precisely, there is new physics beyond the Standard Model which, at higher energies than can currently be tested with accelerators, allows for baryon number violation. (Unlike the CPT symmetry, which follows from the form of the relativistic field equations and ultimately Lorentz invariance, there is nothing especially fundamental about baryon number conservation.) A natural time for this physics to become important would be during the Big Bang, when the Universe was extremely hot. Sakharov therefore set down his three conditions (the existence of baryon number violation, "CP" symmetry violation, and a departure from thermal equilibrium) which if satisfied could lead to "baryogenesis" during the Big Bang, i.e., the dynamical generation of a baryon asymmetry from an initially symmetric universe. Because in Sakharov's days a mechanism of baryon number violation was in some ways as speculative as a mechanism for separating matter and antimatter, the idea languished for several years. As the observational evidence grew that our Universe had a baryon asymmetry not just on the scale of student laboratories but on the larger scales of galaxies and galaxy clusters, e.g., [4], researchers acknowledged there was a problem but did not quite know what to do about it.

With the arrival of GUT theories that explicitly predicted proton decay, i.e., baryon number violation, Sakharov's baryogenesis scenario spring to life again, this time with concrete physical equations to back it up, e.g., see [5]. Further fuel was added to the baryogenesis fire by the realization that many problems in cosmology could be elegantly solved if our Universe had gone through an epoch of "inflation," i.e., a phase of rapid exponential expansion. Unfortunately, the likely baryon asymmetry of the Universe was not one of them as inflation meant that any initial baryon number would be "inflated away," i.e., exponentially reduced by at least $\sim \exp(-210)$ if we wanted to ensure the other good properties of inflation. After inflation, then, the Universe would contain essentially no baryons and antibaryons and have $B \approx 0$, leading to the difficulties discussed above if we are sure baryon number is conserved and parts of the Universe have $B \neq 0$. Sakharov's idea now seemed quite compelling and the number of papers on explicit baryogenesis scenarios exploded – after all, our very existence in the Universe was at stake! As the idea of an inflationary epoch seems to be only gaining ground as new data comes in, some today will even go as far as to say that the Universe's apparent baryon asymmetry is the strongest evidence we have yet for baryon number violating processes and ultimately new physics beyond the Standard Model. (We have only have upper limits on the proton decay rate.)

All this excitement about new physics is, of course, predicated on there actually being a baryon asymmetry in the Universe. A rather crucial question to answer experimentally, therefore, is exactly how baryon asymmetric is our Universe and what are the characteristics of this asymmetry? Is primordial antimatter completely gone from today's Universe, as most baryogenesis scenarios predict, or, as others speculate, do we end up creating distinct domains of matter and antimatter, and if so, what is the distribution of the spatial sizes and baryon numbers of these domains? This then is the main charge of this lecture: to review and explore our evidence for a baryon asymmetry in the Universe. Although we still do not have completely conclusive answers to the questions just posed, now is a particularly interesting time to revisit this topic. If this lecture had been given just two years ago, before the release of the WMAP satellite's data on the Cosmic Microwave Background (CMB), it would have largely been a repeat of Steigman's excellent review of the subject [4], published almost thirty years ago. Besides helping to nail down the space-time parameters of the present-day Universe, e.g., showing that inflation's prediction of a flat universe appears correct, the WMAP satellite provides key information on the average value of the Universe's baryon density as well as on the magnitude and spatial distribution of baryon density fluctuations, which on the largest ("super horizon") scales directly reflect what happened during the epoch of inflation. This information proves to be quite constraining when combined with the COBE satellite's highly accurate measurement of the CMB spectrum and the improved data now available on the primordial abundances of light elements produced during the era of "Big Bang Nucleosynthesis."

This lecture is organized as follows. In section 2, I run through the constraints that we have on the existence of antimatter in the nearby Universe. This section, based on Steigman's review[4], summarizes his main arguments and updates his limits to take into account the improvements in gamma-ray and cosmic detector technology that have occurred since his review was written. In section 3, I introduce some of the key concepts that we will need to understand the cosmology-based constraints discussed in section 4 and 5. I then discuss in more detail why we and objects like baryon-dominated galaxy clusters should not exist in the standard cosmology (a Big Bang that includes an inflationary epoch and only the physics of the Standard Model), and I describe the general constraints that any successful baryogenesis model must satisfy. This is where we shall meet the now-famous Sakharov conditions that, interestingly, were only hinted at in a footnote of Steigman's review. In section 4, the meat of the lecture, I discuss how the recent dramatic improvements in our understanding of cosmology allow us to put significant constraints on the characteristics of the matter and antimatter domains that might inhabit our universe. In section 5, I discuss the tightened cosmological constraints that some specific baryogenesis models must confront even if they do not predict distinct matter and antimatter domains. I summarize the main points of the lecture in section 6.

2. CONSTRAINTS ON ANTIMATTER IN THE LOCAL UNIVERSE

2.1. What Happens When Matter Meets Antimatter?

When matter meets antimatter, what happens next is probably well-known to almost everyone: annihilation and the transformation of the annihilating matter's rest mass into energy particles and radiation with 100% efficiency. To better appreciate what 100% conversion of rest mass energy means, note that a supermassive black hole swallowing up matter at the center of a galaxy is thought to able to convert just $\sim 10\%$ of the rest mass energy that it accretes into radiation. If that black hole, however, manages to swallows just one star's worth (one solar mass) of gas per year it will outshine its entire host galaxy (containing $\sim 10^{10}$ stars!) by a factor $\sim 10^3$ to produce a so-called quasar or Active Galactic Nucleus (AGN) that can be detected out to close to the edge of the observable universe. Imagine then what would happen if just one part in $\sim 10^{13}$ of the matter in gas and stars in our galaxy annihilated every year? (Not surprisingly, early quasar models invoked matter-antimatter annihilation as the energy source.) The products of matter-antimatter annihilation are therefore what we shall use to obtain our most powerful constraints on the presence of antimatter in our Universe. Because the rest mass of neutral matter is dominated by the rest mass of protons, neutrons, and nuclei, and in turn the mass of most gas in the Universe is dominated by protons, we shall largely be concerned with the annihilation of protons on anti-protons. Note that if we can somehow suppress the annihilation of matter with antimatter, we will void most of the constraints discussed here.

Now what does a typical annihilation event look like and how would we recognize it? From [4], pions are the primary products from the annihilation of nucleons (protons and neutrons) with antinucleons. These are unstable, however, and immediately decay. A typical annihilation/decay chain then looks like,

$$N + \bar{N} \rightarrow \begin{cases} \pi^0 \rightarrow \gamma + \gamma \\ \pi^{\pm} \rightarrow \mu^{\pm} + \nu_\mu(\bar{\nu}_\mu) \end{cases}$$

Table I: Matter-Antimatter Annihilation Products

Particle	N^a	$E_{Tot}(Mc^2)$	E_{Av}**(MeV)**
$e^{\pm}$	~ 3	$\sim \frac{1}{3}$	~ 100
γ	~ 3	$\sim \frac{2}{3}$	~ 200
$\nu_e, \bar{\nu}_e$	~ 3	$\sim \frac{1}{3}$	~ 100
$\nu_\mu, \bar{\nu}_\mu$	$\sim 6^d$	$\sim \frac{2}{3}$	~ 100

From [4].

[a] Number of each type produced in an annihilation.

[b] Total energy, in units of the nucleon rest-mass energy, carried off by all particles of the given type.

[c] The average energy of each type of particle: $E_{Av} = E_{Tot}/N$.

[d] Only muon neutrino with $E > m_\mu c^2$ are detectable. $N_{\nu_\mu}(E > m_\mu c^2) \sim 2$.

where the muons are also unstable and decay,

$$\mu^{\pm} \rightarrow e^{\pm} + \nu_e(\bar{\nu}_e) + \nu_\mu(\bar{\nu}_\mu),$$

leading in the end to the production of energetic electrons and positrons, gamma-rays and neutrinos. The typical numbers and mean energies of the final decay products are given in Table 1, also from [4]. These decay products can be mimicked by other processes, in particular the interaction of cosmic rays with interstellar matter and background radiation fields such as the cosmic microwave background (CMB) and the mean stellar radiation field inside our galaxy. One therefore has to be careful in interpreting the detection of a large gamma-ray or positron flux as an indication of annihilation. Still, the production of ~ 200 MeV gamma-rays is not a typical event in our every day life, and if one shook hands with one's anti-neighbor, the results would be quite spectacular and noticeable. As evidenced by the determined experimental effort that was required to find astronomical gamma-ray sources, 200 MeV gamma-ray emission is also not that common in the overall Universe. We can therefore hope (and it is the case) that limits on ~ 200 MeV gamma-ray fluxes from various classes of object will in fact be interesting.

Although we now know what signals to look for, to get the magnitudes of the expected signals we still need to know the rate at which matter and antimatter in contact will annihilate. For the level of accuracy required here (factors of a few), we may use the simple approximations for the annihilation cross-section collected in Steigman's review[4]. At relativistic energies ($E_{nuc} \gtrsim 1$ GeV), one may take the annihilation cross-section to be roughly independent of energy:

$$\sigma_{Ann}(E \gtrsim 1 \text{ GeV}) \approx 5 \times 10^{-26} \text{ cm}^2. \tag{1}$$

In the trans-relativistic regime $0.4 \lesssim E(\text{GeV}) \lesssim 7$, a good fit to the rate coefficient $< \sigma_{Ann} v >$ is provided by

$$< \sigma_{Ann} v > \approx 1.5 \times 10^{-15} \text{ cm}^3 \text{ s}^{-1}. \tag{2}$$

(In terms of this coefficient, the annihilation rate is $n_N n_{\bar{N}} < \sigma_{Ann} v >$ where n_N and $n_{\bar{N}}$ are respectively the nucleon and antinucleon particle number densities.)

Below ~ 400 MeV, one must be careful and there has been some confusion in the past on what rates to use, e.g., below 1 MeV the Coulomb attraction of protons and antiprotons becomes important. For the typical astrophysical applications we will consider here, e.g., gas in galaxy clusters, a Maxwellian (thermal) distribution of even lower energies ($\lesssim 10$ keV) is of interest. For temperatures $10^4 \lesssim T \lesssim 10^8$ K, a good fit for the rate coefficient is,

$$< \sigma_{Ann} v > \approx 10^{10} T^{1/2} \text{ cm}^3 \text{ s}^{-1}. \tag{3}$$

In this temperature range, where protons and antiprotons are still free (instead of being bound up in atoms), one might wonder if a more important contribution to the rate comes from the radiative recombination of the protons and antiprotons to form a proton-antiproton atom (the analog of "positronium" in the case of annihilating electrons and positrons) which then annihilates with an atomic cross-section. Because radiative recombination is an electromagnetic process, however, the overall rate for recombination followed by annihilation is smaller by $(m_e/m_p)^{3/2} \sim 10^{-5}$, where m_e/m_p is the electron to proton mass ratio. Similarly, direct electromagnetic annihilation ($p\bar{p} \rightarrow \gamma\gamma$) which is the dominant process for electron-positron annihilation is reduced by $(m_e/m_o)^2 \approx 3 \times 10^{-7}$. For temperatures lower than 10^4 K, most protons and antiprotons are instead incorporated into atoms. In this case, the annihilation rate coefficient has a magnitude,

$$10^{-10} \lesssim <\sigma_{Ann} v> \lesssim 10^{-9} \text{ cm}^3 \text{ s}^{-1}, \tag{4}$$

for temperatures $10 \lesssim T \lesssim 10^4$ K.

For fun, let us now consider how long a hand would survive if it was placed in contact with an anti-neighbor's hand or one landed a matter probe onto an antimatter planet. A useful expression that we will use later is the characteristic time it takes matter to annihilate. For a background (target) antimatter density $n_{\bar{N}}$, this can be estimated as $t_{Ann} \sim n_N/(dn_N/dt_{Ann}) \sim (n_{\bar{N}} <\sigma_{Ann} v>)^{-1}$. For simplicity, let's assume the anti-planet and your anti-neighbor (as well as yourself) have a characteristic temperature ~ 300 K and a typical mass density comparable to that of water, ~ 1 g cm^{-3}, which corresponds to a number density $n \sim 6 \times 10^{23}$ cm^{-3}. Assuming $<\sigma_{Ann} v> \approx 5 \times 10^{-10}$ cm^3 s^{-1} (eqn. 4), we then obtain an annihilation time $\sim (6 \times 10^{23} \times 5 \times 10^{-10})^{-1} \approx 3 \times 10^{-15}$ sec. This estimate is likely too short, but even if we were to perform a more realistic calculation that includes the feedback from the energetic annihilation products, we would still get something very short, i.e., quite an explosion! Antimatter neighbors can be rather hazardous ... Note that even for interstellar densities, $n \sim 1$ cm^{-3}, the annihilation time is still only ~ 100 years. Antimatter that is not being continually regenerated, e.g., by cosmic ray interactions, should disappear quite quickly from our galaxy.

2.2. Antimatter in the Solar System

Although we may theoretically prefer a baryon symmetric Universe with equal amounts of matter and antimatter, if there is one part of our Universe that is almost certainly matter dominated, it is our solar system. On Earth, if there were significant chunks of antimatter lurking about that we had not created with our particle accelerators, we would have certainly heard about it on the evening news. As far as anti-planets go, we have by now landed matter probes onto several of them that did not immediately vaporize on contact.

Perhaps the strongest argument for matter asymmetry in the solar system, however, is the continuous outflow of particles from the Sun, i.e., the solar wind, that permeates most of it. To zeroth order, one can treat the solar wind as a constant velocity spherical wind, with a particle number flux of $nv \approx 2 \times 10^8 (d/1\text{AU})^{-2}$ cm^{-2} s^{-1}, where d is the distance from the Sun. One can worry about the deflection of the solar wind by magnetic fields around the planets, but usually there are always places, e.g., the magnetic poles, where solar wind particles can leak in (and the moon, at least, has no significant magnetic field and solar wind particles definitely make it onto the surface). The bottom line is that the accretion and subsequent annihilation of the solar wind on anti-planets would make them the brightest gamma-ray sources on the sky. From Steigman[4], if we use the solar wind flux just quoted, a rough estimate for the annihilation flux expected from an anti-planet is then,

$$F_\gamma(\sim 100\text{MeV}) \approx 10^8 (r/d)^2 \text{ photons cm}^{-2} \text{ s}^{-1}, \tag{5}$$

where r is the radius of the anti-planet intercepting the solar wind and d is the distance from us. Let's compute the flux for say, Jupiter, which has $r \approx 7 \times 10^9$ cm and a distance $d \approx 7 \times 10^{13}$ cm. In this case, we get $F_{\gamma,Jup} \approx 1$ photon cm^{-2} s^{-1}. This may not seem like much until we realize that that the typical (one year integration) flux sensitivity of the EGRET gamma-ray detector flown in the early 1990s is $\sim 10^{-7}$ photons cm^{-2} s^{-1} at 100 MeV. Note that the annihilation flux expected from the moon is even larger. Now EGRET did actually detect the moon(!)[6], but the flux was only $\approx 4 \times 10^{-7}$ photons cm^{-2} s^{-1}... (consistent with that expected from the impact of matter cosmic rays on the lunar surface).

2.3. Annihilation Limits for Antimatter on Scales Larger than the Solar System

If we wish to search for the existence and annihilation of significant amounts of matter on larger scales, say our galaxy, the best annihilation channel to use is still the gamma-ray one. (See the discussion in Steigman[4]. Our gamma-ray detectors still have much better detection and localization capabilities at $\sim$ 100 MeV than neutrino detectors do. The other decay products, electrons and positrons, typically do not propagate very far in the galaxy and are also confused with the signal produced from cosmic-ray interactions.)

Let's consider now what we can say about the presence of antimatter mixed in with matter inside our galaxy's gas. Following Steigman[4] again, the galaxy has been clearly detected in 100 MeV gamma-rays, and the observed fluxes are very close to those predicted from a galactic emission model where the emissivity arises from the interaction of cosmic rays with the gas, and, to zeroth order, simply scales with the density of the gas, e.g., [7]. The mean 100 MeV gamma-ray emission per hydrogen atom (which dominates the gas content) needed to explain the observed emission is $S_\gamma/n_H \approx 1.6 \times 10^{-25}$ s^{-1}. If we do not want to overproduce the galactic gamma-rays by having too much annihilation radiation, then we must require that the annihilation flux per hydrogen atom that manages to annihilate be less than this level. Let us assume that the number of antiprotons in galactic gas (as opposed to antiprotons inside dense, slowly annihilating objects like anti-stars) is a fraction f of the number of protons (hydrogen atoms) in the gas, and that both matter and antimatter are distributed roughly uniformly. Noting that we generate roughly $\sim$ 3 photons at $\sim$ 100 MeV energies per annihilation event, the limit on the annihilation emissivity of galactic gas can be written as

$$S_{Ann}/n_H = f < n_H \sigma_{Ann} v > \lesssim (\frac{1}{3})(S_\gamma/n_H), \tag{6}$$

or equivalently,

$$f \lesssim (\frac{1}{3})(S_\gamma/n_H) t_{Ann}, \tag{7}$$

where t_{Ann} is the annihilation time derived above. Substituting in $t_{Ann} \sim 100$ years, we arrive at $f \lesssim 10^{-15}$(!), which should not be a complete surprise since we already noted that primordial antimatter mixed with matter in interstellar gas would not survive very long (compared to cosmological timescales). If we compute the limit on f for annihilation in galactic halo gas, which has a significantly lower mean density, we still arrive at $f \sim 10^{-10}$ [4]. In other words, unless it is all locked up in objects that somehow do not annihilate, there is very little primordial antimatter in our galaxy. Regardless of whether antimatter has been locked up or not, the fact that it is not present in galactic gas alone tells us that our local universe is not symmetric with respect to antimatter and matter. While we may contrive to avoid the annihilation constraint and make the total number of baryons and antibaryons the same in our galaxy, there is no escaping the fact that baryons and antibaryons must have very different spatial distributions.

Even if we give up on the local universe being symmetric, it would still be useful to know how much antimatter is in objects that annihilate away inefficiently. (Note that in a universe with a baryon excess, such "hidden" or non-annihilating antimatter is exactly what we expect: the only survivors of annihilation would be the excess matter and whatever antimatter has been effectively separated from the matter.) If the antimatter ends up effectively in the form of dark matter such as black holes or anti-"quark balls," there is little that we can say except that the total amount of antimatter must be less than or equal to the amount of dark matter we infer in our galaxy, which unfortunately is larger than the amount of visible matter. (There are some constraints from the MACHO gravitational lensing project on the amount of dark matter that can be in discrete low mass objects, but the limit on the total mass in these objects is still larger than the amount of visible matter. [8].)

If the antimatter is instead in the form of anti-stars, what one usually thinks of when contemplating "mirror" universes, then it turns out that there *is* something one can say. To the extent that antimatter stars go supernova, for example, they should periodically spew their contents into the galaxy, and some antinuclei should therefore make their way into the cosmic ray population. Since cosmic rays can traverse much of the galaxy, we can therefore hope to find evidence of even distant antimatter stars by searching for the antinuclei. This is the domain of the "direct antimatter search" experiments, which we discuss in the next section. Turning to a possible gamma-ray signal, the amount of annihilation radiation expected from an antimatter star is indeed very low given the amount of antimatter it contains. For example, if one makes the standard assumption that a star's total annihilation luminosity must be less than its Eddington luminosity, the luminosity for which radiation pressure overcomes the star's gravity, an antimatter star will live billions of years, i.e., effectively forever. Our gamma-ray detectors are quite sensitive, however, and a even a relatively small amount of annihilation radiation may be detectable. Note that some annihilation should almost always occur because stars are not typically at rest with respect to galactic gas, i.e., we have a situation similar to that of anti-planets in the solar wind. Using the Bondi-Hoyle accretion formula to derive the flux of instellar matter gas hitting the surface of a star, we derive the following expected 100 MeV gamma-ray luminosity for an antimatter star[4]:

$$L_\gamma(\text{photons s}^{-1}) \approx \frac{1}{6} n \sigma_{Ann} v \sim 3 \times 10^{35} (M/M_\odot)^2 v_6^3, \tag{8}$$

where $M_\odot$ is the mass of the sun and $v_6 = v/10$ km s^{-1} is the mean stellar velocity relative to the gas. Now EGRET has detected some faint gamma-ray point sources that are still unidentified and seem to be associated with the galaxy. At a flux level $3 \sim 10^{-6}$ photons cm^{-2} s^{-1}, though, there are no obvious unidentified sources. We can therefore constrain the closest antimatter star to be further than ~ 30 pc (parsecs) from us, while there are many known matter stars within this distance, ~ 3000(!), e.g., [9]. Again, matter and antimatter are not distributed symmetrically ... We can even go a bit further than this. Since the overall 100 MeV luminosity of our galaxy in gamma-rays is $\approx 2 \times 10^{42}$ photons s^{-1}, the fraction of stars f_* that are gamma-ray emitting anti-stars must then be $f_* \lesssim 10^{-4}$[4]. Similarly, the fact that our companion galaxy M31 was *not* detected by EGRET at a level of 8×10^{-8} photons cm^{-2} s^{-1} [10] implies that f_* for M31 is also $\lesssim 10^{-4}$. Note that in general EGRET has detected *no* nearby galaxy, including the massive elliptical M87. (The only exception is the very nearby Large Magellanic Cloud, and that flux is consistent with what is expected from cosmic ray-gas interactions.)

There is one more important constraint on the existence mixed antimatter and matter that can be derived using gamma-rays. Galaxy clusters are filled with hot gas whose properties are now very well-measured with X-ray satellites. Following Steigman [4] again, the X-ray luminosity of a cluster (due to thermal bremsstrahlung emission) can be written as,

$$L_x(\text{erg s}^{-1}) \approx 1.4 \times 10^{-23} T_8^{1/2} \int n^2 dV, \tag{9}$$

where $T_8 = T/10^8$ K. If a fraction f of the X-ray emitting gas is antimatter that can annihilate, the predicted cluster gamma-ray luminosity would then be,

$$L_\gamma(\text{photons s}^{-1}) \approx 3 \times 10^{-14} f T_8^{-1/2} \int n^2 dV. \tag{10}$$

Note that the dependence on the gas density distribution is the *same* as in the X-ray luminosity calculation. This removes a huge uncertainty and allows one to place a robust limit on the cluster's antimatter content, ie.,

$$f \lesssim \frac{1}{2} T_8 \left(\frac{F_\gamma}{10^9 F_x}\right). \tag{11}$$

Now EGRET has convincingly detected *no* clusters down to much lower sensitivity levels than quoted in Steigman's review. Accordingly, we can update his limits on f for cluster gas to be $f_{clust} \lesssim 10^{-7} - 10^{-6}$.

In sum, Steigman's gamma-ray based arguments have only strengthened as our detector technology has improved. We can therefore repeat one of his main conclusions: "Clearly, the absence of strong extragalactic gamma-ray emitters requires that if present in the Universe at all, antimatter must remain separated from ordinary matter on scales at least as large as clusters of galaxies."[4]

2.4. Limits from Cosmic Ray "Direct Detection" Experiments

What if for some reason antimatter can't annihilate efficiently with matter because it is not significantly mixed with it, e.g., it is locked up antimatter stars or because it is all sitting in an antimatter galaxy separated from ours by a low density void? Is there any way to unambiguously test and directly test for the existence of this antimatter? If the antimatter were in the solar system, we could land a matter probe on it and see what happened, but unfortunately we don't know how to get out of the solar system yet. An alternative is to try to find something that can travel long distances to us. One useful probe of this type is provided by the energetic cosmic ray particles that rain down on us. For sufficiently high energies, these particles can traverse most of the galaxy (and even escape it) and the fact that they contain elements and isotopes with atomic number $Z > 2$ tells us that the process which accelerates them can sample material enriched by stellar processing. Now let us imagine that we had antimatter stars or an antimatter galaxy nearby. If we want to do as little violence as possible to our original idea of symmetry between matter and antimatter, presumably anti-stars evolve as ordinary stars do, with the more massive ones going supernova and accelerating anti-cosmic rays (in our current understanding of origin of cosmic rays). Thus, anti-stars can lead to the production of anti-cosmic rays and since anti-galaxies are collections of anti-stars, they are overall producers of anti-cosmic rays too. In other words, if a matter-dominated region of the Universe makes cosmic rays, it is not unreasonable that an antimatter-dominated region one makes anti-cosmic rays, and just as with matter cosmic rays, we can use the abundance patterns of the anti-cosmic rays to tell us about the state and origin of the gas they were accelerated from.

Motivated by such thinking, e.g., see [11] for a more recent example, experimenters began to try to measure the antimatter content of cosmic rays. Very early on it was realized, e.g., [12], that one had to be careful in attempting such a measurement because cosmic rays were sufficiently energetic to create antiprotons in their interactions with the interstellar medium. Indeed, this "secondary" flux of antiprotons has been observed, with about 10^{-4} antiprotons per proton above a few GeV in energy, see [13, 14] for recent measurements. Note that if the cosmic ray flux tracks the supernova rate of matter stars, then in a straightforward interpretation, this ratio tells us that the anti-star supernova rate inside the volume sampled by cosmic rays at these energies must be at least 10^{-4} times smaller, and consequently the number of anti-stars should also be $\lesssim 10^{-4}$ times smaller – comparable to what we obtained from some of the gamma-ray annihilations constraints. However, we are not so interested in yet another limit. What we want is a positive piece of evidence that will prove conclusively that at least one antimatter region or anti-star exists. People therefore switched to searches for anti-helium (both ^{3}He and ^{4}He) and heavier elements – see Bob Streitmatter's talk for a history of these efforts.

The reason for this is that the number of antinuclei secondaries produced in cosmic ray-gas interactions is expected to be very small. A relatively easy way to see this is to use Hagedorn's (pre-quark) theory [15] for what happens when a very energetic proton smashes into matter. In his theory, a quasi-thermal fireball is created akin to the Big Bang we discuss below. Continuing with our Big Bang analogy, the relative abundances of nuclei left after the fireball expands, i.e., the branching ratios of the various interaction products, are more or less those obtained using freeze-out abundances given by a magic and constant (for all species) freeze-out temperature $kT \sim m_\pi \sim 160$ MeV. (See below for a definition of "freeze-out" in the context of the Big Bang.) Anyways, the Hagedorn formula for relative branching ratios, which has been checked against experiment, says that

$$N_{\bar{X}}/N_{\bar{p}} \approx \exp[-2(M_X - M_p)c^2/kT_\pi], \tag{12}$$

where $kT_\pi \approx 160$ MeV and $N_{\bar{x}}/N_{\bar{p}}$ is the ratio of the number of X nuclei with mass M_X produced to the number of antiprotons that are also produced in the interaction. (If we increase the total energy of an interaction, we increase the total number of particles produced but do not change the branching ratios.) Taking X to be the lightest helium isotope ^{3}He with $M_X c^2 \approx 3M_p c^2 \approx 3$ GeV, we see that the number of ^{3}He nuclei produced in cosmic ray interactions interactions should be $\approx \exp(-25) \approx 1.4 \times 10^{-11}$ the number of antiprotons or $\approx 1.4 \times 10^{-15}$ times the number of primary protons in the cosmic ray flux, ie., a very small number! Even when this calculation is done more correctly, e.g., [16], one expects the abundance of ^{3}He to be $\lesssim 4 \times 10^{-13}$ that of the primary cosmic ray proton one.

The bottom line is that unless we have another way of making anti-helium, e.g., through "anti-nucleosynthesis" in antimatter stars, none of the past or currently planned antimatter detection experiments should ever see a single anti-helium nucleus. Indeed as summarized in Fig. 1, past experiments have not. Now helium and thus anti-helium can both presumably be synthesized in the Big Bang (see below) if we have separate regions of matter and antimatter. Therefore, the detection of even a single anti-helium nucleus, while very exciting, would not be nearly as

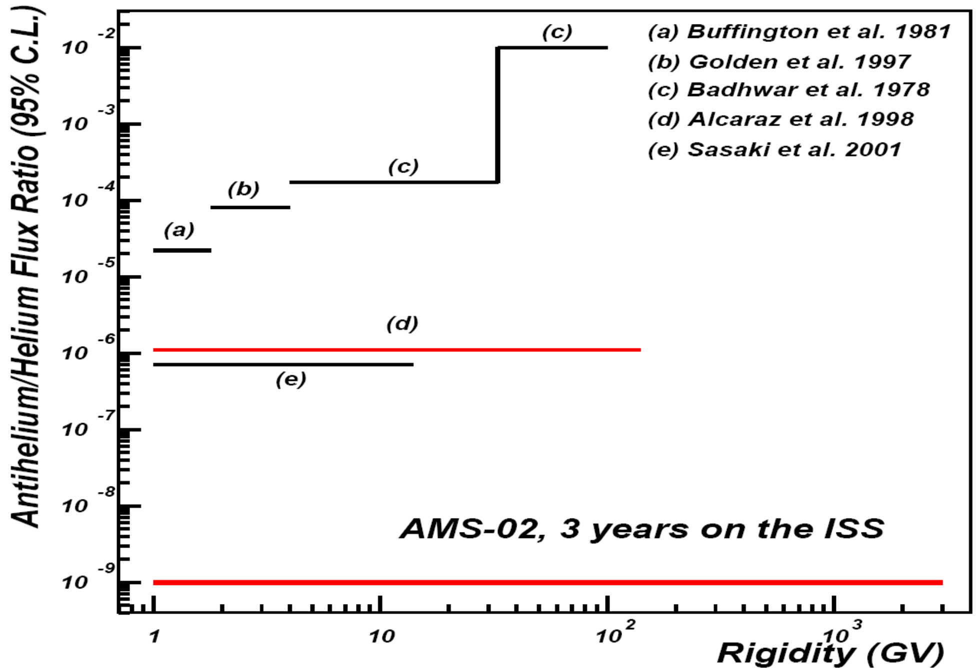

Figure 1: Taken from [17]. A summary of the various limits that have been placed on the anti-helium fraction. The expected sensitivity for the AMS-02 International Space Station experiment is shown as well.

exciting as the detection of single heavier ($Z < -2$) anti-nucleus, which in the standard picture can only be created in significant amounts by stellar processes. In other words, the detection of a single $Z < -2$ nucleus would imply the existence of anti-stars!

Especially if one wants to make use of upper limits, I note that a rather large uncertainty for the antinucleus search technique is still the amount of volume probed by the cosmic rays that are being examined. Clearly, the higher the rigidity (energy) of the cosmic rays that one can check, the longer a propagation distance they will have and the better the chances, say, of detecting a cosmic ray from another galaxy. The planned AMS (AntiMatter Search) experiment on the International Space Station (ISS)[18], also known as AMS-02, will be $\sim$ thirty times better than prior experiments in this regard, e.g., see Fig. 1, but the claim sometimes made that it will see cosmic rays from as far away as $\sim$ 100 Mpc is not completely obvious. For a discussion of the sampling volume uncertainties, and in particular whether cosmic rays can reach us from other galaxies, see [19]. Note that the lack of a gamma-ray detection of the Small Magellanic Cloud (SMC) by EGRET[20] implies a cosmic ray flux ~ 5 times lower than that in our galaxy at the location of the SMC (only ≈ 60 kpc away), so clearly cosmic rays are not escaping that easily from our galaxy, at least in the direction of the SMC. Still, AMS on ISS (AMS-02) will represent at least an order of magnitude improvement over prior experiments and the results are eagerly awaited. Even though the remainder of this lecture will be devoted to reasons why it should see nothing, one never knows...

2.5. Secondary Antimatter in our Galaxy: Another Man's Gold

Though we have been busily trying to rule out the existence of local antimatter, one must remember that it most definitely does exist in our galaxy, e.g., see [13, 14] for the latest antiproton spectra showing that antiproton cosmic rays constitute a fraction $\sim 10^{-4}$ of the proton cosmic flux. As discussed above, matter cosmic rays have enough energy to create secondary antiprotons in their collisions with ordinary matter, so this magnitude of antiproton signal is not unexpected, e.g., see [16] for a recent calculation. The fact that galactic antimatter exists and is actually not so rare is disappointing to some as it means we have yet another annoying background to deal with in our direct

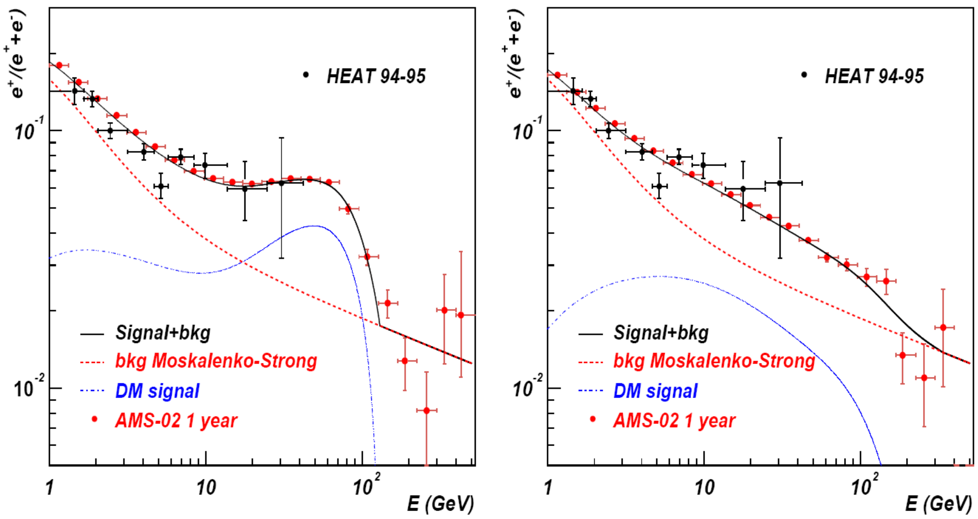

Figure 2: Taken from [17]. The dark crosses show the positron spectrum data points from the HEAT experiment. The lighter crosses show simulated AMS-02 spectra for a one year observation assuming a contribution from annihilating dark matter particles. The dark matter models used are similar to those in Fig. 1 of [21] except that the left panel assumes a supersymmetric relic particle with mass $m_x \approx 130$ GeV, while the right panel assumes a supersymmetric relic particle with mass $m_x \approx 336$ GeV.

detection experiments. One should not forget, however, that one man's background is another man's gold.

If galactic antimatter is not primordial (given how quickly it would have annihilated away), then it must be being created today. As evidenced by the trouble we have gone through to create antimatter on Earth, this is not an entirely trivial exercise and requires high energies and physical conditions that are not typical of what we see around us. Understanding the origin of the antimatter is therefore potentially rewarding. The most "boring" outcome of a antiproton study, for example, is that we are simply lead back to the ancient and still not satisfactorily solved problem of the creation and propagation of primary cosmic rays – but in the process we will likely have gained some important new constraints. A more "exciting" outcome, covered partly by Ted Baltz and Bob Streitmatter in their lectures, is that we instead stumble on important new physics. Rather than representing secondaries from cosmic ray interactions, antiprotons (especially the lower energy ones) could result from the evaporation of primordial black holes or, more in vogue these days, the annihilation of supersymmetric relic particles left over from the Big Bang. Such relic particles are currently considered prime dark matter candidates. For some of the more general recent papers on the topic of annihilation from relic supersymmetric particles, see [22–24].

Since this lecture deals with antimatter, for completeness I will briefly summarize some of the current issues and prospects concerning non-primordial antimatter in the galaxy. One of the earlier problems to surface was the detection of a significant antiproton flux at energies well below 1 GeV, where few antiprotons should have been created by cosmic ray interactions due to the kinematics of the interaction. This generated considerable excitement until it was realized, see [25], that several propagation processes such as ionization losses and non-annihilating scattering events on cold gas could push antiprotons to lower energies. (The observed solar modulation of the low energy antiproton flux was another clue that propagation effects were important.) The excitement for non-cosmic ray researchers has thus been tempered somewhat given all the messy propagation effects that must be dealt with, but [25] have recently suggested that one look instead for low energy antideuterons, which should be much harder to explain as cosmic ray secondaries. Since the production of antideuterons requires the fusion of an antiproton and an antineutron, antideuterons are in general rare. Especially if dark matter antiprotons and antineutrons are present, however, they should be detectable in sufficient numbers given an instrument with the sensitivity of AMS on ISS (AMS-02).

Another antimatter mystery arose with the claim by the HEAT experiment [26] that it had detected an excess flux

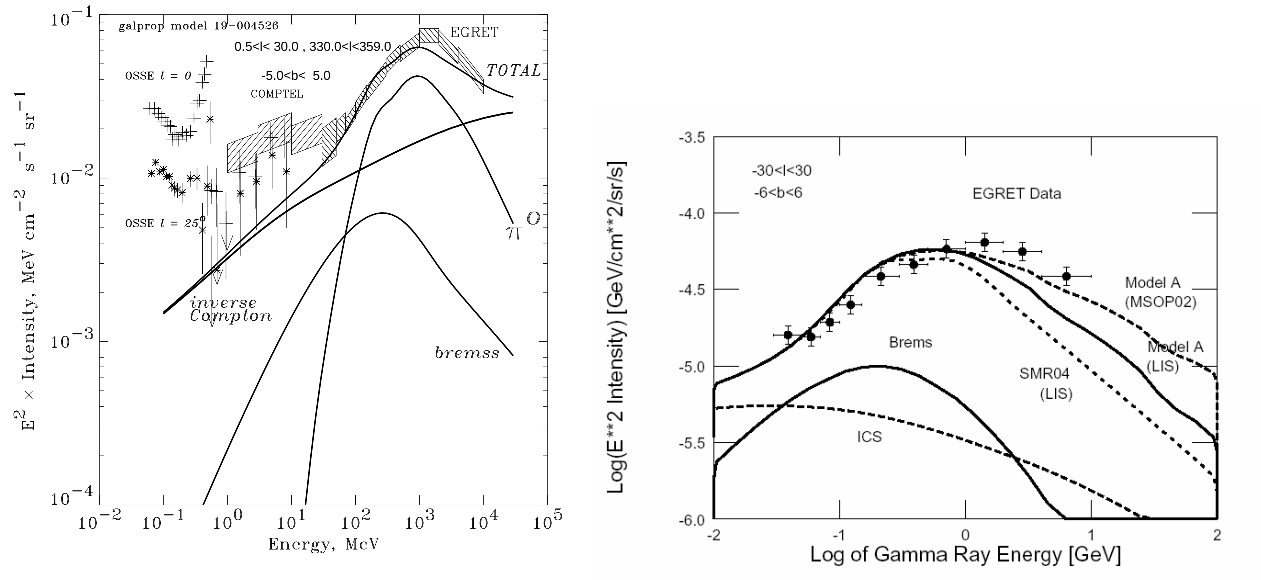

Figure 3: Left panel: Calculation from [28] showing one of their best fits to the observed spectrum of galactic gamma-ray emission, which they presume is due to secondaries from cosmic ray-gas interactions. Even after many parameters such as the energy spectrum of the electron secondaries are varied, the flux at $1-10$ GeV is not very well-matched. Right panel: A preliminary calculation from T. Kamae et al. (in preparation) showing the effects of using improved interaction physics. The curve labeled "Model A (MSOP02)," computed using a more realistic cosmic ray propagation model, new interaction cross-sections, and more standard parameter values, comes quite close to the "excess" GeV points.

of positrons above 10 GeV that was not explained by the standard galactic cosmic ray interaction models, e.g., [7]. Figure 2 shows the HEAT 94 positron spectrum (which was confirmed by the later HEAT 2000 measurements[27]) and the predicted AMS-02 spectra for two dark matter annihilation scenarios, similar to those explored in [21]. Those authors found that they could explain the shape of the excess but not its amplitude without invoking a somewhat *ad hoc* "boost factor" ~ 30 (that might reflect the clumping of the annihilating dark matter). Note that from Fig. 2, AMS-02 should have enough statistics to definitively settle what is going on – *if* we understand the cosmic ray background model. In general, the positron-related signals from annihilating dark matter have received quite a bit of attention in the literature, e.g., see [29–36].

Unfortunately, our poor understanding of the cosmic ray positron background may still be the main limitation in using the positron continuum to probe dark matter physics, and caution is in order. The left panel of Fig. 3 from [28] illustrates one of the latest models for the galactic gamma-ray background due to cosmic ray interactions, which despite many tweaks, still cannot satisfactorily explain the excess of 1-10 GeV gamma-ray flux (and the positron flux) that is seen compared to their models. This could indicate exotic new physics, e.g., decaying dark matter, or simply an improved understanding of old physics. The right panel of Fig. 3, for example, shows a preliminary calculation by T. Kamae et al. using more accurate cosmic ray interaction cross-sections. The GeV excess largely goes away when these cross-sections are combined with a more sophisticated cosmic ray propagation model. (See [42–44] for more discussion of the galactic gamma-ray modeling problem and its implications.) With its vastly improved map of the galactic gamma-ray emission, the upcoming GLAST gamma-ray mission should help significantly in constraining all the expected backgrounds from cosmic ray secondaries.

Some of the annihilation matter scenarios, e.g., [35] and in particular [37, 38], predict distinctive gamma-ray line signatures that avoid all the problems associated with the subtraction of a cosmic ray secondary background. Depending on the exact dark matter model parameters, i.e., if we are lucky, these could be detected by GLAST or the next generation ground-based Cherenkov gamma-ray like HESS and VERITAS, e.g., see [23]. Since the dark matter annihilation goes as the square of the dark matter particle density, any annihilation signal depends critically on the degree of dark matter clumping and concentration. Dark matter annihilation has often been postulated to occur in the galactic halo, but the center of our galaxy, which in optimistic scenarios has a dark matter "cusp" [39, 40], might be a better place to look carefully. Intriguingly, there is a strong electron-positron annihilation line seen from a resolved (few degree) region around the galactic center [41], and very close to the galactic center's black hole, a TeV (continuum) gamma-ray source has been detected (see the lecture by W. Hoffman). Given that the galactic center is a very active place, this emission probably has a non-exotic origin, but one can always speculate...

2.6. Faraday Rotation

As if local antimatter already did not have enough problems, Steigman[4] notes one further way to constrain its presence. Polarized light, e.g., from non-thermal synchrotron sources, that passes through gas with a non-zero magnetic field will have its polarization vector rotated by the process of Faraday rotation. The effect has been measured many times and the amount of rotation, usually expressed in terms of the so-called "rotation measure" is given by the line-of-sight integral $RM \sim \int [n(e^-) - n(e^+)] B_\| dl$. Note that regions dominated by antimatter (positrons) cause a rotation *opposite* to that caused by regions dominated by matter (electrons). The fact that we observe an effect at all means that on average, we can't have equal amounts of antimatter and matter along our various lines of sight. Note that the Faraday rotation effect has also been detected along lines of sight that pass through clusters, eg., [45] strengthening the case for matter-antimatter asymmetry on those scales too. For the case of galactic Faraday rotation we can actually go a bit further because pulsar timing studies, for example, provide us with accurate measures of another line-of-sight integral called the "dispersion measure," $DM \sim \int [n(e^-) + n(e^+)] dl$. If we did not have antimatter, then one could derive an average strength for the parallel component of the magnetic field, $< B_\| > \sim RM/DM$ – which usually agrees reasonably well with estimates derived in other ways. To the extent they agree, we must have $n(e^+) \sim 0$, i.e., the rotation measure we observe is not a delicate cancellation between two large and approximately equal electron and positron density fields. We can actually make a somewhat similar but weaker argument in clusters. Observations of the thermal X-ray emission and the Sunyaev-Zeldovich effect (see below) from clusters give us a measure of the electron density distribution inside the clusters. This can in turn be combined with the Faraday rotation measurement to derive a mean magnetic field (again assuming no antimatter), e.g., [46]. An independent constraint on the magnetic field strength inside the cluster can be obtained from limits on the hard X-ray flux that is Compton upscattered by synchrotron emitting electrons in the cluster, e.g., [47]. To the extent that the two estimates are consistent (which they roughly are), this again rules out a delicate cancellation effect in the rotation measure.

3. SOME KEY CONCEPTS IN COSMOLOGY AND BASIC BARYOGENESIS CONSTRAINTS

The preceeding section has discussed the experimental constraints on the presence of antimatter in the nearby Universe, in particular on size scales ranging from that of the solar system to those of galaxies and clusters. We will now try to improve on these arguments and say something about the presence of antimatter on larger scales and during earlier phases of the Universe's history. In the roughly thirty years since Steigman's review [4], we have made major theoretical and observational strides in our understanding of the overall cosmological parameters and evolution of our Universe. Experiments like WMAP [48], for example, now provide us with measurements of key quantities like the amount of baryonic matter in the Universe that are accurate to several digits. As we shall see, this new information tightens significantly the noose on the various proposals for "baryogenesis" which attempt to explain how and when the excess of matter over antimatter that we represent made its appearance in the Universe.

To understand why our existence is a major problem for those who study the Early Universe and to learn how to probe when the current baryon number was set, we first need to review some of the basic concepts and observational developments that lead to our current understanding of how the Universe evolved. Because such a review could easily fill several other lectures, I must limit my discussion here to only those details required to understand what lies behind the "cosmological" antimatter constraints and some of the current issues concerning them. For readers

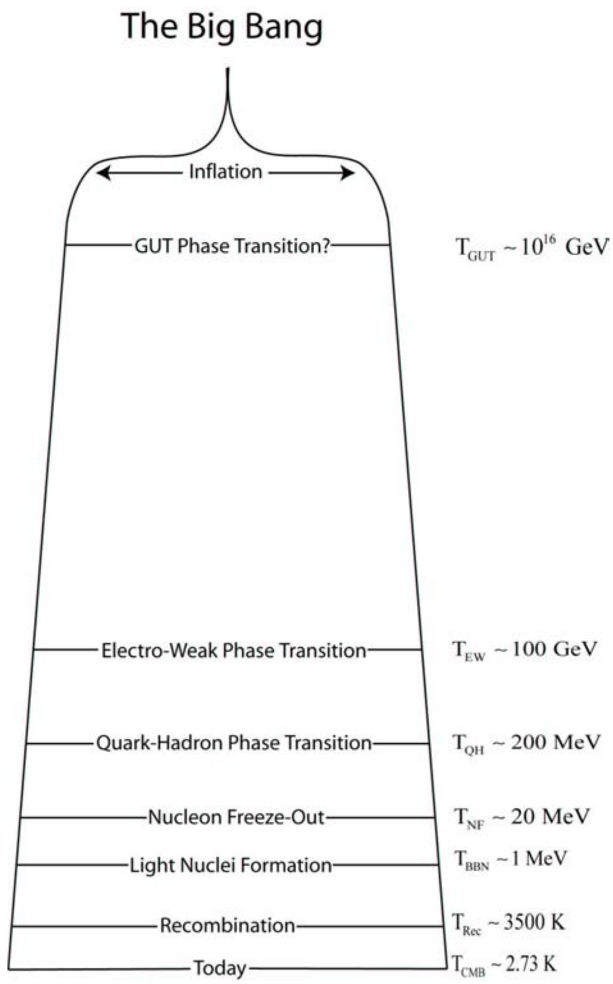

Figure 4: Adapted from [53]. Schematic depiction of how the Universe's radius evolves as time passes and the characteristic temperature of the Universe drops. The characteristic temperatures of some of the key events that occur along the way, e.g., transitions between different states of matter, are shown.

who need to know more, I will try to reference relevant background papers as well as encourage them to consult the several excellent texts written on the subject of particle astrophysics and cosmology, e.g., [49–52]. I will also cite below some review papers on the general particle physics aspects of the baryogenesis problem, but interested readers are reminded to first visit the lecture by M. Trodden.

3.1. The Big Bang

The first basic cosmology concept we must deal with is of course that of the Big Bang, in other words the idea that our Universe is not static in time but instead represents the explosive aftermath of some catastrophic event at early times. For those who did not trust theorists like Einstein (who predicted a dynamic universe but then canceled the dynamics out with the introduction of a "cosmological constant"), the original observational driver for this was the the observation by Hubble and many others that no matter which direction we look, galaxies appear to be receding from us with a speed proportional to their distance from us. This can be understood as the result of the uniform expansion of space-time (the usual analogy is made to the physical distance between spots on the surface of a balloon that is being blown up).

An obvious implication of this conclusion is that if we run the clock backwards, the galaxies that are now flying apart should all coming crashing together in a "Big Crunch." What might such a crunch look like? Well, if one took things to logical extremes, the Universe would be become very hot, with all atoms being ionized, and eventually matter and radiation coming into thermal equilibrium. Figure 4 schematically shows our current understanding of some of the key events we think the Universe has passed through. The various phases between those events reflect the physics we currently think should be dominant at the characteristic temperature of those phases. For example, given our current understanding of QCD and quark-gluon plasmas (now being tested in experiments like RHIC), we think that at temperatures $\gg 200$ MeV, individual hadrons (e.g., protons and neutrons) should not exist but rather that we should have a so-called quark-gluon soup. Only at temperatures below $\lesssim 200$ MeV is it energetically favorable for quarks to group together into hadrons. Similarly for temperatures $\lesssim 3500$ K, it is energetically favorable for electrons and nuclei to "recombine" and become bound in atoms (hence the name "recombination" for the transition). In our current cold Universe, therefore, most matter is neutral, and we see few free electrons. The event labeled "nucleon freeze-out" in Figure 4 is not so much a change in the dominant physics but rather marks the time when the expansion rate of the Universe is so fast that the thermalization processes tying neutrons and protons to the radiation field, for example, are no longer effective (see below), and nucleons consequently drop out of thermal equilibrium with the Universe. For a good overview lecture on the thermal history of the Universe, see [54]. Note that as we go back in time and reach higher temperatures and energies, our knowledge of the relevant physics becomes more and more uncertain as do the location of the events shown. For example, inflation may end without the Universe ever reaching T_{GUT}. (See §5.1 for an example of the problems that occur when the Universe gets too hot, e.g., the overproduction of gravitinos.)

In the discussion that follows, particularly on Big Bang Nucleosynthesis and the Cosmic Microwave Background, I will summarize some of the key evidence we have developed that this picture (Fig. 4) is indeed roughly correct. Since the methods used to collect this evidence tell us about conditions in the early Universe, and the presence of too much antimatter can change those conditions, they will prove useful in our quest to understand when and where antimatter may be present.

3.2. Big Bang Nucleosynthesis

One of the first major hints that the Universe has indeed gone through a very hot phase came from the development of our understanding of nucleosynthesis in stars and the measurement of the abundance of elements in stars and interstellar matter. Although it is not that common on Earth, helium (^{4}He) seemed to be everywhere else, with a very high mass fraction $Y = \rho_{He}/\rho_{tot} \approx 0.25$ (which corresponds to a helium to proton number density ratio of $n_{He}/n_p \sim 0.08$). Such a mass fraction was too high to explain by standard quasi-equilibrium, e.g., stellar, production scenarios, and this lead to the unorthodox but clever proposal of Big Bang Nucleosynthesis (often abbreviated as "BBN;" see [55] for a good introductory lecture on this topic, [56] for a key paper, and [57, 58] for some reviews). That point was that a very hot phase of the Universe would provide another environment besides stars where nucleons could overcome Coulomb barriers to fuse into heavy nuclei, but because the Universe was expanding rapidly it would *not* provide the quasi-equilibrium environment of a slowly evolving star.

As a first pass towards understanding BBN, let us imagine that the Universe were indeed very hot at some point, with a temperature higher than the typical $\sim$ MeV binding energies of nuclei. At this temperature, matter would have been a soup of unbound protons and neutrons and nuclei would not have existed. Now let the Universe expand and cool. For a time, as the Universe's temperature traversed the 100 keV - MeV range, it would be energetically favorable for protons and neutrons to fuse first into deterium and then into into heavier nuclei such as helium. The build-up of heavy nuclei would continue until the process was shut off by a combination of the Universe becoming too cold to overcome the Coulomb repulsion between nuclei and the density of nuclear matter becoming too low to allow a significant number of interactions. Be warned that I have glossed over many important details here such as the role of neutrinos and the neutron decay rate, but the key point, embodied in the so-called "Gamow Criterion", is that the number of heavy nuclei produced, in particular deuterium (the bottleneck element in the nucleosynthetic chain), scales with: (i) the total number density of baryons (ie., the number of fusion target particles), (ii) the relevant nuclear interaction cross-sections, and (iii) the characteristic expansion time of the Universe (which during the nucleosynthesis epoch is set mainly by the energy density of photons which in turn scales with photon temperature). Taking the number density of baryons today and knowing just how much deuterium needed to be produced to match observed abundance patterns, Gamow and collaborators [59–61] were then able to calibrate the expansion rate and photon temperature of the Universe at the time of primordial nucleosynthesis. This in turn enabled them to run the clock forwards to predict that a relic radiation field should be left over from that hot phase of the Universe and that it would have a temperature $\sim$ 5K, an interesting theoretical prediction that was mostly ignored − until that relic background was accidentally discovered by Penzias and Wilson in 1965 with a temperature of 2.7K (see the interpretation paper that immediately followed the discovery by [62]). The rest as they say was history...

The reason I have started off with a discussion of Big Bang Nucleosynthesis instead of the Cosmic Microwave Background (CMB), which is probably the most conclusive proof that something like the Big Bang happened, is that the BBN process enables a direct probe of the physical conditions relevant to matter and antimatter at an epoch corresponding to times only $\sim 1 - 100$ seconds after the Big Bang (redshift $\sim 10^9$ and temperatures $\sim$ MeV), an epoch that lies well beyond the so-called electron "last scattering surface" (photosphere) that produced the distortions in the CMB that we see today and beyond which we cannot see directly via photons. Now, with sufficiently good measurements and modeling of how perturbations in the Universe grow, we *can* use the CMB to learn about conditions in the Early Universe (and we shall make use of this below), but there is no way around the basic fact that the dominant distortions of the CMB reflect what happened when radiation came out of thermal equilibrium with matter and had its last interactions with it. (Once the Universe has cooled enough that electrons recombined with nuclei to form atoms, to zeroth order the background light propagates to us untouched.) This corresponds to temperatures $\sim 10eV$, redshifts ~ 1000, and times $\sim 3 \times 10^5$ years after the Big Bang, and a lot could have happened between the nucleosynthesis epoch and then. For example, one could imagine laying down small-scale spatial perturbations in the density of baryons that could have an important impact on the element abundances predicted by BBN yet would minimally impact the CMB because the perturbations would essentially be gone by the time of electron recombination, e.g., due to damping by particles such as photons with matter-interaction lengths larger than the perturbation scales or because the perturbations involved fluctuations in matter and anti-matter that annihilated away. (For reviews on the topic of spatially "inhomogeneous" BBN, which once garnered considerable attention, as well as other examples of non-standard BBN, see [63–65]. We will discuss this topic further in §4.2.)

3.3. Freeze-Out and Relic Particle Densities

A key concept that I have just skipped over in my discussion of BBN, and that we will need to understand, is that of "freeze-out." For example, one of the important quantities that enters into the BBN calculation is the relative abundance of neutrons versus protons. In principle, this is a quantity that could be fixed by hand as an initial condition, but part of the beauty of the Big Bang picture is that it is not supposed to be. Before the nucleosynthesis epoch (and after the preceeding era of the quark-gluon soup), neutrons and protons interconverted between themselves and maintained thermal equilibrium via weak interaction reactions such as $n + \nu_e \leftrightarrow p + e^-$. As a rule of thumb, thermalizing reactions such as these stop being effective at maintaining a thermal particle distribution when the mean time between interactions, $t_{int} \sim \Gamma^{-1} = (< \sigma_{int} v > n_{target})^{-1}$, becomes longer than the time it takes for the scale factor R of the Universe to double, $t_{expand} \sim R/\dot{R} = 1/H$, where H is the Hubble constant at the epoch we are considering. [The scale factor R tells us the factor we need to multiply our coordinate distances to obtain the physical distance between points. Going back to our Universe-balloon analogy, if we paint spots on a balloon that are initially separated by 1 cm, and then we blow up the balloon/increase its scale factor by a factor two, then spots will now be separated by 2 cm. In particle physics papers, this condition for ineffective thermalization is usually written as $\Gamma < H$.] The rationale for this is that the expansion of the Universe alone will cause particle densities to drop rapidly ($n_{target} \propto R^{-3}$ following the balloon analogy), which in turn will make all interaction times become correspondingly longer: if the interaction rate between two particle species is so slow that they barely manage to interact in a current doubling time, they certainly won't be able to interact in the next doubling time. In other words, if at a given time we find $t_{int} > t_{expand}$ for a particular particle species, then those particles drop or "freeze

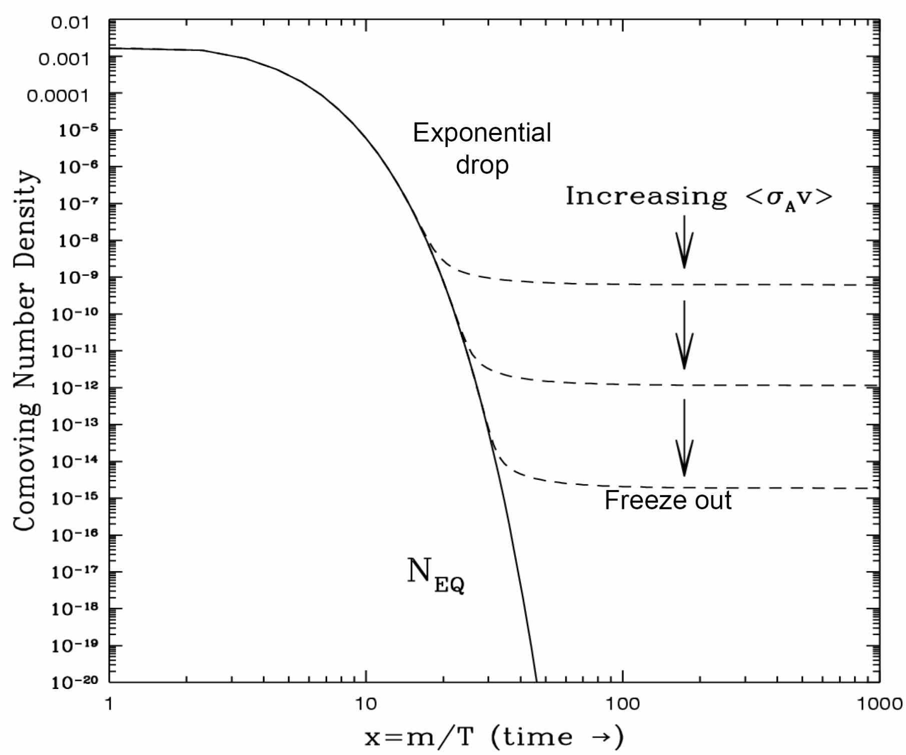

Figure 5: From [66]. The comoving number density of a particle species with a non-zero mass as a function of the Universe's temperature (time). As the temperature drops, the particles become non-relativistic and their equilibrium density, N_{EQ} drops $\sim$ exponentially with temperature. Eventually the rate of thermalizing interactions (characterized by the rate coefficient $< \sigma v >$) becomes so low that the Universe's expansion dominates their density evolution, i.e., their comoving density "freezes out" at a constant value.

out" of thermal equilibrium, and the subsequent evolution of their density is governed simply by the expansion of the Universe. The so-called "relic" density of those particles at a later time will then roughly be the density they had at the time of freeze-out, i.e., approximately the density expected from thermal equilibrium at that time, multiplied by $(R_{fo}/R)^3$ (to account for the Universe's subsequent expansion) where R_{fo} is the scale factor of the time of freeze-out and R is the scale factor of the time we are interested. What this means in practice for the particle density depends on whether the particles were relativistic or not when they froze out, i.e., whether $m_{part}c^2/k_BT_{fo} \ll 1$ or not where T_{fo} is the Universe's temperature when freeze-out occurs. If the particles are relativistic, their density at freeze-out is simply that expected for relativistic thermal matter, i.e., $n_{fo} \propto T_{fo}^3$. If they are non-relativistic, however, their density is given by the Maxwell-Boltzmann distribution, i.e., $n_{fo} \propto (m_{part}c^2/T_{fo})^{3/2} \exp(-m_{part}c^2/k_BT_{fo})$. The key point here is that the density of non-relativistic particles in thermal equilibrium is exponentially suppressed as the temperature drops, see Figure 5. If it were not for freeze-out and the departure from thermal equilibrium, there would be no massive particles left today!

The curves shown in Fig. 5 are for several different characteristic values of the rate coefficient of the interaction assumed to be responsible for maintaining particles in thermal equilibrium. The stronger the interaction (the larger its cross-section), the longer particles remain in thermal equilibrium, the lower their eventual freeze-out density, and thus the lower the density of relic particles we expect today. (See, e.g., [67] for a review on relic particle densities in various scenarios and how to calculate them. Note that since exponential factors are involved, obtaining an accurate estimate of the relic density requires accurately determining the freeze-out temperature. This requires going beyond the heuristic procedure just described and actually solving the relevant Boltzmann equation for the particle distribution in an expanding universe.)

What does this all mean for the protons and neutrons in BBN? Using (i) a total density of target nucleons (neutrons

+ protons) that is extrapolated back in time from what we observe today (see the next section for why we need to do this!) together with (ii) the relevant weak interaction rates (which took many years to measure accurately) plus (iii) a detailed model for the expansion rate of the Universe, one can show that for high temperatures ($\gg 1$ MeV) the weak interaction keeps the number of proton and neutron equal. As the Universe's temperature drops, however, neutrons begin to be suppressed relative to protons because of the neutron-proton mass difference of $\delta mc^2 \approx 1.4$ MeV, eventually reaching a relative freeze-out abundance of $n_n/n_p = \exp(-\delta mc^2/kT_{fo}) \approx 1/6$ where the freeze-out temperature turns out to be ≈ 800 keV, e.g., see [49].

Generalizing beyond BBN, a clever reader will immediately take advantage of this plot to propose a new particle with a mass and an accompanying interaction rate that will give a relic particle density today that is sufficient, for example, to match that inferred for dark matter. This was what was done in the case of the so-called WIMPs (Weakly Interacting Massive Particles, e.g., see [49]) which had masses ~ 100 GeV and thermalized via processes that had a cross-section characteristic of the weak interaction. Somewhat less speculatively, it was pointed out that if a supersymmetric extension to the standard model were correct, the lightest (stable) supersymmetric partner, often a "neutralino," might have the characteristics of a WIMP and hence could accumulate in detectable quantities in dense places like the center of the galaxy (as discussed in §2.5). An even more astute reader will note that we represent relic particles and will not be happy with our estimated relic particle density ... (see §3.6).

3.4. The Cosmic Microwave Background

Let us now turn to the other major probe we have available for baryons and physical conditions in the early Universe, the CMB, and consider some of its general implications. One of the most important features of the CMB that emerged soon after its discovery was its incredible uniformity after one subtracted off a dipole distortion, presumably due to our motion. This implied that the CMB was special and most likely not some local "fog." (The Universe immediately around us is not that uniform, and it would require considerable fine-tuning and coincidence to match the smoothness of the CMB.) For cosmologists, one of the most vexing aspects of this uniformity was that regions on opposite sides of the Universe that could have never been in causal contact in a standard Friedman-Robertson-Walker (FRW) cosmology turned out to have the same temperature to a very high degree of accuracy ($\sim$ a few parts in 10^5). The solution of this so-called "horizon problem" was one of the main motivators for the phenomenological theory of inflation, a piece of the emerging standard cosmological model that was absent, for example, when Steigman's review was written and that we shall see impacts the story of where today's baryons came from. (See M. Kamionkowski's lectures for a much more in-depth discussion of inflation.)

If one postulates an exponential phase of expansion ("inflation") for the Universe, e.g., due to the domination of its energy density by the (false) vacuum energy of some unknown (inflaton) field, then several nice things happen. First, if inflation continues for a sufficient number of e-foldings in the scale factor of the Universe ($> 60-70$), the observable Universe could have started out as a small patch that *was* in causal contact at very early times thereby explaining how regions that were not in causal contact today could have similar temperatures (and it turns out, similar fluctuation spectra). Second, nasty topological field theory defects like monopoles and excessively large density fluctuations would have been "inflated" away leaving a very smooth Universe. Third, unless ones lives in an exactly flat universe, the ratio of the present density of the Universe to the critical one that would close it ($\Omega = \rho/\rho_c$) is a strong function of time, and it was thus puzzling that we just happened to live in era when $\Omega \sim 0.1 - 1$ (the range of values discussed in the pre-WMAP days). Rapid inflation, however, meant that our space time became asymptotically flat, singling out a particular value, $\Omega \approx 1$. Whether or not this matched the actual observed value was another story, but at least there was a clear prediction.

While it solves several problems, the exponential expansion of inflation creates another. At the end of the expansion, the Universe is far too smooth and cold to produce the inhomogeneous structure we see around us today. The second part of the inflation story therefore postulates that the inflaton field decayed to a lower energy state, thereby "reheating" the Universe. (These days some also discuss a "pre-heating" phase involving the resonant production of particles, e.g., [68].) Although there is great argument over the details (e.g., no one knows what the inflaton field is or what its potential energy surface looks like) there are three generic observational predictions from the second part of the story. First, the release of energy by the inflaton decay is not uniform in space but reflects the quantum fluctuations developed by the inflaton field. A detailed examination of the density perturbations resulting from these fluctuations, e.g., see [49–52], shows that the perturbations should therefore: (i) be nearly scale invariant (i.e., have a "Harrison-Zeldovich" power spectrum with index $n \approx 1$ implying roughly equal power on small and large spatial scales), and (ii), reflect a Gaussian amplitude distribution with no spatial correlations (i.e., consistent with Gaussian random noise and not what is expected from alternative models where, for example, topological defects seed structure). Next, the inflaton decay produces fluctuations in energy (perturbations in the space-time curvature) and the distribution of particle species eventually produced in the decay of one of these fluctuations, e.g., the ratio of photons to baryons, should just reflect thermodynamics and particle physics. In particular, if more energy is released somewhere, there should generically be more of every type of particle, i.e., perturbations in different particle species should be correlated and regions with more photons should have correspondingly more baryons. The technical name for these kind of perturbations is "adiabatic" perturbations since such perturbations produce no net change in entropy. (We will revisit the slighty complicated topic of types of perturbations in §4.1 and §5.3.) In sum, then, the perturbations generated by inflation should be "adiabatic, scale free, and Gaussian random." Anything that messes up our basic inflationary scenario, e.g., the presence of an extra scalar field besides the inflaton as speculated in certain baryogenesis models (§5.3), will in general lead to perturbations that do *not* have all three of these characteristics.

Besides general curiosity about the characteristics of our Universe (e.g., its size and whether it is open or closed), predictions such as these strongly motivated the tremendous amount of work that has gone into measuring the space-time geometry of our Universe and the characteristics of the inhomogeneities that we find in it. Before the CMB anisotropy experiments, basically the only tools we had available (besides BBN) were observations of the relatively nearby visible universe using light. A large amount of effort therefore went into studying the clustering of galaxies and stars, trying to find standard candles and standard meter sticks that they could seen at large distance in order to probe the geometry and expansion rate of space-time, and simply trying to up add up all the matter that we could find to see how the implied matter density of the Universe compared to the critical one (e.g., see [50] for an extensive review and discussion). Given, for example, that most of the matter in the Universe turns out not to be visible and that we still don't exactly know how to turn a lump of gas into stars, it should not be surprising that drawing conclusions from these efforts has been difficult and fraught with error.

The detection of the CMB offered an elegant way around many of these problems. If our theory for how the structure we see today formed was correct, i.e., if it is the result of the gravitational collapse of primordial density perturbations, then those perturbations existed when the CMB last interacted with matter. Because at that point the CMB radiation was no longer in thermal equilibrium with matter, we ought to be able to detect the impact of those perturbations through non-thermal deviations in the CMB spectrum (caused, for example, by the gravitational Sachs-Wolfe effect due to photons climbing in and out of the time-changing gravitational potential wells of the perturbations). Given how small those deviations were likely to be (in fact they turned out to be ~ 100 times smaller than initially expected), trying to study them in detail might not have seemed such a smart thing to do. However, there was one very big payoff. In the theory of gravitational structure formation, what we see today is a highly non-linear and messy function of the initial perturbations, and unfortunately we know that physicists are only really good at linear perturbation theory. When the perturbations made their imprint on the CMB (at $z \sim 1000$), however, the Universe was in a much simpler state as not enough time had passed since the end of inflation for the density perturbations to have become large, i.e., non-linear ($\delta\rho/\rho \sim 1$). At that time, there were no collapsed objects such as stars and galaxies and all the headaches associated with them. The study of CMB distortions therefore provides us a much cleaner snapshot of the initial density fluctuations and makes it relatively straightforward to test, for example, whether they are consistent with Gaussian random noise or not. Also because we are in the linear regime, it is possible to predict quite accurately (after several graduate student theses of work) how initial perturbations grow (be they adiabatic or not) and map onto the CMB distortions. In particular, because in the Big Bang-inflation scenario all perturbations start their growth at a very well-defined time and from a well-defined mean background field, one can show that the angular power spectra of the CMB distortions should have distinctive features called the

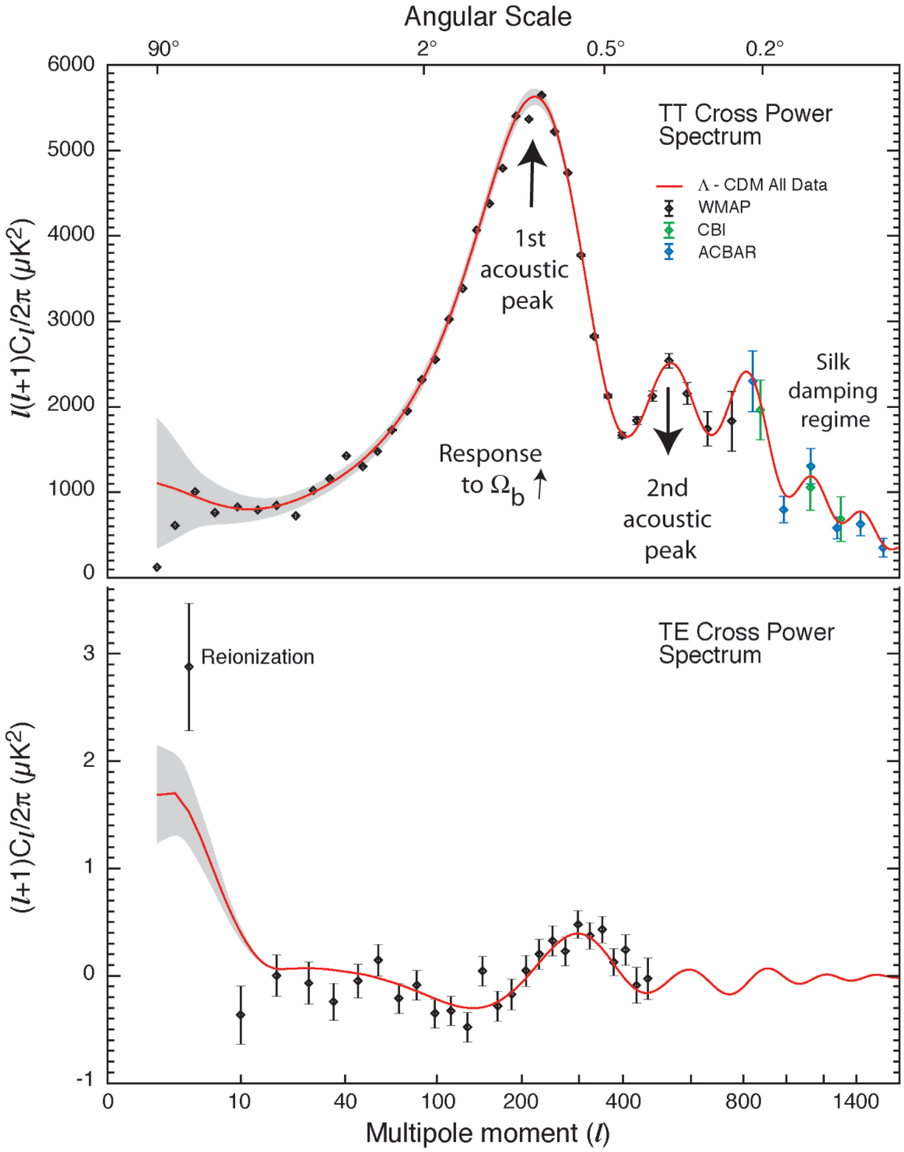

Figure 6: Angular power spectrm of CMB temperature fluctuations (top panel) and angular cross-correlation spectrum of CMB temperature-E mode polarization (TE) fluctuations (lower panel) obtained using the first year of data from the WMAP mission as well as data from ground-based experiments (for small scales, $l \gtrsim 600$). The arrows indicate the change in amplitude of the first and second acoustic oscillation peaks in response to an increase in the baryon density, Ω_b. On angular scales $l \gtrsim 800$, adiabatic baryonic fluctuations are damped by photon diffusion ("Silk damping") near the epoch of recombination. Adapted from figure by NASA/WMAP team.

"acoustic oscillation peaks" – see Fig. 6 which shows the latest observational data on the CMB fluctuations.

For the purposes of this lecture, all one needs to know about the squiggles and peaks in Fig. 6 is (i) we can learn about the space-time geometry of the Universe by using the characteristic physical wavelength of acoustic oscillations at recombination as a standard meter stick whose apparent angular size on the sky is given to us by the location of the oscillation peaks in the CMB power spectrum (the l^{th} harmonic on the x-axis corresponds to an angular scale of π/l radians, or $\sim 1°$ for the first peak), and (ii) the amplitudes of the peaks probe the physics of how the perturbations propagate and couple to the CMB. (For those who must know more about the squiggles, good places to start are the reviews of [69] and [70].) Of particular interest to our topic, changing the abundance of baryons in the Universe changes the relative amplitude of the first to the second peak in the CMB power spectrum. The CMB fluctuations can therefore be used as a "baryometer."

The CMB is not just useful as a probe of the perturbations that led to the structure we saw today. Well before the tiny distortions we have talked about could be detected, people measured the average spectrum of the CMB and found that it was a Planckian (i.e., a thermal blackbody) to a very high degree of accuracy (see [71] for a discussion of the most recent results provided by the FIRAS instrument on the COBE satellite). This is of course expected in the Big Bang picture we have been discussing since photons are massless, and at freeze-out they consequently still have a relativistic thermal particle distribution. One can then show that if the Universe expands adiabatically and nothing further happens to the photons after freeze-out, the form of their thermal distribution (a Planckian) is maintained except that one must replace the original photon temperature by an appropriately "redshifted" one, $T \propto R^{-1}$, where R is the Universe's scale factor. Reading through again the reasoning for obtaining a Planckian, we see that one of the key requirements is that "nothing happened" since freeze-out. The fact that we observe a Planckian therefore places a powerful constraint on the non-standard things that might have happened.

To better understand the nature of this constraint, let us first consider the coupling of photons to other particles to see exactly when and how they "freeze-out." If we want to exchange energy with photons and thermalize them, the particles of interest must of course couple electromagnetically. Of these particles, the ones that will determine photon freeze-out are the ones that couple most strongly to photons and that maintain their density for as long as possible while the temperature of the Universe drops, i.e., from the general freeze-out discussion above, we want the particles with the lowest mass. Scanning through the possibilities, we find that electrons and positrons are the particles of interest. These undergo a dramatic drop in density due to pair annihilation once the temperature of the Universe drops below $kT \sim m_e c^2 \sim 1$ MeV, and pair production is no longer energetically possible. Interestingly, as we shall see shortly, baryons do not reach their freeze-out density and presumably because we must maintain a charge-neutral universe, neither will electrons. From the neutrality condition, the number density of electrons surviving annihilation should be the same as baryons, $\sim 10^{-9}$ the density of photons. This number is small but until a redshift $z \sim 10^7$ (temperature $T \sim 10^7$ K), the electron density is high enough that Compton scattering between electrons and photons keeps their energies tightly coupled and bremsstrahlung emission and absorption allow the photon number to adjust to maintain a thermal (Planckian) distribution for the photons [72]. Thus, even though a large amount of energy was dumped into photons during the era of electron-positron annihilation, the energy is thermalized and the net effect is simply to boost the temperature of the Planckian photon distribution. (This explains, for example, why the relic neutrino background, which does not see any of the electron-positron energy, is expected to be colder than the CMB. Note that the thermalization of the annihilation radiation would have been fairly effective even if electrons had reached their freeze-out density [72].) Any other energy dumped into the Universe during that time would suffer a similar fate as long it coupled electromagnetically, i.e., it would simply boost the temperature of the CMB but not distort the Planckian shape of its spectrum.

Below redshift $\sim 10^7$, however, the electron density has dropped sufficiently due to expansion that bremsstrahlung (a higher order process) no longer works effectively. This means that Compton scattering is the only process left to redistribute energy between photons and electrons (and effectively, other photons). Unfortunately, Compton scattering conserves photon number, and this in turn means that if we add energy to the Universe, e.g., in the form of hot electrons, it is impossible to maintain a thermal photon distribution since for a Planckian, the photon number goes as T^3 and should increase with increasing energy density (but can't!). For redshifts above $\sim 10^5$, the electron density is high enough that a photon will Compton scatter many times off the background electrons before reaching us. In this limit, the photon energy distribution that results is best described as a Bose-Einstein distribution with a non-zero chemical potential μ which scales linearly with the amount of energy added (see [71, 72]). For lower redshifts (and especially after electrons recombine into atoms), the probability for a photon to scatter on its way to us is small, the mean energy of background electrons and photons can be very different, and the CMB spectral distortion that results is then best described by the "Sunyaev-Zeldovich effect" [73], i.e., an overall shift of the energy of CMB photons by a factor e^{4y} where $y = \int [k(T_e - T_\gamma)/m_e c^2] d\tau_e$. (Here, $\int d\tau_e$ represents a line integral over electron optical depth, i.e., $\int \sigma_T n_e(z) dl$ including appropriate cosmological factors.) In reading the literature, please note that the factor of 4 sometimes appears in the definition of y, and that y is often given in the limit $T_e \gg T_\gamma$ as would

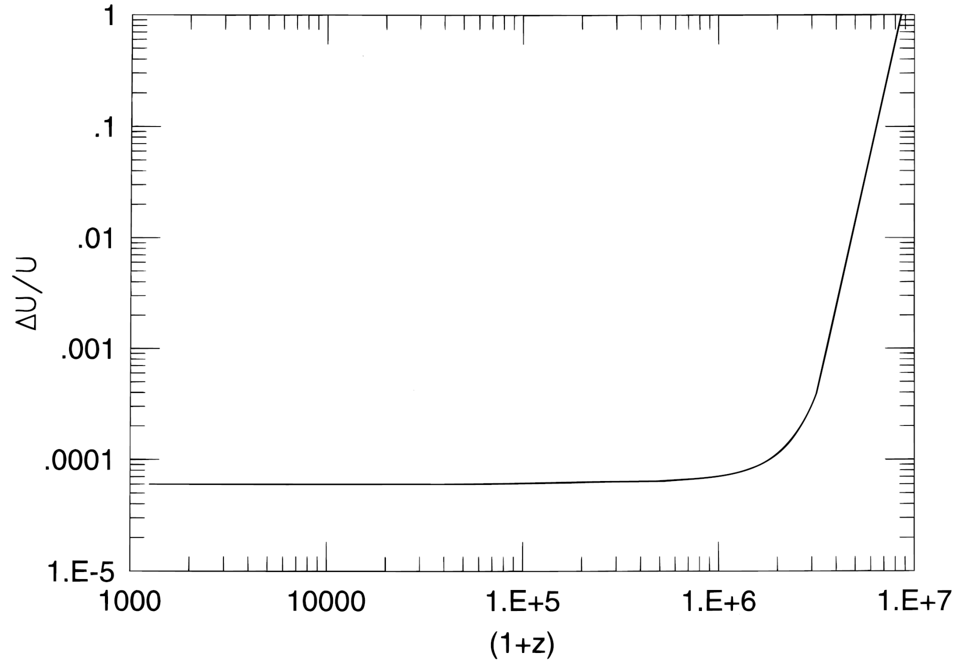

Figure 7: From [71]. *Relative* limit ($\Delta U/U$) on the amount of energy ΔU that can be added to matter that couples with the CMB as a function of redshift. Exceeding this limit causes spectral distortions of the CMB larger than the limits obtained by the FIRAS instrument on COBE. Here U is the energy of the CMB, which decreases with temperature as T^{-4}. The *absolute* limit on ΔU therefore also decreases strongly with redshift.

apply, for example when dealing with hot electrons in a foreground galaxy cluster. In physical terms, the so-called Compton y parameter is the average number of scatterings a photon undergoes as it propagates towards us ($\propto \tau_e$ for $\tau_e \ll 1$) times the mean fractional photon energy change per scattering ($\Delta\epsilon/\epsilon \approx 4k_BT_e/m_ec^2$ for $T_e \ll m_ec^2$ and $T_e \gg T_r$.) Note that we do not require exotic energy release or physics for meaningful spectral distortions to occur. The simple non-linear growth of the Universe's initial density perturbations, e.g., to form a cluster, leads to detectable distortions as CMB photons pass through the hot cluster gas.

Because looking for Sunyaev-Zeldovich distortions may prove one of the best ways to find galaxy clusters (highly useful for cosmology) and because we would be very interested if we found evidence for non-Gaussian CMB fluctuations (that were not due to foreground structure like clusters), one can be sure that the CMB will be scrutinized very carefully in upcoming years. In particular, if we have chunks of the Universe annihilating away, we might hope to spot them by their effect on the CMB, not just by their gamma-ray emission. (See [74] for a schematic calculation of the distortions caused by annihilation at the boundaries between large antimatter and matter domains. Their predicted effect is small, though, and could only be seen by a next-generation CMB experiment like Planck.) Needless to say, no unexpected CMB features have been found so that for now we must content ourselves with global limits on the amount of excess energy released into the Universe since, roughly, the electron-positron annihilation epoch. The current limits on this are summarized in Fig.7 from [71].

Here I have focused on distortions of the CMB spectra, but I note that in future years will be able to use the CMB as a very useful probe in another way, namely by mapping out its polarization structure (e.g., see section XXXX below on delayed re-ionization). The WMAP mission already made a first attempt at this, and polarization studies are one of the main goals of the upcoming Planck mission and several proposed ground-based experiments.

3.5. Cascading

Before returning to the problem of baryons in the Universe, we should add one more concept to our toolkit. Some of the forms of non-standard energy release in the Universe that particle physicists talk about today involve the decay of massive GUT scale ($\sim 10^{15-16}$ GeV) particles, whose decay products have energies far above the mean temperature when the decay occurs. Another very relevant example for us, where we have particles with energies far above the background temperature, is the possible annihilation of matter and antimatter after baryon freeze-out, e.g., during the BBN era. In this case, the typical annihilation products are particles with energies ~ 100 MeV while the background temperature is less than ~ 1 MeV. What happens to the energy carried by these "supra-thermal" particles, e.g., whether it is eventually thermalized and joins the background, depends exactly on when it is released, in what form, and what energy loss/absorption processes are effective at that time.

Generically, however, one finds that any high energy electrons or positrons will cool very quickly on background radiation via Compton scattering. Note that the characteristic cooling time for an electron interacting with the CMB in the Thomson-limit is,

$$t_C \sim \gamma/\dot{\gamma} \sim \Big[\frac{4}{3}\gamma c\sigma_T\big(\frac{U_{rad}}{m_e c^2}\big)\Big]^{-1} \approx 3.9\times 10^{17}(E/100\ \mathrm{MeV})^{-1}(T/2.73)^{-4}\ \mathrm{s}, \tag{13}$$

where U_{rad} is the CMB energy density. This is to be compared with the Universe's current expansion time of $t_{exp} \sim 4.2\times 10^{17}$ s. Since the expansion time decreases as a much slower power of temperature than t_C (e.g., $t_{exp} \propto T^{-2}$ in the radiation-dominated era), the cooling time for even 100 MeV electrons quickly becomes much shorter than the expansion time as we go back to earlier epochs . Note that at very high electron energies where the Klein-Nishina effect reduces the effectiveness of Compton scattering, higher order processes such as triplet pair production, $e\gamma \rightarrow e(e^+e^-)$, kick in and keep energy loss times short. The net result of this rapid Compton cooling is the production of a Compton-upscattered radiation spectrum that may contain non-trivial amounts of energy at photon energies *much lower* than the original electron energy.

Now what happens to very high energy photons, produced either directly by decay or annihilation or indirectly by Compton upscattering? Photons with energies above $E_{\gamma\gamma} \sim 100$ TeV $(T/2.73K)^{-1}$ have enough energy to pair produce on the thermal background radiation and are quickly transformed into electron and positrons. A 100 TeV photon today, for example, can only propagate ~ 10 kpc before pair producing. The new electrons and positrons that are created in turn cool by Compton scattering, and the net result is an electromagnetic pair cascade (not unlike what happens when a high energy particle hits a calorimeter). This cascade will degrade the energy of the few primary particles, transforming into the form of many secondary photons at energies less than $E_{\gamma\gamma}$, and even lower if secondary background radiation fields, e.g., from stars, are present (as they are today).

This production of lower energy photons from initially very high energy particles is sometimes forgotten but can have quite important implications, as we shall see. For example, if today some GUT-scale particles are still present and decaying, we can put strong constraints on their abundance by using detectors operating at much lower energies, e.g., GLAST, to detect their cascade radiation, e.g., [75–77]. At earlier times, even if the cascade radiation is eventually thermalized and we can't observe it, the low energy photons produced by Compton cooling and pair cascading can be much more damaging to their surroundings than the original high energy particles. As an example, the atomic photo-ionization cross-section drops with photon energy as ϵ^{-3} once the photo-ionization threshold is crossed, so one might not be too concerned by the presence of a few extra 100 MeV photons at the epoch of recombination. Once cascading is taken into account, though, the effects of those photons may be much more dramatic, e.g., delaying the epoch of reionization, e.g., [78] (§4.3). Similarly, in the BBN era, cascading can make energy quickly appear at the $\sim$ MeV energies needed to efficiently photo-dissociate nuclei – which leads to very important constraints on GUT scale particles that survive until that epoch (§5.1).

3.6. Why Baryons Are a Problem: Basic Constraints on Baryogenesis Scenarios

The combination of information from the CMB distortions and information from other cosmological probes such as large-scale galaxy clustering (provided by massive surveys like the Sloan Digital Sky Survey) has turned out to be

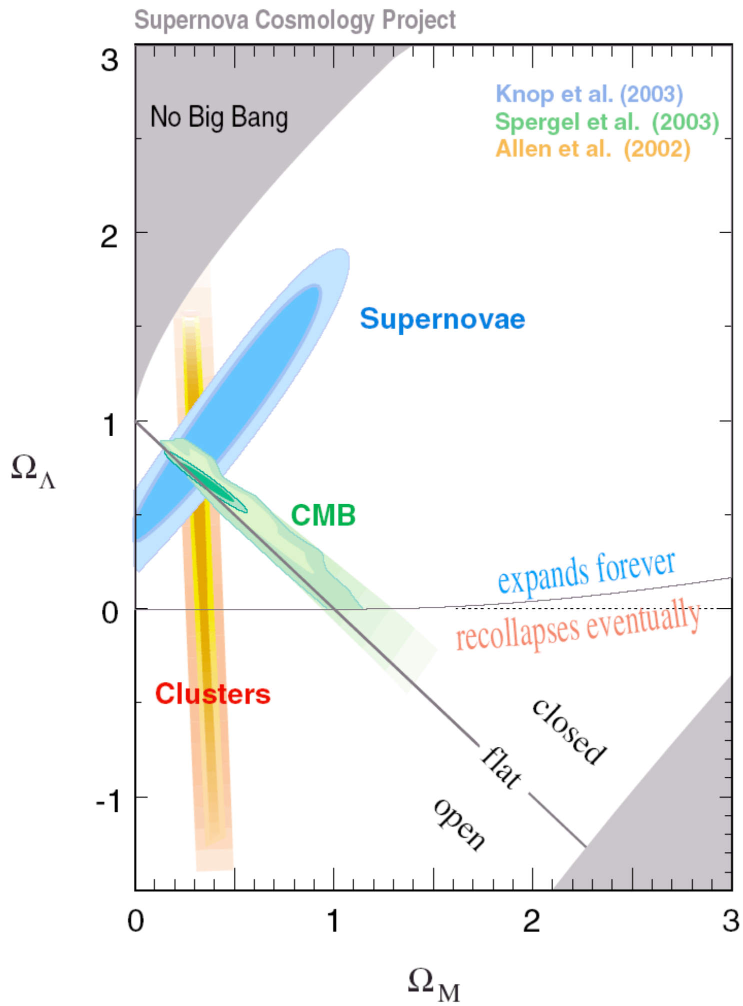

Figure 8: From [79]. The new "concordance" cosmology: the experimental constraints from the CMB measurements (WMAP), studies of clusters, and Type Ia supernova searchs are plotted on the plane Ω_Λ vs. Ω_M where Ω_Λ is the energy density of the cosmological constant ("dark energy") and Ω_M is the total energy density of matter (dominated by dark matter and not baryons).

spectacularly successful. After years of fierce battles and confusion, we finally seem to be converging on a precision "concordance" cosmology that has $\Omega_{tot} = 1$, including a dominant contribution $\Omega_\Lambda \approx 0.7$ from a mysterious "dark energy" and insignificant contribution, $\Omega_{baryon,CMB} \approx 0.04$, from us baryons [48]. (See Fig. 8 for a summary of the different observational constraints. Please note that for the remainder of the lecture we ignore dark energy as it only affects the Universe's dynamics at late times and thus does not impact our main conclusions.) Perhaps more interestingly, all the generic predictions from inflation appear to have been borne out [80], confirming the basic soundness of our overall picture. We thus are entering an era not unlike that of the standard model in particle physics where we can explain most experimental data very well. This is all excellent – until we realize that neither the standard particle physics model nor the new concordance cosmology are consistent with the abundance of matter we see today, let alone the absence of antimatter...!

To see why this is so, let us follow the standard scenario and assume that at earlier times baryons and anti-baryons are no different from any other particle and consquently are subject to the thermalization and freeze-out arguments above. In particular, we will initially have have equal amounts of baryons and anti-baryons that are homogeneously mixed as they condense out of the quark-hadron soup. Later, as the Universe expands and cools below ~ 1 GeV, the baryons and anti-baryons become non-relativistic and their relative thermal production and annihilation rates lead to an exponential suppression of their density until they hit freeze-out and decouple from the thermal background. What is their relic freeze-out density? Looking through Particle Data Group summaries or using an order magnitude estimate for the proton-antiproton interaction cross-section $\propto m_\pi^{-2}$, one finds that at GeV energies, the relevant $< \sigma_{int} v >\sim 1.5 \times 10^{-15}$ cm^{3}s^{-1}. Working through the detailed freeze-out calculation, e.g., [4, 51, 67], one arrives at a freeze-out temperature of $T_{fo} \approx 22$ MeV, and thus that the expected relic baryon to photon density ratio is $n_b/n_\gamma = n_{\bar{b}}/n_\gamma \sim 10^{-18}$!

This is a very small number, but is it in fact a problem? To see if it is, we need to experimentally determine the density of baryons and anti-baryons today, a task that is not as easy it sounds. Specifically, we are after the value of one of the old Holy Grails of observational cosmology, the ratio,

$$\eta = n_B/n_\gamma = (n_b - n - \bar{b})/n_\gamma, \tag{14}$$

where n_b and $n_{\bar{b}}$ are the baryon and anti-baryon densities respectively, n_B is the net baryon number density, and dividing by n_γ, the CMB photon density, removes the temperature (redshift) dependence of n_b and $n_{\bar{b}}$ after freeze-out. (Why people look for the difference between baryon and anti-baryon number should become clear shortly. η is a is a measure of the present baryonic charge, B, of the Universe) In the preceeding paragraph I cheated and gave you the answer. To make the baryon problem more robust (in case one does not believe in the magic of CMB observations) and to better appreciate the historical difficulties in measuring η as well as develop a feel for the numbers, let us assume we only know about what we can see with our eyes in the local Universe, and that for now we will not distinguish between matter and antimatter. Since we clearly live in a rather biased location in terms of matter density, we need to sample densities in a much larger volume of space. One of the main ways to do this has been through optical galaxy surveys that add up all the starlight in a set volume, thereby obtaining the mean stellar luminosity density of the Universe. From blue light (B band) studies, for example, the observed luminosity density is $< L_B >\approx 1.9 \times 10^8 L_\odot h$ Mpc^{-3} [81], where $L_\odot = 4 \times 10^{33}$ ergs^{-1} is the luminosity of the sun and h is today's Hubble constant ($H_0 = \dot{R}/R$) normalized by the value 100 km s^{-1} Mpc^{-1}. To convert this to a mass density, let us assume that all matter is in stars, that all stars have the mass to luminosity ratio of the sun (2×10^{33}g$/L_\odot$), and then to correct for the fact that the first two assumptions are not necessarily true, introduce an infamous fudge factor called the "mass-to-light ratio," (M/L). We thus arrive at the crude matter density estimate, $\rho_m \approx 1.4 \times 10^{-32} h(M/L)$ g cm^{-3}, where (M/L) is in solar units. Given that a lot of matter could be locked up in interstellar or intergalactic gas or borderline stars called "brown dwarfs" or faint galaxies that are all hard to detect, plus the confusion introduced by the clear presence of dark matter in galaxies, it should not be surprising that determining (M/L) has been one of the more vexing problems in observational cosmology. For now, let's be conservative and use $(M/L) \approx 2$ based on samples of local stars or stellar evolution/population theory (i.e., we are implicitly assuming that all mass is in the standard stars we know about). Using the critical density, $\rho_c = 3H_0^2/8\pi G \approx 2 \times 10^{-29} h^2$ g cm^{-3}, we then arrive at the following estimate for the contribution of starlight-related matter to the Universe's closure density: $\Omega_{b,\star} = \rho_\star/\rho_c \approx 0.0014 h^{-1}$. This value falls far short of the $\Omega_{tot} = 1$ expected from inflation, and using the best fit value of $h \approx 0.71$, falls a factor n ≈ 20(!) below the current best WMAP-era measurement for Ω_b – a clear indication that baryons are not distributed in a completely straightforward manner. Nevertheless, let us proceed and use this conservatively small number to estimate n_b/n_γ ($\sim n_{\bar{b}}/n_\gamma$?). Using the fact that the CMB density today is 412 photons cm^{-3}, we finally arrive at $n_{b,\star}/n_\gamma \approx 1.4 \times 10^{-11}$. Note that the best WMAP-era value is $n_{b,tot}/n_\gamma \approx 6 \times 10^{-10}$, a value we shall use below. Oops... The standard Big Bang prediction is too small by a factor of 10^7!

The problem is actually much worse than this since I ignored the distinction between matter and antimatter, and in the standard Big Bang/freeze-out argument, we have equal amounts of matter-antimatter that are mixed. While it is true that homogeneously distributed matter and antimatter no longer annihilate significantly after freeze out, in the standard picture the density of matter or antimatter does not remain uniform. Gravitational instability amplifies the initial perturbations to produce local density enhancements that are much higher than the mean overall density. For example, the typical interstellar density in our galaxy of $n \sim 1$ cm^{-3} is a factor $\approx 4 \times 10^6$ higher than the observationally inferred mean baryon density, $\approx \Omega_{b,tot}(\rho_c/m_p) \approx 2.6 \times 10^{-7}$ cm^{-3}. Even galaxy clusters, which are among the hottest and most diffuse gravitationally collapsed objects, have mean particle densities within a virial radius ($\sim$ 1 Mpc) that are $n \sim 10^{-3}$ cm^{-3}, a factor $\sim 10^3$ higher. Using the annihilation cross-section discussed above for 10^8 K cluster gas, we obtain an annihilation time for mixed matter-antimatter cluster gas $\sim (n_{cluster} < \sigma_{ann} v >)^{-1} \approx 10^{-3} \times 10^{-14}$ s $\approx 10^{17}$ s, which is *less* than the current expansion time, $\approx 4 \times 10^{17}$ s. (Note that central densities of clusters can be $\sim 10 - 100$ higher.) As noted before, the annihilation time for mixed antimatter-matter interstellar gas is correspondingly shorter. In other words, even if we managed to inject baryons and anti-baryons into the Universe after freeze-out, to the extent that they are mixed and equal in number, they would be gone today and collapsed structures such as the Sun and ourselves would not exist! This argument also tells us that if the Universe did have an intrinsic excess of matter over antimatter, the expected amount of primordial antimatter surviving in our galaxy today is ~ 0 since it should have annihilated during non-linear structure formation – unless we play some of the special games discussed below, e.g., see [82].

Leaving aside the expected baryon density problem for a moment, the standard Big Bang scenario also fails to explain the embarassing fact that our local universe indeed seems to have very little antimatter, i.e., we naively expected that $\eta \propto (n_b - n_{\bar{b}}) = 0$, while experimentally we are finding $\eta \sim 6 \times 10^{-10}$. Let us examine the expectation that $\eta = 0$, i.e., that the net baryonic charge of the Universe is $B = 0$, a bit more closely as it is a somewhat more profound statement than one might first think. First, we have the problem of naturalness. A non-zero value of the net baryonic charge B could represent an initial condition of the Big Bang, and in the early days this was indeed assumed to be the case, but it seems a rather arbitrary choice to the extent we say that there should be no essential difference between baryons and antibaryons. Some may now object that experimentally we know that the phenomenon of CP violation exists (whether or not we understand it), ie., particles and antiparticles are *not* the same since they can interact and decay differently. Indeed, CP violation lies behind many modern baryogenesis scenarios for producing $B > 0$ today. Note the following, however. While the CP symmetry may be broken, however, we do not expect CPT to be broken on the microphysical level, implying for example that the baryon and antibaryon masses that appear in the freeze-out should be exactly the same. (See, however, [83] for an example of "effective" CPT violation when coupling to an external field is considered.) We can use the CPT symmetry and unitarity, another basic assumption of field theories, to prove the following interesting theorem. Barring any explicitly conserved non-zero quantum charges, if the Universe is in a state of thermal equilibrium (as the spectrum of the CMB suggests it once was), the expectation value for the net baryon charge must be $< B >= 0$ [51, 84–86] – even if CP violation tells us that particle and anti-particle interaction rates are in fact different. This is because in thermal equilibrium, we are by definition in a static situation, and there is effectively no arrow of time: we can't have a total net production or destruction rate of any particle species. If B is strictly conserved (as it is in the standard model) and has a non-zero initial value, this of course violates the assumption of no non-zero conserved quantum charges, and in the thermal phase we end up with a chemical potential for baryons and antibaryons. Our theorem then simply tells us that we won't change the net baryon number during the thermal phase. Note one other interesting thing. In GUT theories, for example, baryon number is in fact not a strictly conserved quantity. At the very high temperatures where baryon number violation is possible, any initial non-zero value of B will then be wiped out as one approaches thermal equilibrium [86]. If we remain in thermal equilibrium as the B violation shuts off due to the Universe's cooling (a big if as we shall see), then initial conditions can't help us and we will not have a net B today. In other words, arguments such as these show that if we don't rely on mysterious initial conditions, generating a baryon asymmetric ($\eta \neq 0$) Universe is not a completely trivial endeavor.

As in the early days, let us therefore throw up our hands in disgust and just assume the necessary initial conditions. After all, as we go back in time before the freeze-out epoch when baryons were relativistic, we will have $n_b \sim n_{\bar{b}} \sim n_\gamma$ (modulo the usual g counting factors). This implies that the fractional excess of baryons we need to explain our existence, $(n_b - n_{\bar{b}})/(n_b + n_{\bar{b}}) \sim (n_b - n_{\bar{b}})/n_\gamma = \eta \sim 6 \times 10^{-10}$, is actually a very small number. In other words, we are not deviating very much from our idealized scenario, and the required initial baryon excess is thus a minor detail we should not worry about too much. Unfortunately, now that we are fairly certain the Universe went through an inflationary phase, such a viewpoint is no longer tenable. Inflation and a non-zero initial baryon number are simply not compatible, e.g. see [87]. After all, one of the nice features of the inflation scenario is the fact that nasty topological defects or unwanted charges or perturbations will be "inflated away." The current consensus seems to be that we need an exponential growth phase that goes through approximately ~ 70 e-foldings. This means that any pre-existing charges, e.g., an initial non-zero baryon number, will be reduced by a factor $(R_{final}/R_{initial})^3 \sim \exp^{210}$! The initial baryon excess required is therefore not a small number but huge. One could in principle tolerate this except for the following problem. Let us start with the small relative baryon excess that we see today. When baryons are relativistic, the energy density associated with this excess baryon charge is to order of magnitude $\rho_{b,excess} \sim \eta\rho_{tot}$. Now what happens as we go back in time and hit the epoch of inflation? The total energy ρ_{tot} will grow as R^{-4} until we reach the time of inflaton decay. Before this time, though, we are in the inflation phase, which is driven by an energy density that must be approximately *independent* of the Universe's scale factor (or we won't have exponential growth). What happens to the energy density associated with the excess baryons? If baryon number is a strictly conserved quantum number and we don't make radical changes to the particle physics and relativity, then the excess baryon energy density (in whatever form it may be at high temperatures) will still go as R^{-4} – even during the epoch of inflation since we can't make the baryonic charge go away. If at the end of inflation the ratio of the excess baryon to the inflation-driving energy density is $\sim \eta$, then after going back only ~ 6 inflationary times the excess baryon energy density dominates over the the inflaton-associated energy density, and prior to that we couldn't have been in an inflationary phase since the excess baryon energy density has the wrong dependence on scale factor R. To summarize then, if we go forwards in time, sustained inflation is likely to wipe out any initial baryonic charge yet we apparently have a non-zero one today. If we go backwards in time, we can't have sustained inflation unless we effectively wipe out today's net baryonic charge. Since we have yet to conclusively observe a proton decay, if today's Universe really is baryon asymmetric, many therefore consider the apparent baryon asymmetry of our Universe to be the strongest evidence we have yet for the existence of B-violating processes and the new physics this implies.

If our Universe is in fact only locally baryon asymmetric but globally symmetric, i.e., $<\eta>=0$ when averaged over sufficiently large scales, then the argument we just have given fails. It is therefore important to see if such a possibility is tenable. (It would allow us to salvage at least a bit of the aesthetically pleasing idea of a symmetric universe.) Indeed, as long as our understanding of the baryon-antibaryon annihilation process is correct (and we have tested it up to GeV energies in the laboratory), the only possibility for avoiding $<B>\neq 0$ today and the non-standard physics this implies seems to be to put the matter and antimatter in distinct spatial domains so it cannot annihilate effectively and we can live in a region dominated by one sign of baryonic charge. This idea of effectively separating matter from antimatter before freeze-out (so that the subsequent annihilation can no longer happen) is actually one of the oldest models for baryogenesis and has been the subject of much attention, e.g., [88–97] including Steigman's pre-inflation theory review [4] which strongly criticized the baryon symmetric models of his day. Because the possibility of anti-worlds next door to ours is still an intellectually fascinating one, we will spend considerably more time (in the next section) discussing the constraints on models with matter and antimatter domains.

For now, though, we'll just summarize the main problems and features of such models, starting with Steigman's main objections, many of which are still valid today. First, Steigman noted that the domain sizes assumed in published models were rather arbitrary, e.g., chosen to generate enough redshifted annihilation radiation to explain the cosmic gamma-ray background [98], and lacked a clear, self-consistent physical mechanism behind them. Second, the one known separation mechanism that almost certainly operated, thermal fluctuations in baryon and anti-baryon number, again failed miserably to generate the required baryon asymmetry. To see why this is so, note that before freeze-out the number of baryons and antibaryons is $\sim n_\gamma$. In any given volume V, then, the number of baryons (or antibaryons) is $N_b \sim n_b V \sim n_\gamma V \sim N_\gamma$, the number of photons, and we therefore expect Gaussian, thermal fluctuations over this volume to produce a net baryon excess $\delta N_b \sim N_b^{1/2} \sim N_\gamma^{1/2}$. The effective η_{th} for that volume is then the excess thermal baryon density $(\delta N_b/V)$ divided by the photon number density $(n_\gamma = N_\gamma/V)$, i.e,. $\eta_{th} \sim (\delta N_b/V)/n_\gamma \sim n_\gamma^{-1/2}V^{-1/2} \sim \delta N_b^{-1} \sim 10^{-70}$(!) for a typical galaxy cluster which observationally contains an excess number of baryons $\sim 10^{70}$. As I have just stated it, this argument is technically flawed because a cluster is a collapsed object. Since photons today are decoupled from matter, the volume occupied today by a cluster's excess baryons is much smaller than the volume V used above to calculate, say, N_γ. However, if we go back to the BBN era, photons are still tightly coupled to matter, and the typical primordial density perturbation is still very linear. The relevant η_b for primordial nucleosynthesis in the progenitor of our cluster is then the one we just estimated, η_{th}, which would imply that clusters (which are closed box systems from a nucleosynthetic point of view) should have very little helium, something we would have noticed by now. Another way to phrase this argument is the following. If we are to have a cluster-sized baryonic thermal fluctuation (with $\delta N_b \sim 10^{70}$), then for it to be reasonably likely, we must consider a volume of space so large that it contains $\sim 10^{140}$ photons, which is more than the number of CMB photons in the observable Universe , $\sim n_{cmb}(z=0) \times (c/H_0)^3 \sim 412$ cm$^{-3} \times (10^{28}$ cm$)^3 \sim 4 \times 10^{86}$... The third generic problem that Steigman noted is that even if one achieves the required separation before freeze-out, it is often hard to avoid some remixing of matter and antimatter at later times. For published models, the likely level of re-mixing lead to a violation of the much cruder BBN and CMB constraints available then, e.g., [4, 99–101]. We will shortly re-examine the problems caused by remixing or, in the most optimistic case, simple contact between the surfaces of matter and antimatter domains. They are not entirely unlike the problems faced by anti-stars or anti-planets. Such objects do not live in a perfect vacuum, e.g., anti-stars are surrounded by interstellar gas, and unless matter and antimatter repel today (there is no laboratory evidence for this), some annihilation is inevitable and sufficient to place constraints on their numbers.

As Steigman himself noted, his first objection (the lack of a good physical model) is not really a fundamental one but rather reflects a lack of sufficient "imagination" on the part of theorists. Indeed, since his review, there has been quite a bit more creative thinking, resulting, for example, in the "island universes" model of Dolgov et al., e.g., see [82, 87] where we have high density regions of matter or antimatter separated by much lower density voids that conveniently get around the matter/antimatter domain contact problem since the density in the voids is too low for significant annihilation to occur. Exactly how the newer particle physics models for matter-antimatter domain production work is beyond the scope of this lecture, but typically we have something like a phase angle in the Langrangian describing the relative amount of baryons to antibaryons. (See also [83] [102] for similar realizations of such particle physics models.) In the initial very high temperature Universe, CP is a good symmetry and this phase is essentially random and with a zero expectation value. At later times, the CP symmetry is broken, e.g., in a first order phase transition, which leads to domains of positive and negative baryonic numbers much like the aligned positive and negative spin domains in a ferromagnet. One might worry about the problem of domain walls (with huge energies) between the baryon and anti-baryon dominated regions, but apparently sufficiently clever theorists can speculate their way out of this, e.g., [87, 103]. While such a scheme thus sounds quite promising, it also suffers a potentially fatal flaw: if the the baryonic phase domains develop *after* inflation, there is again no way to match the observed local baryon asymmetry. If the domains are the result of interactions as they are in a ferromagnetic, then they must be causally connected and have a size smaller than the causality horizon at the time the phase transition occurs ($\sim cH = cR/\dot{R} = ct_{exp}$). Moreover, if we are to avoid an annihilation catastrophe, the phase transition (effective baryon-antibaryon separation) must occur well before the freeze-out epoch, ie., at temperatures $T > 20$ MeV. Following [87], the maximum domain size today is thus $l_{causal} \sim 5 \times 10^{16}(T_{\Delta B}/1GeV)^{-1}$ cm, where $T_{\Delta B} > 20$ MeV is the temperature when the symmetry breaking occurs. This is a rather small scale when compared to the present-day size of a galaxy ($\sim 10^{22}$ cm) or a cluster ($\sim 10^{25}$ cm) and the even larger sizes that such objects had before their baryons underwent gravitational collapse, i.e., we can't be explain the large-scale baryon asymmetry inferred from observations! Fortunately, this is one case where inflation may actually help us by allowing us to inflate the causal domains to the sizes we actually need. Although a universe that is globally but not locally symmetric is not the preferred scenario today, it is therefore not completely implausible and the consequences of such a universe thus bear further investigation.

Before we investigate them, however, let us summarize what currently *is* the preferred general scenario for baryogenesis. (Useful reviews of modern baryogenesis models can be found in [53, 84, 87, 104–107], and M. Trodden's lecture.) We saw above that the most likely result of inflation is a universe with no net baryon number, a condition that is likely to persist as long as we are in thermal equilibrium. But what if our preceding discussion was in fact an academic one since we are never *exactly* in thermal equilibrium? After all, we only need to generate a very small asymmetry, the expanding Universe definitely has an arrow of time associated with it, and particle species are constantly freezing out, i.e., definitely dropping out of thermal equilibrium. Thinking along these lines led Sakharov [3] to propose his famous three conditions for generating a baryon asymmetry without imposing it as an initial condition:

(i) we must be able to effectively violate baryon number B since we observe a non-zero B today, (ii) we must be able to violate the CP symmetry or we have no way to differentiate baryons and anti-baryons, and (iii) the Universe must be out of thermal equilibrium at some point or the theorem we gave above applies. (See [87] for a discussion of some more exotic scenarios showing why these conditions are in fact sufficient but not strictly necessary.) Although we are not sure exactly how these conditions are met and fit together to produce the observed baryon asymmetry, meeting them does not seem that hard. As noted, the Universe definitely has an arrow of time and CP violation is an observed phenomenon (although the currently known examples do not appear sufficient to produce the magnitude of the currently observed asymmetry). Besides the inflation argument, we also have reasonably well-motivated examples of particle physics that predict baryon number violation, e.g., the proton-decay GUT theories or the so-called "sphaleron" mechanism operating in a possible electroweak phase transition. Some fairly complete scenarios for how the asymmetry might actually arise have been worked out (e.g., see M. Trodden's lecture), but the devil so far has been in the details. We either fail to explain the actual magnitude of the observed asymmetry because of experimental constraints on the relevant particle physics, e.g., supersymmetry, parameters (see the lecture by Carlos Wagner), or we run into observational constraints such as the limits on gravitino production described in the last section. Perhaps all we need here too is just a bit more imagination. Better insight into new physics, e.g., as provided by the discovery of supersymmetry in next generation experiments, would also certainly help. The resulting baryogenesis scenario would be quite elegant, dynamically generating the observed asymmetry as a natural consequence of particle physics and the Big Bang, explaining why we see very few primordial antibaryons today without invoking problematic matter and antimatter domains, and making our location in the Universe like every other.

4. COSMOLOGICAL CONSTRAINTS ON THE EXISTENCE OF MATTER AND ANTIMATTER DOMAINS

If we can come up with a force that cannot be constrained by current laboratory experiments and that *forever* repels matter from antimatter on larger than laboratory scales, then in principle we can prevent matter-antimatter annihilation and the powerful signal it generates by separating matter and antimatter into distinct spatial domains. By ensuring that the baryons in these domains are sufficiently tightly bound together, either by gravity or another unknown force, we may also be able to prevent communication between these domains and thus kill a direct detection signal too. Depending on the degree of fine-tuning allowed, one might therefore conclude that one can never completely rule out the existence of anti-worlds via experiment. While this conclusion is not doubt true in an absolute sense, it is interesting that a creative theorist already has to work harder than one might have first thought. The reason is that our observations of the overall Universe have improved and are improving at a tremendous pace, and there is less and less freedom remaining to deviate from the standard picture that is emerging.

On small scales, for example, the AMS Shuttle direct detection experiment could plausibly have detected antihelium from anti-stars if they existed, e.g., [11], but saw nothing. Even better experiments of this type are coming. If we are convinced that anti-stars are relatively common and that not all antimatter ended up automatically in black holes, then at some point we would probably have to pay the ugly price of suppressing supernova explosions in anti-stars (which otherwise disperse antimatter on galactic scales). If we don't want to do violence to stellar microphysics, we can always assume that anti-stars only had low masses and always turned into white dwarfs, but then we have simply transferred the problem to some of other set of physics and the cycle of fine tuning begins.

On scales comparable to or larger than galaxy clusters, ie., the scales one must focus on if one does not invoke new physics on smaller scales and thus accepts the direct detection/annihilation constraints we have discussed, matter and antimatter domains inevitably represent large scale structure *not* expected from adiabatic and Gaussian random primordial fluctuations. With the very large-volume galaxy surveys that now exist (e.g., the Sloan Digital Sky maps galaxy clustering over much of the sky out to redshifts $z \sim 0.1$) and the firm detection and characterization of the CMB fluctuations in the WMAP all-sky maps, our knowledge of large scale structure has improved radically in the last few years. Although I will not quote any firm limits here and one may occasional hear talk of unusual structure, e.g., as possibly indicated by apparently low quadrupole moment reported by WMAP, if anything obvious like the early claims of periodic galaxy redshift structure had been confirmed, we would have heard about it. Baryogenesis models such as those summarized in [105] which predict a periodically varying baryon/antibaryon density structure or have island universes $\sim 10^{10}$ light years in size "floating in a sea of dark matter" are therefore probably already ruled out, whether or not they have problems with annihilation or invoke new physics to avoid it. In general, any new baryogenesis scenario should be checked for potential large-scale structure implications and then compared against the new data. (It goes out without saying that topological defects such as domain walls should be avoided in any putative baryogenesis scenario.)

Two other important points regarding the large scale structure of the Universe should be noted. First, through large scale surveys and very deep exposures, we have directly detected galaxies and quasars out to redshifts $z \sim 7$ [109, 110] when the Universe was less than 10^9 years old or $\sim 1/14$ of its current age in the concordance cosmology. As of now, there appear to be no discontinuities, say, in the densities of objects as a function of redshift. Also, the highest redshift quasars, many of which are bright enough to study in detail, appear to have properties such as elemental abundances and emission line velocity profiles that are very similar to those of nearby quasars. In other words, as we probe back very far in time and distance, we have yet to run into any obvious discontinuities or "edges" in our Universe and to a high degree, the basic physical parameters and constants of the Universe appear to be the same in very high redshift and very low redshift objects. Again, this is not necessarily what one expects in a non-homogeneous universe with large scale domains.

Secondly, although the following claim needs more work to back up rigorously, it appears that regions of the Universe that are separated by large distances do *not* have radically different histories, i.e., more or the less the same things seem to have happened on similar timescales everywhere in the Universe. This is of course a fundamental prediction of the Big Bang theory, but this also does *not* follow if we split up the Universe into domains. In particular, if we allow the sign of the baryon charge B (η) to change between domains, then why should the magnitude of the charge ($|\eta|$) be the same in different domains? In the model of [82], for example, the magnitude does vary, producing the expected consequences for the yields of primordial BBN elements in regions of large $|\eta|$ [111]. (Remember, the overall baryon or antibaryon density is one of the main parameters in BBN theory. The model of [82] cleverly arranges to have large fluctuations of $|\eta|$ occur only in very small scale regions so that global BBN predictions are not altered from the standard one – but this is not the case in all models.)

Do we see evidence for primordial abundance variations? Fig. 9 shows the best current set of deuterium to hydrogen abundance measurements obtained by measuring by the strength of foreground deuterium absorption features in the spectra of bright background quasars. The deuterium to hydrogen abundance ratio is often considered to be the most reliable BBN "baryometer" because it has a strong monotonic dependence on the baryon density (unlike, say, the helium 4 to hydrogen ratio, which is always ≈ 0.25), and perhaps more importantly, we think that deuterium can only be destroyed after it is created in the BBN. (The solid curve in Fig. 9 shows how the deuterium fraction should drop as function of the overall metallicity of the absorbing gas, i.e., the amount of stellar burning that has gone on.) This removes many of the systematic problems that plague attempts to measure the primordial abundances of other elements. We will discuss the accuracy of the BBN baryon density estimate further below, but for now, the reader should note the following. The names next to the data points in Fig. 9 give the sky coordinates of the background quasars, and we see that they are scattered across the sky, in particular we have quasars in both hemispheres. Also, the matter (antimatter?) clouds responsible for the deuterium absorption features are located over a range of redshifts $z \sim 2 - 3.5$ In other words, even though we only have 5 points available, we are indeed looking in very different places (and thus in possibly different domains), yet we only see a factor ~ 2 scatter in the measurements, corresponding to only a 30% scatter in the estimated baryon density. One might not be overly impressed by this – except that the BBN baryon density estimated from the mean of these measurements happens to agree almost exactly with that obtained from the CMB acoustic peak measurements [48]! This strongly suggests that the observed deuterium abundance scatter is largely experimental since the baryon density dependence of the CMB fluctuations reflects physics that is very different and operates at much later times compared to the physics of the BBN baryon density dependence. Note also that the CMB measurement results from the averaging of many regions located over

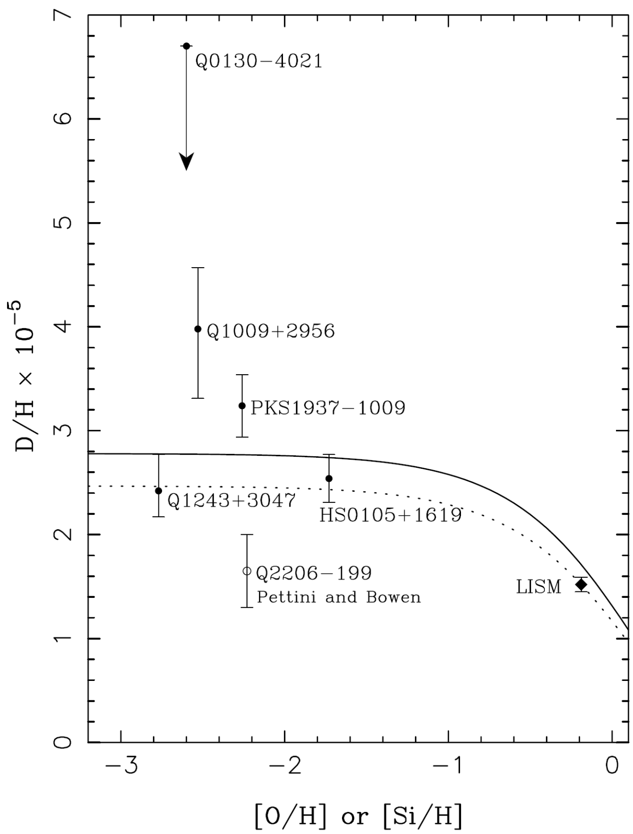

Figure 9: From [108]. The circles are measurements of deuterium abundance from the detection of foreground absorption features in the spectra of bright quasars. The measurements are plotted as a function of the abundances of silicon or oxygen in the foreground absorbing gas, indicators of the amount of stellar reprocessing that has gone on in the gas. The curves show a closed box stellar-chemical evolution model for the expected dependence of the deuterium abundance on these indicators, i.e., the quasar measurements should be very close to the true primordial abundance. The diamond data point labeled "LISM" is the deuterium abundance measured in our local interstellar medium. The solid curve is normalized to the mean of the quasar values while the dotted curve is normalized to the LISM value.

the *entire* sky. Primordial abundance variations can also have significant effects in the later Universe, e.g., enhanced gas cooling due to higher abundances of deuterium or lithium can completely change the mass spectrum of the first stars [112]. A systematic offset in the mean baryon density also changes the overall rate at which stars and galaxies evolve, yet there are currently no major discrepancies reported between groups studying various aspects of galaxy evolution that cannot be explained by the large scale structure fluctuations naturally expected in the standard Big Bang scenario.

In sum, even if we could forever suppress the annihilation and communication between antimatter and matter domains (there is currently no compelling scenario that does this), we have just seen that arbitrarily slicing up the Universe into regions of positive and negative baryonic charge can cause other problems. The slicing must be done in a clever way that mimics the standard "one domain" model of the Universe.

Now what if after the domains are created, matter and antimatter turn out to be subject to the standard laws of physics and can mix and annihilate? Note that this is actually the case for most proposed matter and antimatter domain scenarios: the domains are created by special high temperature physics, e.g., a phase transition or Omnes' matter-antimatter repulsion[90], that should not be important at the $\sim$ GeV-TeV energies and temperatures we can currently access in our laboratories. Since we now know our physics rules (including the fact that annihilation is allowed), if we are given a set of domain initial conditions we can in principle compute exactly the evolution and interaction of the domains, and not surprisingly, much stronger constraints usually result. As noted above, the problematic consequences of "remixing" the matter and antimatter in various domain scenarios were recognized early on [4, 99–101, 113, 114] and used to rule out several of the early scenarios,. To keep the presentation manageable I will focus here only on the modern versions of remixing constraints. Looking through the list of current cosmological probes presented above, we may expect three main difficulties when "too much" annihilation occurs on cosmological scales: (i) if the bulk of the annihilation occurs at late times, i.e., after recombination, we can see directly the annihilation photons as well as the additional radiation Compton upscattered by annihilation electrons and positrons and the intensity of this overall radiation may violate the current limits on a diffuse gamma-ray background, (ii) if the annihilation occurs below redshifts $z \sim 10^7$, we may violate the CMB limits on energy release and distort the CMB spectrum (since the limit is expressed as a fraction of the CMB energy, which drops with time relative to the energy density of baryons, the later the annihilation occurs, the stronger the constraint), and (iii) over a range of redshifts effects such as the annihilation of the neutrons needed for nucleosynthesis, the photodissociation of nuclei by annihilation cascade radiation, and the preferential annihilation of heavier nuclei may wreak havoc with the predictions of Big Bang Nucleosynthesis.

4.1. Gamma-Ray Background Constraints

Let us examine first the constraints resulting from the requirement that we not overproduce the gamma-ray background. As shown by Cohen and collaborators [115], this constraint can be surprisingly powerful if we make use of the CMB information and our current understanding for how the structure we see today formed. Consider first the Universe as we see it today. We now know that matter (and presumably antimatter) is arranged in a filamentary network with high density nodes at the locations where the filaments meet (corresponding, e.g., to clusters) and low density voids in between the filaments. By drawing our domain boundaries appropriately, e.g., locating the domains entirely in voids or centering them on the nodes and making them sufficiently large, one can probably make the annihilation signal expected today quite small, especially if we do not require the matter and antimatter domains to occupy equal volumes of space. A more concrete way to say this is to note that typical galaxy clusters have most of their mass within $\sim$ a few Mpc of their center but the typical separation between clusters is crudely ~ 30 Mpc. If we declared every other cluster to be inside an antimatter domain and put the domain boundary at $\sim 5 - 10$ Mpc from the cluster center (where the density of matter is low enough that the annihilation time exceeds the expansion time), we have in effect created a set of island universes bounded by voids, and thus we may avoid annihilation constraints. This is why we appear to be stuck if we wish to experimentally prove that the Universe is also baryon asymmetric on scales larger than $\sim$cluster scales. As [115] note, though, we are ignoring how the structure we see was created. If we run time backwards what do we see happening to large structure? Baryons move out of the high density nodes, back into the filaments and voids. (The baryons in cluster actually come from an initially much larger region than they appear to occupy today.) In the standard picture, at least, the contrast between the highest and lowest density points decreases, and by the time we reach back to the epoch of reionization, almost all density perturbations are back in the linear regime and the Universe, as reflected in the CMB, looks incredibly smooth down to a few parts in $\sim 10^5$. The voids that separated our island universes must be filled in with matter or antimatter and gone! There is no choice if we don't want to abandon the currently preferred theory of structure formation and its several successful predictions. Because the voids are filled in, matter and antimatter must come into contact and annihilation at some time in the past becomes inevitable.

This statement is so sweeping in its implications that we must look very closely for loopholes. In fact, there are some very important caveats that one must add. First, as discussed in [115], the technique of using CMB fluctuations to tell us about the Universe at the epoch of recombination has definite limitations. In particular, the sudden decrease in the Compton scattering optical depth of the Universe as electrons recombine into atoms is *not* an instantaneous phenomenon, occurring between redshifts $\approx 1000 - 1200$. In other words the "surface of last scattering" has a finite thickness with a characteristic comoving half-width measured in today's physical distance units [ie., using today's scale factor $R(t_0)$] of ~ 15 Mpc. If we look at the CMB fluctuation spectrum on the angular scales corresponding to sizes smaller than this (roughly $l \gtrsim 1400$ in Fig.6), the effect of features, e.g., voids, smaller than this size scale will be washed out, and thus we cannot say much about their existence. Note the lack of strong CMB constraints on these small angular scales increases dramatically the uncertainty in the small-scale end of the "standard" theory of structure formation. To check if we understand what is going on, we have to rely instead on galaxy clustering studies and plausible guesses for the initial perturbations that drive small-scale structure formation. The situation is further complicated by the fact that the perturbations are dominated by the cold dark matter, which we currently do not understand and in fact might not act so cold on small scales. In addition, on scales smaller than $\sim 5-10$ Mpc today, messy non-linear baryonic feedback effects such as the input of energy from stars become important in determining what we actually see on the sky in terms of galaxies. The bottom line is that we still do not understand well how small-scale structure arose and there could be surprises in store for us, [117].

At first sight, then, it appears that we can say nothing firm about what the baryon distribution looked like at recombination on scales smaller than 15 Mpc. In particular, matter domains and antimatter domains could still be separated by $\lesssim 15$ Mpc, with very little resulting annihilation. After a bit of thought, though, one might wonder if this statement can really be true since there is no obvious feature at ~ 15 Mpc scales in galaxy clustering statistics, e.g., as one might have expected from domains with an empty 15 Mpc buffer zone around them. Indeed, this likely means that the cold dark matter density fluctuations do *not* have empty zones with a characteristic size ~ 15 Mpc. However, since baryons have turned out to be a negligible component of the overall density budget and thus are not the main factor determining the large-scale structure fluctuations we see today, it turns out we have a bit more "wiggle room" in what we can do with baryons, at least on very small scales. This flexibility is what we need to understand better to see if annihilation is inevitable, and I will spend a bit of time on it because the discussion will prove useful in the next section too.

To define the wiggle room available, we first need to understand better what happens with baryons in the standard picture (see e.g., [49] for a much more in-depth discussion). Let us focus on late times when baryons are the only particles that remain coupled to photons but the total energy density is still dominated by the radiation. Looking at the linearized equations for the evolution of density perturbations in this two-component, radiation and baryon, Universe we find that there are two "normal modes" in the relevant second order equations: the *adiabatic* mode (discussed above) where the relative matter and radiation density perturbations are proportional, ie., $\delta\rho_m/\rho_m \propto \delta\rho_r/\rho_r$ and the total energy density is changed ($\delta\rho \neq 0$) while the entropy remains constant (hence the name adiabatic), and the *isocurvature* mode where the matter and density perturbations are exactly out of phase so that the total energy density (which determines the space-time curvature) remains constant ($\delta\rho_m = -\delta\rho_r$ so $\delta\rho_{tot} = \delta\rho_m + \delta\rho_r = 0$) while entropy changes. In general, an arbitrary radiation-baryon perturbation can be decomposed as a combination of these two modes. Of the two modes, the adiabatic one initially grows fastest, comes naturally out of scenarios like inflation, and is the one that usually comes to mind first when talking about "density" perturbations. (Strictly speaking, there is no overall density perturbation in the isocurvature mode since $\delta\rho_{tot} = 0$.) Hence, "adiabatic perturbations" are the ones that have received the most attention, and in the current standard cosmology scenario, the primordial perturbations that influence the CMB are assumed to be solely of this type. (See [116] for a precise and general discussion of all the possible primordial fluctuations modes that are possible, e.g., when we do not restrict ourselves to considering just baryons and radiation.)

As long as radiation and baryons remain tightly coupled one can show that both modes have oscillatory solutions (the "acoustic oscillations," but with an offset in phase between the modes), in other words the baryon density perturbations of interest to galaxy and cluster formation do not grow. What happens next, as electrons recombine and baryons and radiation decouple, is one of the main differences between the two types of perturbations and key to our discussion. The baryons and photons are effectively coupled by Compton scatterings, and thus on size scales smaller than a scattering mean free path (which gets larger and larger as one approaches recombination) one can no longer consider the photons and baryons as a single fluid. Consider what happens now if we have a positive adiabatic density fluctuation with a linear size smaller than say a few Compton mean free paths. Because the density enhancement is adiabatic, it is hotter and has a higher density than its surroundings. Photons, which are now relatively mobile, will therefore diffuse out of the high density peak into the lower temperature, lower density surroundings, dragging baryons with them in the process. The net result of this photon diffusion, called

Silk damping[118], is a low-pass filter that suppresses perturbations in baryon density and temperature that have a comoving (present day) size scale $\sim$ 15 Mpc – see Fig. 6 for the effect of this damping on the small-scale CMB power spectrum (Although a detailed solution of the relevant Boltzmann equations is required to obtain this scale, it is not a coincidence that it is similar in size to the thickness of the last scattering surface since since the thickness is also defined by the transition from tightly coupled to free-streaming, uncoupled photons.) Cohen *et al* [115] are very pleased with this outcome as it means that any voids that are too small to detect in the CMB ($\lesssim$ 15 Mpc) in size are effectively filled in by the time recombination is over. Conversely, any voids large enough to have survived Silk damping (and empty enough to avoid problems with annihilation) would be detectable as abnormally large CMB temperature fluctuations – and as such fluctuations are not seen, such voids do not exist. Therefore, even if today one can conveniently locate all antimatter-matter domain boundaries inside voids, that cannot be true at the time of recombination because the baryon density then should be exceedingly uniform on *all* scales – at least for adiabatic perturbations.

Now before we turn to the possible escape route provided by isocurvature baryonic perturbations, let us continue following the fate of baryons in the adiabatic perturbation picture. Our goal is to see how long the baryons will remain uniform, i.e., for how long $\delta\rho_b/\delta\rho_b < 1$ so we can guarantee that voids have not formed at domain boundaries. Once we know the time when baryonic perturbations start to go non-linear, we can compute what is a *minimum* annihilation signature by integrating the number of annihilations that occur from recombination up until that time. (If a domain boundary happens to instead lie on a positive density enhancement, for example, the annihilation rate in fact does not go to zero as the density enhancement becomes non-linear but rather it will increase dramatically.) Note that this annihilation signature has *no* free parameters except for the domain configuration, i.e., the volume filling factor of the antimatter (or matter) domains and their typical surface area to volume ratio. This is because the (complicated) structure and dynamics of the annihilation zone at the interface between matter and antimatter domains can be calculated from first principles once one knows the pressure and density of the baryons and antibaryons on each side of the interface. Since we are only considering times when baryonic perturbations are linear, the relevant pressures and densities are essentially the same for baryons and antibaryons and are simply given by the mean background density and pressure in the standard Big Bang scenario. (Because Compton cooling of annihilation electrons can be important, the background radiation field actually also matters – but this is known too in the standard scenario.)

So exactly when do adiabatic baryonic perturbations go non-linear in the standard scenario? If baryons were the only matter component and had the mean density inferred from BBN and the small fluctuation amplitudes at recombination inferred from the CMB, the somewhat troubling answer is never(!), especially on scales less than 15 Mpc where Silk damping has operated. Therefore something else has to drive the non-linear perturbations we see today. The current answer is that cold dark matter, which dominates the matter density, is responsible. Because baryons and photons remain tightly coupled until recombination, baryonic perturbations cannot grow significantly until the coupling is broken (oscillations result unless the perturbation is extremely massive). Cold dark matter, on the other hand, is presumed to have decoupled from baryons and photons at very early times and its perturbations can start growing as soon as the mean matter density exceeds that of the relativistic constituents of the Universe (i.e., photons and neutrinos). This happens before the time of recombination, giving cold dark matter perturbations a relative head start. Also, because cold dark matter essentially couldn't care less about photons after matter starts to dominate the density of the Universe, there is no Silk damping problem. The non-linear baryonic structure we see today is then not caused by primordial baryonic fluctuations that have grown but by the gravitational infall of baryons at late times into the potential wells of the much larger cold dark matter fluctuations. Now, the cold dark matter perturbation power spectrum at lates times has the interesting property that the smaller scales (masses) are the ones that go non-linear first. Hence, the process of structure formation mediated by cold dark matter is often described as a "hierarchical" one, where large objects form from the coalescence of smaller objects. The baryons, however, have a pressure and are relatively hot (their temperature stays close to that of the CMB until $z \sim 100$), i.e., they resist gravitational collapse onto cold dark matter fluctuations that involve too little mass. (To see why, idealize a cold dark matter perturbation just starting to go non-linear, $\delta\rho_{CDM}/\rho_{CDM} \sim 1$, as ball of mass $M \sim \rho_{CDM}\lambda^3$ where λ is the size scale of the perturbation. If baryons near the perturbation have a typical thermal velocity that significantly exceeds the escape velocity from this ball, they won't be significantly by the perturbation. Pursuing this train of thought gives rise to the so-called "Jeans mass" criterion.) In addition, even if baryons become bound to dark matter perturbations, they cannot undergo a cosmologically interesting runaway collapse if they are not able to radiate away their energy (cool) on less than one expansion time. Putting everything together, we see that the first collapsed baryonic objects start appearing at redshifts ~ 20, e.g., see [50]. This is the lower redshift bound used by Cohen *et al.* in their calculation [115] and is quite conservative because the $\sim$ Mpc-sized structures and voids we were considering above do not appear until much later.

As we shall shortly see, the result of their calculation implies that if our Universe is globally baryon symmetric, we can only afford to have ~ 1 matter or antimatter domain in our observable Universe, i.e., for all practical purposes we live in a baryon asymmetric Universe. Such a conclusion, of course, does not please those who prefer a more lively Universe but the only way around it is to argue that matter and antimatter do not inevitably come into contact by recombination. Baryon density perturbations dominated by that other mode, isocurvature, may in principle allow this. To see why, let us first see exactly what an isocurvature perturbation is and whether it can be physically generated. The reader interested in details is cautioned that the term "baryon isocurvature perturbation," is often employed in the literature in a rather imprecise way, and that initially isocurvature ($\delta\rho_{total} = 0$) perturbations do not strictly remain so after pressure forces become important (i.e., they can induce density fluctuations with $\delta\rho \neq 0$). At very early times, the mathematical definition given above, $\delta\rho_{total} = 0$, is precise enough and indicates that we are considering perturbations that only change the relative abundances of the species that contribute to ρ_{total} but not their total total energy. While it is clear how one might realize an exactly adiabatic perturbation, e.g., from inflation, it is not so obvious why particle physics produces changes in, say, baryon density that are exactly out of phase with changes in the photon density, as required for an isocurvature perturbation. A much more realistic scenario, that might actually happen, is that a phase transition, or whatever mechanism we like, sets up a baryon density field that has a varying spatial amplitude (and of interest here, also a possibly varying sign of the net baryonic charge) that has no initial correlation with the photon field. The minute this happens we have a spatial varying baryon to photon ratio, which cannot be represented as a purely adiabatic fluctuation. At the same, the baryon and photon density initially do not compensate each other, so we do not have an isocurvature perturbation either. Note what happens, though, if early on we create an excess baryon density field with the order of magnitude required to explain what we see today ($|\eta| \sim |n_b|/n_\gamma \sim 10^{-9}$). As we argued above, the energy density associated with these excess baryons at early times is a negligible fraction $\sim \eta$ of the total energy density. This means the largest total energy density perturbation we can induce through these excess baryons has $\delta\rho_{tot}/\rho_{tot} \sim |\eta| \ll 1$, i.e., we have an "almost" isocurvature perturbation. Morever, if we took all the energy in the baryons and somehow transferred it to the photons, the maximum temperature change we what induce would be $\delta T/T \sim \frac{1}{4}|\eta|$, ie., the temperature hardly changes. For this reason, an overall fluctuation of the type I have just described is often called an "isothermal" perturbation. The key difference in this type of perturbation compared to an adiabatic one is that we are now allowed to have baryon voids that are *not* accompanied by a photon voids and in some realizations, e.g., [82], the effective temperature of baryons in positive baryon density enhancements may initially be even lower than that of the surroundings. If we have such a situation (a baryon island floating in a photon sea), photons have no incentive to diffuse out of the baryon overdensity, and if anything, they want to diffuse into it. In other words, if one does a proper analysis (e.g., see [50]), one finds that Silk damping does *not* operate effectively for isocurvature/isothermal perturbations.

This is the possible loophole we have been looking for, but is it of relevance in cosmology as we understand it today? One advantage of having baryonic perturbations that are isocurvature/isothermal is that the small scale perturbations of this type survive recombination because their Silk damping is small. Hence, they undergo runaway gravitational collapse as soon as they become non-linear on large enough scales to satisfy the Jeans mass instability criterion ($\sim 10^6$ solar masses, the mass of only a globular cluster). Models based on baryonic isocurvature perturbations, e.g., [119], therefore tend to have an early appearance of small structures and stars. Historically, this fact has been used to increase the amplitude of the inferred perturbation spectrum at the small scales (~ 10 Mpc). Unfortunately, isocurvature baryonic perturbations also have a potentially serious problem. If one compares

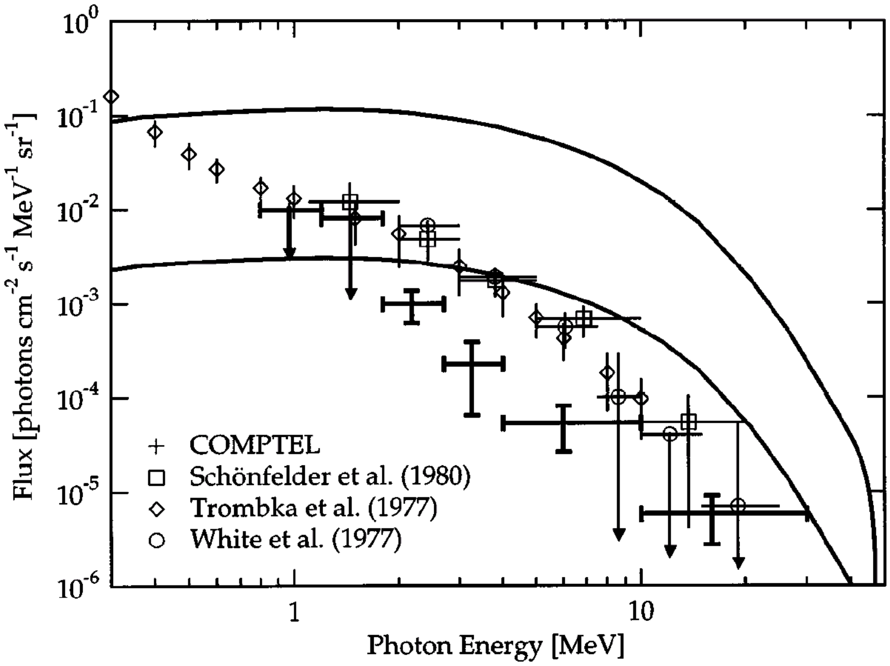

Figure 10: From [115]. Comparison of the $\sim$ MeV extragalactic diffuse gamma-ray background (the COMPTEL points are the most recent) with the total gamma-ray flux produced by the "unavoidable" annihilation between matter and antimatter domains during the redshift interval $20 \lesssim z \lesssim 1000$. The upper curve is for a characteristic domain size $d_0 = 20$ Mpc. The lower is for $d_0 = 1$ Gpc. The curves[115] are computed assuming a universe that is symmetric on large scales, i.e., with matter and antimatter domains that are comparable in size and equal in number.

the CMB fluctuations produced by isocurvature vs. adiabatic baryonic density perturbations of the same amplitude ($\delta\rho_b/\rho_b$), the fluctuations due to isocurvature perturbations are larger and at a different place in the angular power spectrum (e.g, [69] [70]). In other words, on the >10 Mpc scales that one can study well with the CMB, isocurvature perturbations lead to a distinctive CMB signature that can be used to rule them out. With proper fine tuning, this potential problem can be remedied to some extent by noting that if we create enough early stars, they will reionize the Universe, effectively creating another scattering surface that can conveniently wash out the the inconvenient CMB signatures of isocurvature perturbations [120]. This is not necessarily a great cure, though, since *all* CMB fluctuations will tend to be washed out.

So what does the new CMB data tell us about the possible existence of isocurvature fluctuations? Interestingly, a signature of reionization *has* been detected by WMAP (in the non-zero polarization seen at the largest angular scales, bottom panel of Fig.6), and it is not entirely clear that enough early star formation occurs in the standard cosmological scenario to explain it! This has led some, e.g., [117], to invoke the addition of a somewhat *ad hoc* isocurvature component to the total density perturbation spectrum on very small scales. (Note, though, that the error bars on the amount and epoch of reionization are actually quite large and we do not completely understand early star formation, so that there may actually be no problem.) On the other hand, what is clear from current data is that the overall effects of reionization are *not* that strong and on >10 Mpc scales, one can rule out that purely isocurvature baryonic or cold dark matter perturbations explain what we see [80]. One of the reasons why is that we have data from experiments with higher angular resolution than WMAP (see Fig.6) that can probe into the regime where Silk damping is expected to start being important for adiabatic perturbations (at roughly $l > 800$) – and we see the expected damping signal! (The error bars do get large, though, in the regime of interest to Cohen *et al.*) Considering a mixture of adiabatic and various isocurvature perturbations (not just in baryons, a more realistic case), the general conclusion is that the adiabatic mode component dominates [80] except in the case of some particularly clever combination of modes [121]. However, while combinations of modes with a non-negligible non-adiabatic component can be found that are consistent with the data, the fit to the data generically does *not* improve enough to justify the introduction of the new free parameters associated with the extra modes. In other words, Occam's razor would currently say that the observed CMB fluctuations are predominantly adiabatic.

If one then assumes that the processes responsible for generating perturbations on large scales (which are subject to CMB constraints) are the same as those responsible for perturbations on small scales (as they are in the standard inflation scenario), the isocurvature loophole is largely closed. The big catch here is that we currently do not have enough evidence to conclusively prove this reasonable-sounding assumption. The fact that cosmological constraints derived from galaxy and cluster studies fit into the concordance cosmology picture means that the structure formation assumptions made to derive those constraints are probably correct and that we in fact understand what is happening on comoving scales down to ~ 1 Mpc. On scales smaller than this, though, something new and "beyond the standard model" could be happening. Information on these scales may eventually be provided by studies of foreground Lyman α absorption systems in quasar spectra (these) and by the up-coming Planck mission that has much better angular resolution plus polarization sensitivity than WMAP. In the meantime, scenarios such as that by Dolgov & Silk [82] (see also [122]), where large baryonic fluctuations are produced on very small scales, are still somewhat allowed. The reason I say "somewhat" is that isocurvature perturbations with significant baryon voids do bad things to the BBN predictions (next section). In order to avoid these potential problems, Dolgov & Silk have to postulate that most of the Universe has the standard adiabatic, matter-dominated state and have to allow themselves the freedom to hide as much as necessary of the antimatter by collapsing it into black holes or quark nuggets. They do warn (predict), though, that the antimatter that does not make it into black holes is detectable via the kinds of annihilation signals we are discussing here.

The bottom line is that modulo the possible (but currently not compelling) loophole just discussed, the results of the Cohen et al. calculation [115] represent a powerful constraint on the existence of antimatter domains. In their specific calculation, designed for a symmetric $B = 0$ universe, [115] assumed that the Universe can be divided evenly into matter and antimatter domains with a similar characteristic constant comoving size d_0. They then integrated the equations for the annihilation zone at the domain interfaces from $z \approx 1000$ to $z \approx 20$ to compute the amount of annihilation energy per unit area of domain interface that could go into Comptonizing the CMB and producing gamma-rays. Since annihilation between domains is a surface effect, the total annihilation signal expected then scales as $\sim 1/d_0$, which, for domains that are not too oddly shaped, is roughly the surface area to volume ratio of the domains. The CMB distortion due to Compton scattering was not found to be interesting for current experiments, but as shown in Fig.10, the level of gamma-ray flux predicted was. To avoid overproducing the current best limits on the MeV background (from COMPTEL), we require $d_0 \gtrsim 1$ Gpc, i.e., a significant fraction of the characteristic size of the observable Universe ($\sim ct_{exp} \sim 4$ Gpc today). If we do not have a symmetric universe, then the situation of course is more complicated. If we make the simple generalization that the Universe is filled with a few antimatter domains that are in contact with and have the same density as the surrounding matter, then we can re-use the result of [115] by rescaling it by the volume filling factor of the antimatter domains, ie., we have the constraint $(d_0/f) \gtrsim 1$ Gpc where f is the filling factor and d_0 is the characteristic size of the antimatter domain. (Let there be N antimatter domains and let V be the volume of the Universe. Then $f \sim (Nd_0^3)/V$, and the annihilation gamma-ray flux $\propto Nd_0^2 \sim fV/d_0$.) If the typical antimatter domain size is ~ 20 Mpc, then the domains must occupy a fraction of the observable universe that is $f \lesssim 1/20$. If we follow [82], for example, and put the antimatter into very dense nuggets or black holes, with much higher densities than the surrounding matter, we can of course make the annihilation gamma-ray constraint go away. But then, as we discussed above, we are adding something new to the standard scenario for structure formation... and only in a suitably small fraction of the Universe.

Overall, this discussion of the gamma-ray background constraint illustrates why it is so important to get the best possible limits on the extragalactic gamma-ray background and then remove the contribution from any non-exotic sources (e.g., AGN) if one does actually detect a diffuse background. With its greatly improved sensitivity and angular resolution as well as a scanning observational mode (which guarantees much more uniform sky coverage), GLAST should represent a major step forward in unraveling the nature of the gamma-ray background, at least in the 100 MeV range which is relevant for constraining low redshift domains. (See [28, 42–44] for a discussion of some

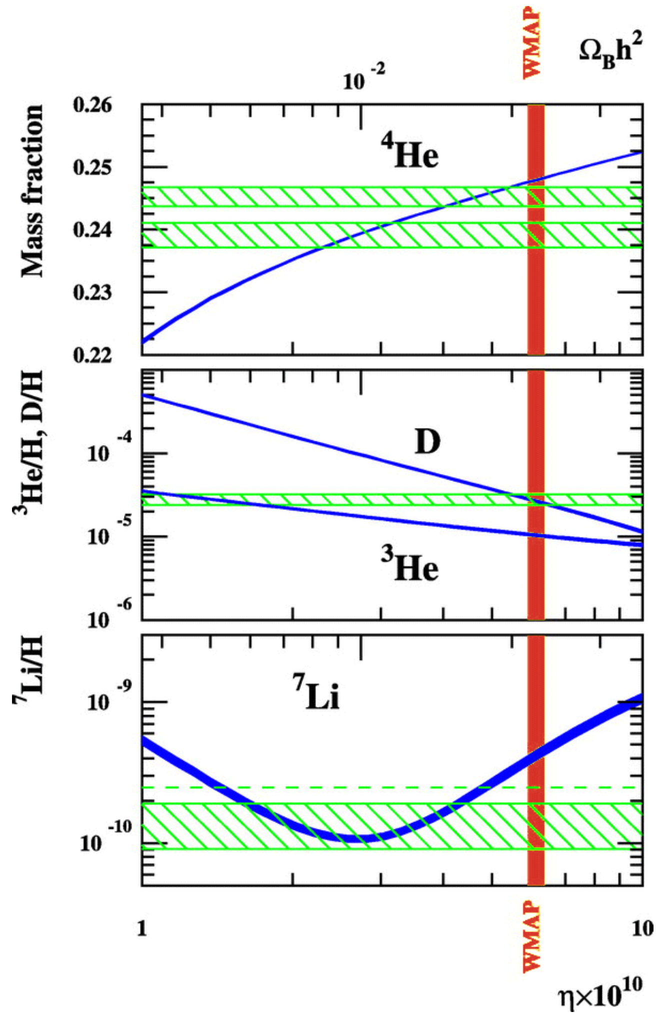

Figure 11: From [124]. Primordial abundances of the light elements as a function of the baryon-to-photon ratio, η, or equivalently the baryon density, $\Omega_b h^2$. The curves show the results from a current standard BBN calculation, where the thickness of the curves represents the 1σ theoretical uncertainties of the calculation (e.g., due measurement errors in the relevant nuclear cross-sections). The hatched regions show the best current observational estimates for the primordial abundances, see [124] for details. The vertical bar represents the 1σ constraints on η provided by WMAP[48].

of the current controversies on exactly what diffuse background was detected by EGRET – it depends critically on the correct removal of the not-so-well determined galactic gamma-ray emission.) Stranger things have happened, and instead of cutting off at redshifts below ~ 20 as Cohen et al. conservatively assumed, the domain boundaries might be very badly aligned with respect to the collapsing matter and the annihilation rate might actually increase significantly with time instead. For a sufficiently small antimatter filling factor f, the $z = 20 - 1000$ unavoidable annihilation constraint goes away, and we can't exclude that we have a few domains annihilating away today. Instead of constraining domain boundaries, GLAST might then end up detecting them via anisotropies in the gamma-ray background, e.g., [123] (although EGRET saw nothing of the type predicted in that paper).

4.2. Constraints from Big Bang Nucleosynthesis

As previously noted, primordial elemental abundances from BBN are very sensitive to changes in the physics or the physical conditions during the BBN epoch. To the extent that the observed abundances agree with those predicted from standard BBN, this allows one to place strong constraints on deviations from the standard scenario, the most

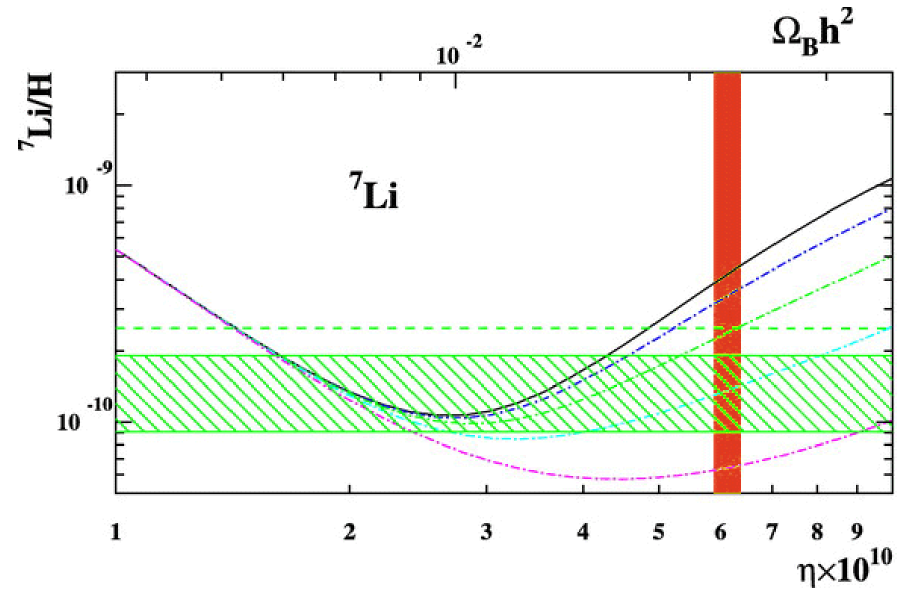

Figure 12: From [124]. Same as the bottom panel of the previous figure, except that the theoretical curves are computed assuming different values of the poorly measured ^{7}Be$(d,p)2^4$He reaction cross-section, see [124] for more.

famous perhaps being the constraint on the number of neutrino generations ($N_v \leq 3.3$, e.g., see [49]). To assess how well the abundances agree with the latest standard BBN prediction, see Fig. 11 from [124] (see also [125]). The hatched regions show the range of values allowed by observations, while the solid curves show the expected abundances as a function of η (baryon density), probably the most uncertain parameter in standard BBN before WMAP. Clearly some systematic problems remain, especially with the lithium abundance, since no value of η gives predictions that lie exactly within all the observational bounds. Current thinking is that this is probably due to problems in reconstructing the primordial abundances from the observations (which except for deuterium is not at all straightforward) or possibly errors due to poorly determined cross-sections that are not accounted for in the thickness of the theoretical abundance curves, e.g., see Fig.12 for the effect of a potential uncertainty in the lithium production chain. If these possibilities are eventually ruled out, then we actually have positive evidence for non-standard physics, not just constraints on it, e.g., see [126]. For now, though, it is probably a safer bet that standard BBN is close to the truth given the spectacular agreement between the values of η determined by deuterium and CMB measurements.

The importance of the CMB constraint on η, which is based on independent physics, cannot be stressed enough, and is perhaps the most important development in the BBN field since the improvements in the measurements of deuterium and lithium abundances and the relevant nuclear cross-sections and rates. An interesting example of the CMB constraint's impact that is also relevant to our discussion of domains is the animated discussion in the late 1980s on the topic "inhomogeneous" BBN. At the time, the QCD-hadron phase transition was speculated to have produced nearly isothermal (isocurvature) baryonic fluctuations of exactly the type we have just been discussing. (See [63–65] for reviews and [127–129] for some of the original papers). In particular, the net baryon density constrast between regions that remained in the hot quark phase and regions that had nucleated to form hadrons (nucleons) could be quite large for not obviously unreasonable QCD parameters (e.g., ~ 60 in [129]), which sounds a bit like baryon islands surrounded by voids. (Note, though, that the QCD phase transition discussed was *not* envisioned as changing the sign of the net baryonic charge, i.e., there were no antimatter islands. Also, the discussion then focused on very small-scale inhomogeneities.) Excitement arose because [127] had pointed out that if high and low density baryons regions had the right sizes, the significantly greater diffusion length of neutrons compared to protons (which are charged and couple strongly to the background photon-electron plasma) would lead to the creation of neutron rich and neutron poor regions. This fact, when combined with appropriate density contrasts between the high and low density regions, allowed one in turn to reproduce reasonably well the helium and deuterium abundances (but alas not the lithium ones) for values of the averaged baryon density as high as the critical density, i.e., one could arrange for $\Omega_b \sim 1$ to be consistent with the abundance data! (The stubbornly low value of Ω_b required for standard BBN to work was one of the main impediments to the success of a purely baryonic universe scenario, which was still being seriously contemplated then.) Subsequent improvements in calculations and abundance observations showed that in fact one could not use this specific mechanism to push the value of Ω_b much above the range allowed by standard BBN, and interest in the topic died. The point, though, is that if we had we known then that $\Omega_{b,tot} \approx 0.04$, the entire discussion would have been moot. In sum, unless something funny has happened to change the baryon density between the BBN epoch and the time when the CMB temperature fluctuations were imprinted, the CMB measurement completely removes one of the main degrees of freedom of BBN scenarios, standard or otherwise. (And if something has happened, we would then have to explain the major coincidence between the standard BBN baryon density and the CMB one.) All BBN-related constraints therefore tend to tighten considerably. A very dramatic example of this will be seen below when we discuss the possibility of large lepton number due to neutrino degeneracy (§5.2).

Returning to our discussion of distinct matter and antimatter domains, if we wish to prevent catastrophic annihilation by effectively separating matter and antimatter, we must in general do so before nucleon freeze-out. If we then want to avoid subsequent annihilation by having sufficiently large voids or low density regions between the domains, e.g., as desired above to avoid the gamma-ray background constraint, the overall baryon density field must at some point have *non-linear* fluctuations (e.g., $\delta\rho_b/\rho_b \sim 1$, if we have a void). Since typical baryonic perturbations do not grow significantly before recombination, these fluctuations (e.g., voids) must have been part of the initial baryon density field, i.e., in the technical jargon we are dealing with "primordial non-linear baryonic isocurvature (isothermal) perturbations." If we already know how to create a baryon density field where the sign of the baryonic charge varies, creating one that also has regions of $\sim$ zero charge is presumably not a problem if one does so at sufficiently early times (but see [130] for mechanisms that can damp out non-linear baryonic isocurvature fluctuations on small scales if one is not careful). What is a serious problem, however, is that our desired baryon density field then has large fluctuations during the BBN epoch, i.e., we have an example of inhomogeneous BBN. The outcome of inhomogeneous BBN studies (see in particular [131, 132] for studies of the effects of large-scale baryon inhomogeneities) is that large amplitude baryon density fluctuations present during BBH epoch generally lead to primordial BBN abundance predictions that cannot be reconciled with the data unless the low density phase dominates and matches what we see today and the high density phase is sufficiently hidden in black holes or dark matter, e.g., [128] (which is indeed the case in the Dolgov & Silk [82] scenario). Given that we are not allowed to play with the overall baryon density any more, the constraints on such fluctuations from BBN considerations are that much stricter today. Note that even Dolgov & Silk-type scenarios, if they are not perfect in hiding high density region, can be constrained by searches for systems with anomalous abundance patterns. (I know of no compelling examples for such systems, but searches admittedly have not been comprehensive. Note that since very high density BBN can lead to the production of elements beyond lithium, observers should not automatically restrict themselves to looking at low metallicity systems.)

Given the problems that primordial baryonic voids create for BBN and also the CMB (if they have scales $\gtrsim 10$ Mpc today), let us give up on the idea of keeping matter and antimatter domains separated and ask a different question. The gamma-ray constraint tells us we probably don't have too many large-scale antimatter domains from recombination onwards, but what about well before then, e.g., see [133, 134]? In particular, it could be that while we don't create strong baryon density constrasts, we may still make antimatter regions, i.e., our domain creation mechanism can change the sign of the baryonic charge density but not significantly its magnitude. If these antimatter regions are sufficiently small in size, one can show they would have annihilated away by today (or recombination), leaving us with our presently asymmetric Universe. This is the case studied by Rehm & Jedamzik [135, 136] and Kurki-Suonio & Sihvola [138–140]. (See Rehm's online thesis [137] for a good review of the physics.) Even if we equate the densities of the matter and antimatter regions so as to minimize the inhomogeneous BBN problems just discussed, those authors show that the existence of antimatter domains will still create problems for the BBN abundances (and eventually the CMB spectrum) if enough annihilation occurs – and some annihilation is inevitable

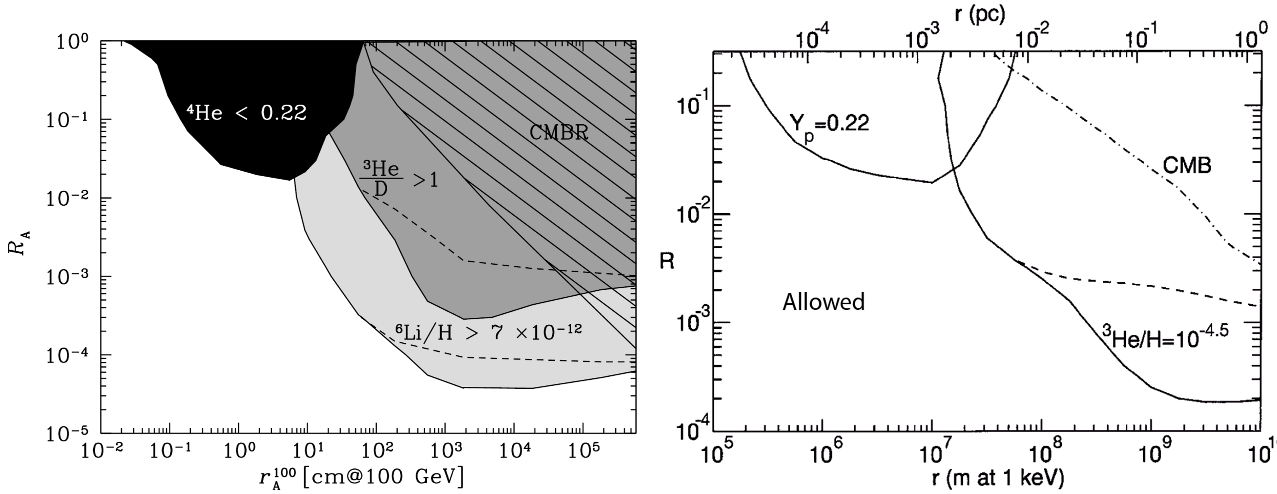

Figure 13: Left panel: Limits on the presence of antimatter in the Universe from the BBN calculation of [137]. r_A, which determines when the antimatter domains annihilate, is the assumed characteristic size of the domains measured in comoving coordinates defined when the Universe's temperature was 100 GeV. R_A is the overall ratio of antimatter to matter in the Universe. The allowed parameter region for the domains lies below below all the curves. The dashed lines indicate the results when ^{4}He photodisintegration is ignored. The hatched region represents the (usually weaker) antimatter annihilation limit imposed by the requirement that spectral distortions of CMB not exceed the limit from [71]. Right panel: Similar calculation by [139]. The axes are the same except the antimatter domain region size r is now measured in comoving coordinates set when the Universe's temperature was instead 1 keV (so that $r_A = 10^{-6} r$).

now that we are letting domains touch. The type of problem can be classified by the characteristic size of the antimatter domains or, equivalently, by the epoch when they annihilate away (smaller domains annihilate earlier). Domains with a comoving size today smaller than $\sim 2\times10^{-5}$ pc will be gone before nucleosynthesis and have no BBN constraint. Domains larger than this can cause problems. For regions annihilating just before the lock-up of neutrons into ^{4}He is complete (when the Universe reaches a temperature $\sim$ 80 keV in standard BBN), the main effect is due to neutrons and antineutrons (which have a much longer diffusion length than protons) crossing domain boundaries and annihilating. This decreases the neutron-to-proton ratio, lowering the predicted primordial ^{4}He abundance. If the typical antimatter domain annihilation occurs after this, there are not many free neutrons left and the main effect is due to the protons and antiprotons. These can annihilate with themselves, releasing energy that, after cascading, photo-disintegrates the ^{4}He. In addition, antiprotons can also disrupt ^{4}He by annihilating away one if its protons. The net result is the production of energetic ^{3}He and deuterium nuclei, which can exceed the observational bounds on these elements. In addition, the presence of energetic ^{3}He can lead to the production of appreciable amounts of ^{6}Li. A summary of work on this problem is shown in Fig. 13. The results of the two groups basically agree and significantly extend the constraints possible from CMB spectral distortions alone. (The fact these calculations do not agree exactly, though, is indicative of the complexity of the calculation and suggests that rechecking the gamma-ray background calculation of Cohen *et al.* [115], which has several elements in common with them, might be a worthwhile exercise given its importance.) The bottom line is that *if* we cannot keep matter and antimatter separated from the time they are created onwards, then the combination of annihilation constraints from BBN, the CMB, and the gamma-ray background tells us that antimatter domains did not constitute a significant fraction ($\gtrsim$ 10%) of the total baryonic matter from the time the Universe had a temperature $\sim$ 1 MeV and onwards. (Note that while Dolgov & Silk-type [82] scenarios might be able to sneak through some of these constraints, such scenarios generally do not claim to have equal amounts of antimatter and matter either.)

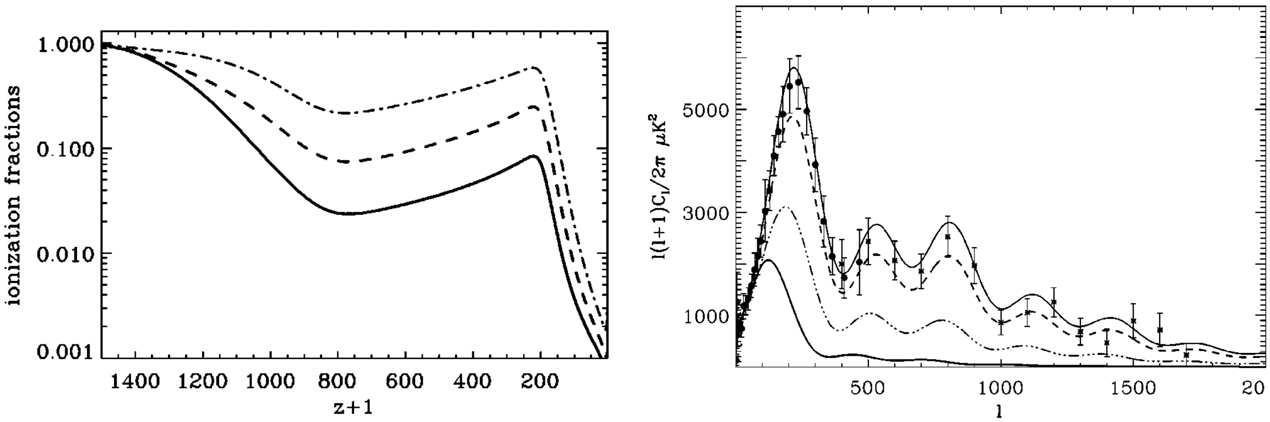

Figure 14: From [78]. The left panel shows the ionization history of the Universe as a function of redshift for three different small antimatter domain models (motivated by [82]). Though small, the domains are assumed to be sufficiently large that the bulk of their annihilation occurs *after* the standard epoch of recombination (ionization fraction=0) which occurs at $z \sim 1000$. The result of the excess energy released by the domain annihilation is to delay recombination until $z \sim 200$. The effect of this delay on the predicted CMB fluctuation spectrum, a strong reduction in the fluctuation amplitudes not unlike that from an early reionization epoch, is shown in the right panel. (In the two panels, curves with the same line style correspond to the same underlying antimatter domain models.)

4.3. Limits on Small Antimatter Domains From Delayed Recombination?

I would like to conclude this section by mentioning one other type of CMB-related constraint, proposed by Naselsky & Chiang [78] (see also [141]), that is relevant to our discussion. If we have a Dolgov & Silk[82] scenario that produces antimatter objects with masses in the range $\sim 10 - 10^7$ solar masses (that are not collapsed and hidden), then such objects will annihilate away and release the bulk of their energy at the epoch of recombination. Not only does this create a CMB spectral distortion of the type discussed above, but another side effect of the annihilation (and cascading) is to keep the Universe ionized for longer than the standard recombination estimate, e.g., see left panel, Fig.14. To zeroth order, this is equivalent to introducing an an early reionization epoch, which damps the the CMB spatial anisotropy signal, right panel Fig. 14, and creates a strong polarization signature. The exact CMB signatures are, of course, highly dependent on the exact model we choose for the antimatter objects, in particular their mass distribution which determines the time profile of annihilation energy release. Following Naselsky & Chiang, if the antimatter objects evaporate (annihilate) at a characteristic redshift z_{ev}, then the WMAP anisotropy and polarization data puts a limit on the fraction of baryonic mass contained in such objects, $f_{abc} \leq 1.6 \times 10^{-5}[(1 + z_{ev})/200]^{3/2}$ [78], i.e., again small. The Planck mission, with improved angular and polarization sensitivity, could improve this limit by another factor 5-10.

5. CONSTRAINTS ON CURRENT BARYOGENESIS MODELS WITH NO ANTIMATTER DOMAINS

“Standard” baryogenesis models avoid all the problems we have been discussing by creating a net baryon excess that is always positive and uniform and thus has no antimatter domains. Given that an early Universe effect is probably responsible (and in particular one that occurs well before the BBN and recombination epochs), one might therefore think that we have exhausted our cosmology tool kit and that it is now safe for theorists to return to their ivory towers. Alas, the observations have improved so much recently that this is no longer quite true. I would like to finish by mentioning three specific examples of this.

5.1. Baryogenesis from the Decay of Massive GUT-Scale Particles

One of the first well-motivated baryogenesis scenarios, e.g., [5, 84, 104], was based on the realization that GUT theories predicted baryon number violation, thereby providing us with a concrete realization of the first of Sakharov's three conditions. If the Universe ever reached a temperature close to $kT_{GUT} \sim M_{GUT}$ ($\sim 10^{16}$ GeV), then massive GUT scale "X" particles (heavy bosons) could be produced whose subsequent B- and CP-violating decays at later times would provide the excess baryon number needed to explain the observed asymmetry. The key to making the model work was to have the right number of X particles around. This could in principle be achieved by using the freeze-out mechanism. If X particles were initially relativistic and in thermal equilibrium, one could then tune their mass and interaction rates to produce the required relic freeze-out density as the Universe expanded and cooled (an example of how the Universe's expansion can be used to satisfy Sakharov's third condition that we somehow break thermal equilibrium). Unfortunately, the scenario ran into two problems. First, as we noted above, baryogenesis had to happen towards or after the end of inflation, ie., presumably after the Universe reheated, and inflation constraints implied that the initial reheat temperature probably was significantly lower than T_{GUT}. Very few X particles would be produced in the first place. Secondly, a universe sufficiently hot to produce X particles could produce other nasty massive particles, e.g., long-lived massive gravitinos if one invoked supersymmetry in the inflationary scenario. If one were not careful, these particles could last until the BBN epoch (or even longer in some theories which explain UHE cosmic ray production).

Because of the cascading process we have discussed, the decay products of these long-lived particles could photo-dissociate ^{4}He and deterium, again ruining the apparently good agreement between the standard BBN calculation and observations. A considerable amount of work went into computing the resulting constraints. (See [142–156]. Some even suggested that a bit of gravitino decay was actually a good thing because it could help reconcile BBN with higher values of the baryon density, $\Omega_b \sim 1$ [157, 158].) Before WMAP, the standard lore from the BBN constraints was that one could not tolerate reheat temperatures much above $10^{10} - 10^{11}$ GeV, well below the GUT scale – and a big problem for GUT baryogenesis models. As summarized in [84], clever ways around the difficulties imposed by this maximum temperature limit have been found, e.g., involving a "pre-heating" where non-linear effects (parametric resonances) can lead to copious production of supra-thermal particles [68]. However, some of the conclusions should be probably be rechecked because with the inclusion of the WMAP η constraint along with improved, the BBN constraints are now much tighter. The latest limit on a reheating temperature quoted in [150], for example, is now $T_{reheat} \lesssim 10^6$ GeV (!) for a gravitino mass ~ 100 GeV. This is low enough to potentially cause problems not just for baryogenesis scenarios but also some inflationary models. In any event, theorists should be aware that gravitino (massive late-decaying particle) constraints from BBN have improved significantly recently, possibly to the point of being annoying again.

5.2. Baryogenesis and Neutrino Degeneracy

A detailed discussion of baryogenesis via leptogenesis or the Affleck-Dine baryogenesis mechanism is beyond the scope of this lecture. (For more details, see the reviews of, e.g,. [53, 84, 104]). All we need to know here is that it can sometimes be useful to have a large lepton number in the Universe and that in some Affleck-Dine scenarios, e.g., [159], a large lepton to baryon number density $n_L/n_B \sim 10^8 - 10^9$ is expected. Since charge conservation requires that the number of electrons be equal to the baryon number ($n_e = n_b$, i.e., small compared to n_γ), the natural place to hide large lepton numbers is in the neutrino sector. An excess number of a certain neutrino species is usually characterized by its "degeneracy" parameter $\xi_\nu = \mu_\nu/kT$ where μ_ν is the chemical potential for that species which appears in the Bose-Einstein distribution for that neutrino, $f_\nu(E) = 1/[1 + \exp(E/kT - \xi_\nu)]$. (See [63, 160] for a more general discussion of the definition and importance of neutrino degeneracies.) One possible way to constrain neutrinos degeneracies has been through BBN, but one of the problems has been the tradeoff that one can make by adjusting the electron neutrino degeneracy (which controls the critical neutron-proton ratio when BBN starts) versus the degeneracies in the τ and μ neutrinos (which affect the expansion rate of the Universe via their energy density and determine how long the BBN epoch lasts). One can play the two degeneracies off each other, for example,

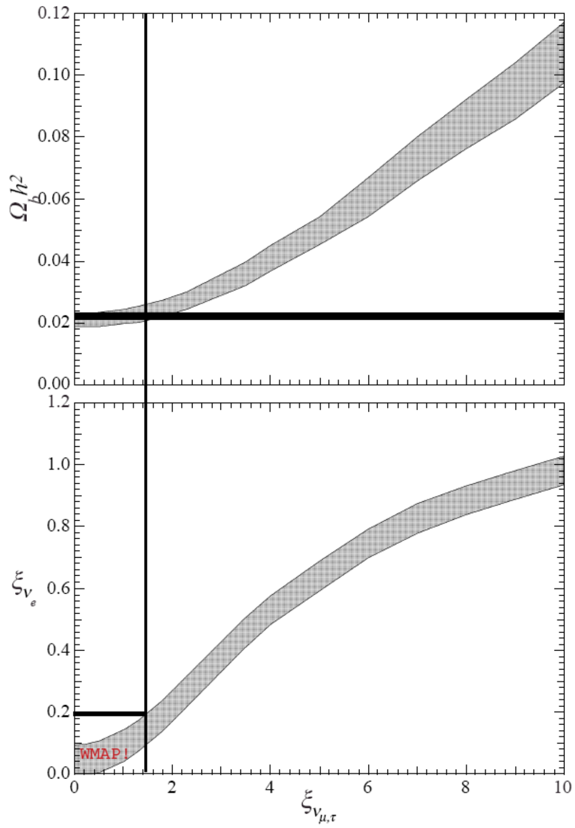

Figure 15: From [161]. An example of some of the degeneracies that enter BBN constraints once non-standard physics is added, in this case neutrino degeneracy. The panels show the allowed values of baryon density ($\Omega_b h^2$) and electron neutrino degeneracy parameter, ξ_{ν_e}, for which the constraints from the observed light-element abundances are satisfied as a function of the μ and τ neutrino degeneracy parameter, $\xi_{\nu_{\mu,\tau}}$. The heavy lines show roughly the constraints imposed by the independent WMAP determination [48] of $\Omega_b h^2$, i.e., $\xi_{\nu_e} \lesssim 0.2$, which removes most of the parameter space originally available.

to significantly change the value of the baryon density, Ω_b, required for BBN to match the observed primordial light element abundances. This is clearly seen in Fig.15 taken from [161]. If we now (naively) add the WMAP Ω_b constraint to these plots, much of the phase space (and freedom for theorists) is gone! Neutrino degeneracy can significantly affect the CMB fluctuation power spectrum (and thus the inferred value of Ω_b), so what I have done is technically wrong and a full joint CMB-BBN likelihood analysis should be carried out, e.g., as in [161]. The results of [161] are now somewhat dated since neutrino oscillations, which are now confirmed, likely imply $\xi_e \approx \xi_\mu \approx \xi_\tau$. More importantly, though, they are pre-WMAP. See [162] for an analysis including WMAP results, with possibly interesting conclusions. (If there are only three neutrino species, then $\xi_e \lesssim 0.1$. If there is in fact a neutrino degeneracy $\xi_e \sim 0.1$, [162] note that BBN *does* allow for the possibility of extra relativistic particle species, e.g., sterile neutrinos, although the range is restricted by the WMAP data.) The bottom line to take away, again, is that the combination of improved CMB and BBN data provide potentially powerful constraints that may impact your work.

5.3. CMB Constraints on Baryogenesis from a Scalar Field Condensate

A popular class of baryogenesis models that currently does not suffer from too many experimental constraints is the so-called Affleck-Dine scenario, e.g., [104]. In its most general version, e.g., see [163], we have a scalar field that carries baryon and lepton number and has a potential with "flat" directions (e.g., in supersymmetry) that can be easily excited, say at the end of inflation. The excited field can form a coherent "condensate" that decays at

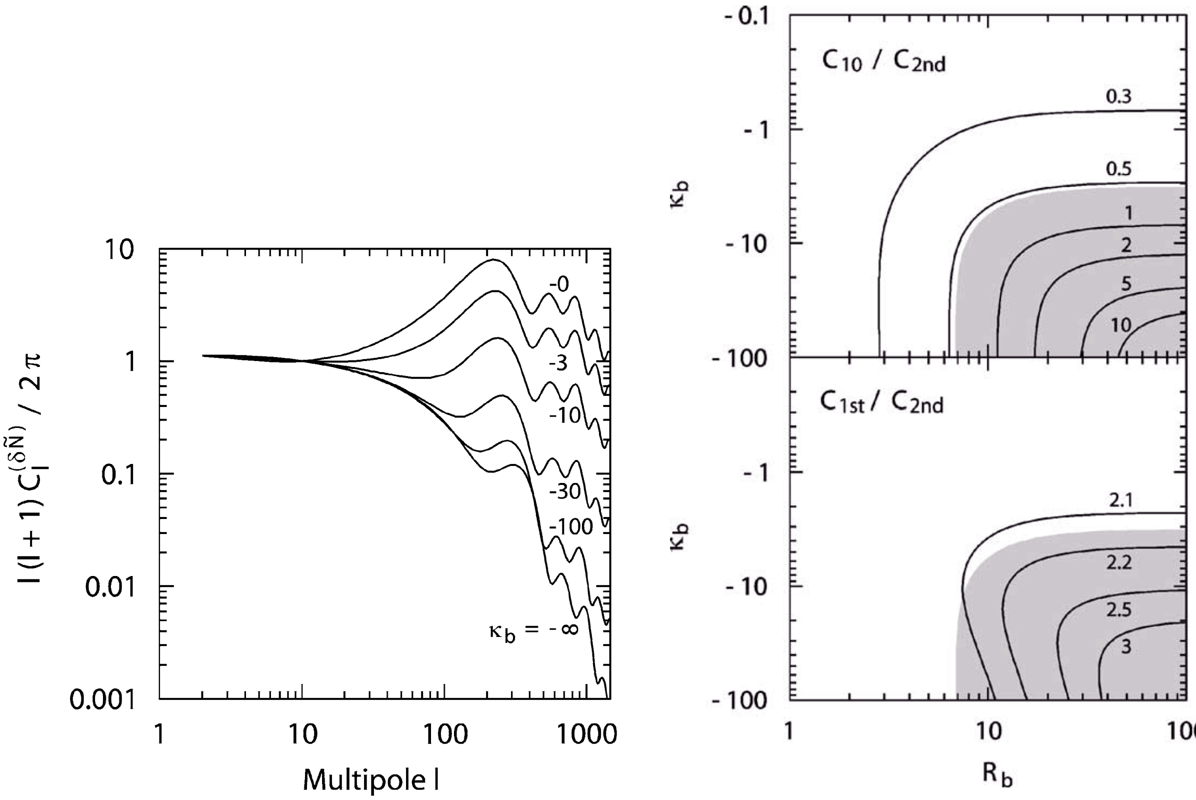

Figure 16: From [163]. Left panel: The angular power spectrum for CMB temperature fluctuations produced by a right-handed sneutrino scalar field (responsible for baryogenesis), where the parameter $\kappa_b = 0$ corresponds to a purely adiabatic perturbation spectrum and $k_b = \infty$ corresponding to a purely isocurvature one. Inflaton and sneutrino fluctuations are assumed to be uncorrelated so the total observed fluctuation spectrum is the sum of the standard adiabatic inflaton one and the one shown here. Right panel: constraints on the parameters κ_b and R_b. The parameter R_b describes the relative importance of the scalar field to inflaton perturbations. The shaded regions are excluded by pre-WMAP data. The WMAP results should significantly tighten these constraints, as will future results from the Planck mission.

later times, efficiently (too efficiently in some cases) producing baryon number. Exactly when the decay occurs is somewhat of a free parameter, but it could be around or even before the time of the decay of the inflaton, the main particle field responsible for inflation. The main point for us is that we now have two important scalar fields to deal with at early times, the inflaton and the condensate field, that are formally *independent*. Normally, we think of one scalar field (the "inflaton") dominating everything, e.g., the dynamics of inflation, but in a realistic particle physics model with several fields, this is not completely obvious (especially given that we don't know what the inflaton is). What could really be happening is that we have a mixed or "hybrid" inflation where the inflaton may dominate at first but other scalar fields, generically called "curvatons" these days, e.g., [164, 165], also become important at some stage. Although not usually discussed in these terms, baryogenesis scenarios involving primordial scalar field condensates are in fact examples of hybrid inflation scenarios.

Now from a practical point of view, why do we care about hybrid inflation? Because it means two (or more) scalar fields could be contributing to the primordial perturbation spectrum coming out of inflation. In particular, if the fields are oscillating and decaying independently, their decay products could be created out of phase – which breaks the key prediction of adiabatic fluctuations coming out of inflation. A scalar field condensate that is solely responsible for our excess baryon number is a clear example of this. Let us say that at early times the decay of the inflation produces most of the energy density of the Universe but it produces *no* net baryon number. To the resulting hot soup (with its inflaton generated adiabatic fluctuations) we now add small perturbations (in terms of the total energy density) due to the decay of the scalar field condensate. Because the condensate scalar field is not the inflaton, there is no guarantee, for example, that regions of high excess baryon density (created predominantly by the condensate decay) will lie in regions of high photon density (created predominantly by the inflaton decay). Going back to our long-winded description of different types of possible primordial perturbations, one sees that I have just described a mechanism for generating nearly isocurvature (isothermal) baryonic perturbations. (Note this is a very different prediction from the GUT particle baryogenesis case, where the GUT particles are produced in the reheating due to the inflaton decay and presumably track the inflaton fluctuations.) In general, what we generate from hybrid inflation (scalar field baryogenesis) are perturbations that can be decomposed as a mixture of adiabatic and isocurvature fluctuation modes, e.g., see [163] for a more general study and [122, 166] which focus on Affleck-Dine scenarios. Fig.16 from [163] shows how the CMB angular power spectrum can be affected depending on exactly how important the scalar condensate is (as measured by their parameter κ_b, with $\kappa_b = 0$ corresponding to a purely adiabatic perturbation spectrum and $k_b = \infty$ corresponding to a purely isocurvature spectrum). Such changes in the CMB power spectrum are already strongly constrained by WMAP (e.g., see [80] and discussion above) and will be further constrained by Planck. In sum, although there is undoubtedly plenty of parameter space remaining to scalar condensate scenarios, some of it has already been closed off and one needs to start paying attention to the CMB constraints.

6. CONCLUSIONS

Because it is 100% efficient at converting rest mass energy into energetic particles, in particular ~ 100 MeV gamma-rays, the annihilation of even small amounts of antimatter with matter provides a powerful signal that we can use to constrain its occurrence. The constraint on the amount of annihilation that may be occurring can in turn be used to constrain the density of antimatter and/or the extent to which matter and antimatter are in contact. For example, if the solar wind and the moon had opposite signs of baryonic charge, the moon would be the brightest gamma-ray source on the sky due to the ensuing annihilation. As it is not, and astronauts have walked on its surface with no ill effects, we can be quite sure the moon and the solar wind are predominantly made of matter. Another location where annihilation would occur extremely quickly, on the timescale of only $\sim$ a few hundred years, is the interstellar medium of our galaxy. The fact that we still find a significant interstellar medium today therefore tells us that it is again dominated by particles of one sign of baryonic charge, presumably matter. The limits on the gamma-ray flux from the interstellar medium in fact tell us that the density of any antimatter that is mixed in with matter must $f \lesssim 10^{-15}$ smaller than that of the matter. Similar reasoning leads to the conclusion that the intracluster gas in nearby galaxy clusters, which exists on spatial scales $\sim$ Mpc, must also be predominantly made of matter.

Do constraints such as these then mean that the total amount of antimatter in our galaxy or a galaxy cluster is very different from the amount of matter? Strictly speaking the answer is no since one may always claim that all antimatter is hidden away in very dense, collapsed objects such as stars or black holes that do not annihilate efficiently with external matter. Such a claim, be it reasonable or not, is hard to absolutely refute empirically. (Note, however, that we can constrain the relative number of antimatter stars in our galaxy to be small given our failure to detect cosmic ray antinuclei from antimatter stars that go supernova as well as our failure to detect the faint annihilation radiation expected from the impact of the interstellar medium on such stars.) What we can be sure of, though, is that matter and antimatter in the nearby universe are distributed in very different ways. This is enough to cause us a serious problem. To the extent that CPT is a good symmetry and that accelerator experiments tell us that baryons and antibaryons are always created or destroyed in pairs and empirically behave in very similar ways up to GeV-TeV energy scales, it is hard to see how such a dramatic spatial asymmetry could have arisen in the late stages of the Big Bang, when the Universe's temperature was less than $\sim$ a GeV. This points us to something that happened earlier in the Universe's history. For those interested only in "low temperature" physics, it is sufficient to treat the observed asymmetry between baryons and antibaryons as an initial condition. However, as we become increasingly sure that inflation has happened (e.g., because of the "concordance cosmology" picture emerging from recent observations), we are driven to the conclusion that the asymmetry ultimately *cannot* be an initial condition.

Rather it must have been dynamically generated by new physics some time from the end of inflation onwards.

Because the physics responsible for this "baryogenesis" is currently unknown, theorists (initially) had few constraints on their imagination. Accordingly there has been considerable speculation as to what spatial distribution of matter and antimatter is actually produced by baryogenesis, e.g., see the discussion on baryogenesis and large-scale structure in [87]. Note that once we allow for baryon number and CP violation (Sakharov's first two conditions), we do not have to invoke any mysterious spatial transport mechanisms to pull matter and antimatter apart. For example, by using a variant of the symmetry breaking mechanism discussed in [83] and doing some fine-tuning with regards to inflation, we can envision dynamically creating domains, e.g., after a phase transition, that have quite different magnitudes and sign of total baryonic charge, i.e., induce significant spatial variations in $\eta = (n_b - n_{\bar{b}})/n_\gamma$. As long as our local patch of the Universe could be fit inside one of these domains, considerable freedom remained for theorists because we could not say much about the characteristics of other possible domains and, thus, the outcome of a specific baryogenesis scenario. For example, the comparison of Big Bang Nucleosynthesis predictions with observed light element abundances provides a strong constraint on η. In the early days, however, most of the abundance determinations came from studying stars and gas near our galaxy, so that the constraint on η was really a local one. This situation has changed dramatically now that we have measured primordial deuterium abundances in very distant (redshift $z \sim 2$) clouds scattered across the sky, and the WMAP satellite has provided an independent, sky-averaged estimate for the value of η during the epoch of recombination. The various deuterium measurements agree with each other to within plausible observational errors, and the mean value of η deduced from these measurements agrees with the one obtained by WMAP. This tells us nothing dramatic happened between the epochs of nucleosynthesis and recombination and makes it much harder to accomodate significant spatial variations in η.

Besides the fact that we have a clear baryon asymmetry in our local universe (be it in the relative numbers of baryons vs. antibaryons or be it simply in their spatial distributions), the main message of this lecture is that improved observations of the type just discussed mean that the epoch of unbridled freedom for baryogenesis models is finally ending. Our ability to look beyond our local environment and study the universe on large-scales and earlier times has grown dramatically in the last few years. For example, we can now see individual galaxies out to redshifts ~ 7, close to the edge of the observable Universe, and we now have surveys of millions of galaxies that allow us to accurately characterize the spatial distribution of visible baryons on $\sim$ Mpc to ~ 100 Mpc scales. With the arrival of ground-based and space-based instruments that can study the CMB with high precision, we can do even better and effectively image the entire Universe at the epoch of recombination ($z \sim 1000$) to detect and study the fluctuations that grew into the large-scale structure we see today. The results of these improved observations imply, in particular, that the Universe looks quite homogeneous on large scales. To date, we have no evidence for the discontinuities one might expect if our universe were divided up into domains with differing baryonic content. Every part of the sky we have looked at generally seems to have had the same evolutionary history, and as noted, one universal value of η so far seems consistent with what we see. The homogeneity of the Universe is perhaps most clearly demonstrated by the CMB. When one goes back far enough in time that the density perturbations responsible for today's large-scale structure should be small, i.e., in the linear regime, the Universe is indeed incredibly uniform, to a part in a few times 10^5 at the epoch of recombination.

This uniformity of the CMB has a very important implication for baryogenesis scenarios that predict the existence of matter and antimatter domains, especially if one couples it to the fact that that the primordial light element abundances agree (modulo likely systematic errors) with the standard BBN ones computed using the average CMB value for η. While today we may be able to slice up the Universe in such a way as to have density voids between most domains, thereby preventing annihilation, we generally cannot do so in the past. The minute we admit that matter and antimatter must have been in contact, the "unavoidable" annihilation that occurs, e.g., [115], enables us to place strong constraints on the domains. Too much annihilation between domains will cause problems such as predicting a diffuse gamma-ray background that is too high, causing excessive distortions of the CMB spectrum, washing out CMB fluctuations due to a delayed epoch of reionization, or causing the BBN light element abundance predictions to disagree with observations. Except for the escape route of putting antimatter into small-scale domains that effectively decouple early on from the rest of the Universe, e.g., by forming black holes or dark matter[82], the general conclusion is that if matter and antimatter domains exist, their characteristic size scale must be a significant fraction of the size of the observable Universe or else antibaryons cannot be a significant fraction ($\gtrsim$ 10%) of the total number of baryons in the Universe from the time of Big Bang nucleosynthesis onwards. The simplest and cleanest hypothesis consistent with this conclusion is that our Universe contains precisely one domain that is matter dominated and has no antimatter left today.

The overall result from our recent observational breakthroughs has been the development of a "concordance cosmology" which, like the Standard Model of physics, has well-measured parameters and can satisfactorily explain many observations at the same time. Partly because it is based on the Standard Model, however, this concordance cosmology also does not provide an explicit recipe for baryogenesis and it thus fails to predict our existence. Whatever we do to rectify this situation must be accordingly subtle so as not to upset the significant successes of the Standard Model and the concordance cosmology. One non-trivial example of such a success is the prediction of the primordial light element abundances. This prediction is generally known to be quite sensitive to the presence of non-standard physics during the nucleosynthesis epoch, e.g., the decay of relic gravitinos. The agreement of the concordance cosmology prediction with observations thus provides strong constraints on such physics. What may not be completely appreciated, however, is that the independent determination of the baryon density by the CMB experiments effectively removes the main degree of freedom in the BBN calculation that can be played with to enlarge the allowed phase space for one's favorite non-standard physics. BBN-derived constraints therefore tend to be much tighter now than they were in the pre-WMAP era. Whatever non-standard things baryogenesis does or invokes should thus be finished or well-hidden by the epoch of nucleosynthesis ($T \sim 1$ MeV). In a similar vein, the CMB fluctuations observed by WMAP are consistent with being predominantly due to the adiabatic, Gaussian-random perturbations expected from standard inflationary scenarios that involve one scalar field. The degree to which this is in fact the case will be explored further by next-generation CMB experiments like Planck. Baryogenesis scenarios which invoke a different scalar field may therefore end up in conflict with observations if the baryon density perturbations produced by that scalar field are not sufficiently well-correlated with the fluctuations produced by the scalar field responsible for inflation.

Acknowledgments

I would like to thank the Max-Planck-Institut für Kernphysik for its generous hospitality during my sabbatical stay when this lecture was written. I would also like to thank Yale University for providing a senior faculty fellowship which made my sabbatical possible. Finally, I would like to thank the organizers for organizing a highly stimulating lecture series and for their infinite patience in waiting for this manuscript.

References

[1] P.A.M. Dirac, Proceedings of the Royal Society **117A**, 610 (1928).

[2] http://athena.web.cern.ch/athena/

[3] A.D. Sakharov, JETP Letters **91B**, 24 (1967).

[4] G. Steigman, Annual Reviews of Astronomy and Astrophysics **14**, 339 (1976).

[5] E.W. Kolb and M.S. Turner, Annual Reviews of Nuclear and Particle Science **33**, 645 (1983).

[6] D.J. Thompson, D.L. Bertsch, D.J. Morris, R. Mukherjee, Bulletin of the American Astronomical Society **28**, 1306 (1996).

[7] I.V. Moskalenko and A.W. Strong, Astrophyical Journal **493**, 694 (1998).

[8] C. Alcock et al., Astrophysical Journal **550**, 169 (2001).

[9] W. Gliese and H. Jahreiss, *Preliminary Version of the Third Catalogue of Nearby Stars*, Astronomisches Rechen-Institut, Heidelberg (1991).

[10] P. Sreekumar et al., Astrophysical Journal **426**, 105 (1994).

[11] K.M. Belotsky, Yu. A. Golubkov, M. Yu. Khlopov, R.V. Konoplich, and A.S. Sakharov, Physics of Atomic Nuclei **63**, 233 (2000).
[12] P.K. Suh, Astronomy and Astrophysics **15**, 206 (1971).
[13] S. Orito et al., Physical Review Letters **84**, 1078 (2000).
[14] M. Aguilar et al., Physics Reports **366**, 331 (2002).
[15] R. Hagedorn, Nuclear Physics B **24**, 93 (1970).
[16] P. Chardonnet, J. Orloff, and P. Salati, Physics Letters B **409**, 313 (1997).
[17] From AMS-02 presentation by F. Rapauch, http://www.accms04.physik.rwth-aachen.de/~rapauch/talks/rbtalk.pdf, (2004).
[18] http://ams.cern.ch, http://cyclo.mit.edu/~bmonreal/
[19] S.P. Ahlen, P.B. Price, M.H. Salamon, and G. Tarlé, Astrophysical Journal **260**, 20 (1982).
[20] P. Sreekumar et al., Physical Review Letters **70**, 127 (1993).
[21] E.A. Baltz, J. Edsjö, K. Freeze, P. Gondolo, Physical Review D **65**, 063511 (2002).
[22] E.A. Baltz and P. Gondolo, Physical Review D **67**, 063503 (2003).
[23] L. Bergström, P. Ullio, and J.H. Buckley, Astroparticle Physics **9**, 137 (1998).
[24] E.A. Baltz and L. Bergström, Physical Review D **67**, 043516 (2003).
[25] F. Donato, N. Fornengo, and P. Salati, Physical Review D **62**, 043003 (2000).
[26] S.W. Barwick et al., Astrophysical Journal Letters **482**, 191 (197).
[27] S. Coutu and the HEAT Collaboration, in Proceedings of the 27th ICRC, (2001).
[28] A.W. Strong, I.V. Moskalenko, and O. Reimer, Astrophysical Journal **537**, 763 (2000).
[29] J. Silk and M. Srednick, Physical Review Letters **53**, 624 (1984).
[30] S. Rudaz and F. Stecker, Astrophysical Journal **325**, 16 (1988).
[31] J. Ellis, R.A. Flores, K. Freeze, S. Ritz, D. Seckel, and J. Silk, Physics Letters B **214**, 403 (1989).
[32] F. Stecker and A. Tylka, Astrophysical Journal Letters **336**, L51 (1989).
[33] D. Eichler, Physical Review Letters **63**, 2440 (1989).
[34] M. Kamionkowski and M.S. Turner, Physical Review D **43**, 1774 (1991).
[35] M.S. Turner and F. Wilczek, Physical Review D **42**, 1001 (1990).
[36] A.J. Tylka, Physical Review Letters **63**, 840 (1989).
[37] P. Ullio and L. Bergström, Physical Review D **57**, 1962 (1998).
[38] L. Bergström and P. Ullio, Nuclear Physics B **504**, 27 (1997).
[39] L. Bergström, J. Edsjö, P. Gondolo, and P. Ullio, Physical Review D **59**, 043506 (1999).
[40] P. Gondolo and J. Silk, Physical Review Letters **83**, 1719 (1999).
[41] P. Jean et al., Astronomy and Astrophysics **407**, 55 (2003)
[42] D.D. Dixon, D.H. Hartmann, E.D. Kolaczyk, J. Samimi, R. Diehl, G. Kanbach, H. Mayer-Hasselwander, and A.W. Strong, New Astronomy **3**, 539 (1998)
[43] U. Keshet, E. Waxman, and A. Loeb, Journal of Cosmology and Astroparticle Physics **617**, 281 (2004).
[44] A.W. Strong, I.V. Moskalenko, and O. Reimer, Astrophysical Journal **613**, 956 (2004).
[45] K.-T. Kim, P.P. Kronberg, P.C. Tribble, Astrophysical Journal **379**, 80 (1991).
[46] T.E. Clarke, P.P. Kronberg, H. Böhringer, Astrophysical Journal Letters **547**, 111 (2001).
[47] Y. Rephaeli and D. Gruber, Astrophysical Journal **606**, 82 (2004).
[48] D.N. Spergel et al., Astrophysical Journal Supplement **148**, 175 (2003).
[49] E.W. Kolb and M.S. Turner, *The Early Universe*, Addison-Wesley Publishing Company, Reading (1993).
[50] P.J.E. Peebles, *Principles of Physical Cosmology*, Princeton University Press, Princeton (1993).
[51] G. Börner, *The Early Universe: Facts and Fiction*, Springer, Heidelberg (2003).
[52] T. Padmanabhan, 1993, *Structure Formation in the Universe*, Cambridge University Press, Cambridge (1993).
[53] W. Bernreuther, hep-ph/0205279 (2002).
[54] M. Bartelmann, http://www.ita.uni-heidelberg.de/~msb/Lectures/thrmHst1.pdf
[55] H.W. Rix, http://www.mpia-hd.mpg.de/homes/rix/BBN_Lect.pdf

[56] R.V. Wagoner, W.A. Fowler, and F. Hoyle, Astrophysical Journal **148**, 3 (1967).
[57] S. Sarkar, Reports on Progress in Physics **59**, 1493 (1996).
[58] D.N. Schramm and M.S. Turner, Reviews of Modern Physics **70**, 303 (1998).
[59] G. Gamow, Nature **162**, 680 (1948).
[60] R.A. Alpher and R.C. Herman, Nature **162**, 774 (1948).
[61] R.A. Alpher and R.C. Herman, Physical Review **75**, 1089 (1949).
[62] R.H. Dicke, P.J. Peebles, P.G. Roll, and D.T. Wilkinson, Astrophysical Journal **142**, 414 (1965).
[63] R.A. Malaney and G.J. Mathews, Physics Reports **229**, 4 (1993).
[64] K. Jedamzik, http://www.mpa-garching.mpg.de/~jedamzik/papers/non-sbbn/non-sbbn.html
[65] K. Jedamzik and J.B. Rehm, Physical Review D **64**, 023510 (2001).
[66] J. Feng, http://www.hep.ps.uci.edu/~jlf/research/presentations/0405aps.pdf (2004).
[67] G. Steigman, Annual Reviews of Nuclear and Particle Science B **29**, 313.
[68] E.W. Kolb, A.D. Linde, and A. Riotto, Physical Review Letters **77**, 4290 (1996).
[69] M. White, D. Scott, J. Silk, Annual Reviews of Astronomy and Astrophysics **32**, 319 (1994).
[70] W. Hu and S. Dodelson, Annual Reviews of Astronomy and Astrophysics **40**, 171 (2002).
[71] D.J. Fixsen, E.S. Cheng, J.M. Gales, J.C. Mather, R.A. Shafer, and E.L. Wright, Astrophysical Journal **473**, 576 (1996).
[72] R.A. Sunyaev and Ya. B. Zeldovich, Astrophysics and Space Science **7**, 20 (1970).
[73] Ya. B. Zeldovich and R.A. Sunyaev, Astrophysics and Space Science **4**, 301 (1969).
[74] W.H. Kinney, E.W. Kolb, and M.S. Turner, Physical Review Letters **79**, 2620 (1997).
[75] P.S. Coppi and F.A. Aharonian, Astrophysical Journal Letters **487**, L9 (1997).
[76] S. Lee, Physical Review D **58**, 043004 (1998).
[77] D.V. Semikoz and G. Sigl, Journal of Cosmology and Astro-Particle Physics **4**, 3 (2004).
[78] P.D. Naselsky and L.-Y. Chiang, Physical Review D **69**, 123518 (2004).
[79] R.A. Knop et al., Astrophysical Journal **598**, 102 (2003).
[80] H.V. Peiris et al., Astrophysical Journal Supplement **148**, 213 (2003).
[81] G. Efstathiou, R.S. Ellis, and B.A. Peterson, Monthly Notices of the Royal Astronomical Society **232**, 431 (1988).
[82] A. Dolgov and J. Silk, Physical Review D **47**, 4244 (1993).
[83] A.G. Cohen and D.B. Kaplan, Physics Letters B **199**, 251 (1987).
[84] A. Riotto and Trodden, M., Annual Reviews of Nuclear and Particle Science **49**, 35 (1999).
[85] D. Toussaint, S.B. Treiman, and F. Wilczek, Physical Review D **19**, 4 (1979).
[86] E.W. Kolb and S.Wolfram, Nuclear Physics B **172**, 224 (1980).
[87] A.D. Dolgov, Physics Reports **222**, 6 (1992).
[88] E.R. Harrison, Physical Review Letters **18**, 1011 (1967)
[89] E.R. Harrison, Physical Review **167**, 1170 (1968).
[90] R. Omnès, Physical Review Letters **23**, 38 (1969).
[91] R. Omnès, Astronomy and Astrophysics **10**, 228 (1971)
[92] R. Omnès, Astronomy and Astrophysics **11**, 450 (1971)
[93] R. Omnès, Astronomy and Astrophysics **15**, 275 (1971)
[94] R. Omnès, Physics Reports C **3**, 1 (1972).
[95] J.J. Aly, S. Caser, R. Omnès, J.L. Puget, and G. Valladas, Astronomy and Astrophysics **35**, 271 (1974).
[96] R. Aldrovandi, S. Caser, R. Omnès, and J.L. Puget, Astronomy and Astrophysics **28**, 253 (1973).
[97] F.W. Stecker and J.L. Puget, Astrophysical Journal **178**, 57 (1972).
[98] F.W. Stecker, D.L. Morgan, and J. Bredekamp, Physical Review Letters **27**, 1469 (1971).
[99] B.T. Jones and G. Steigman, Monthly Notices of the Royal Astronomical Society **183**, 585 (1978).
[100] R.A. Sunyaev and Ya.B. Zeldovich, Astrophysics and Space Science **9**, 368 (1970).
[101] A. Ramani and J.L. Puget, Astronomy and Astrophysics **51**, 411 (1976).

[102] F.W. Stecker, Nuclear Physics B **252**, 25 (1985).
[103] A.K. Mohanty and F.W. Stecker, Physics Letters B **143**, 351 (1984).
[104] M. Dine and A. Kusenko, Reviews of Modern Physics **76**, 1 (2004).
[105] A.D. Dolgov, Nuclear Physics B (Proceedings Supplement) **35**, 28 (1994).
[106] A. Dolgov, Nuclear Physics B (Proceedings Supplement) **95**, 42 (2001).
[107] A. Dolgov, Nuclear Physics B (Proceedings Supplement) **113**, 42 (2002).
[108] D. Kirkman, D. Tytler, N. Suzuki, J.M. O'Meara, and D. Lubin, Astrophysical Journal Supplement **149**, 1 (2003).
[109] R.J Bouwens et al., Astrophysical Journal Letters **616**, 79 (2004).
[110] X. Fan et al., Astronomical Journal **128**, 515 (2004).
[111] S. Matsuura, A.D. Dolgov, and S. Nagataki, Progress of Theoretical Physics **112**, 971 (2004).
[112] V. Bromm, P. Coppi, and R. Larson, Astrophysical Journal **543**, 1 (2002).
[113] F. Combes, O. Fassi-Fehri, and B. Leroy, Astrophysics and Space Science **37**, 151 (1975).
[114] J.J. Aly, Astronomy and Astrophysics **64**, 273 (1978).
[115] A.G. Cohen, A. De Rújula, and S.L. Glashow, Astrophysical Journal **495**, 539 (1998).
[116] M. Bucher, K. Moodley, and N. Turok, Physical Review D **62**, 083508 (2000).
[117] N. Sugiyama, S. Zaroubi, and J. Silk, Monthly Notices of the Royal Astronomical Society **354** 543.
[118] J. Silk, Astrophysical Journal **151**, 459 (1968).
[119] P.J. Peebles, Astrophysical Journal Letters **315**, 317 (1987).
[120] W. Hu, E.F. Bunn, and N. Sugiyama, Astrophysical Journal Letters **447**, 59 (1995).
[121] M. Bucher et al., Physical Review Letters **93** 081301 (2004).
[122] M.Yu. Khlopov, S.G. Rubin, A.S. Sakharov, Physical Review D **62**, 083505 (2000).
[123] Y.-T. Gao, F.W. Stecker, M. Gleiser, and D.B. Cline, Astrophysical Journal Letters **361**, 37 (1990).
[124] A. Coc, E. Vangioni-Flam, P. Descouvemont, A. Adahchou, and C. Angulo, Astrophysical Journal **600** 544 (2004).
[125] R.H. Cyburt, B.D. Fields, K.A. Olive, Physics Letters B **567**, 227 (2003).
[126] K. Jedamzik, Physical Review D **70**, 063524 (2004).
[127] Applegate, J.H., Hogan, C.J., and Sherrer, R.J., Physical Review D **35**, 1151 (1987).
[128] K.E. Sale and G.J. Mathews, Astrophysical Journal Letters **309**, 1 (1986).
[129] C. Alcock, G.M. Fuller, and G.J. Mathews, Astrophysical Journal **320**, 439 (1987).
[130] K. Jedamzik and G.M. Fuller, Astrophysical Journal **423**, 33 (1994).
[131] K. Jedamzik and G.M. Fuller, Astrophysical Journal **452**, 33 (1995).
[132] C.J. Copi, K.A. Olive, and D.N. Schramm, Astrophysical Journal **451**, 51 (1995).
[133] M. Giovannini and M.E. Shaposhnikov, Physical Review Letters **80**, 22 (1998).
[134] M. Giovannini and M.E. Shaposhnikov, Physical Review D **57**, 2168 (1998).
[135] J. B. Rehm and K. Jedamzik, Physical Review Letters **81**, 3307 (1998).
[136] J.B. Rehm and K. Jedamzik, Physical Review D **63**, 043509 (2001).
[137] J.B. Rehm, Thesis, http://edoc.ub.uni-muenchen.de/archive/00000420/01/Rehm_Jan.pdf (2000)
[138] H. Kurki-Suonio and E. Sihvola, Physical Review D **63**, 3508 (2000).
[139] H. Kurki-Suonio and E. Sihvola, Physical Review Letters **84**, 3756 (2000).
[140] E. Sihvola, Physical Review D **63**, 3001 (2001).
[141] Naselky, P. and I. Novikov, Monthly Notices of the Royal Astronomical Society **334**, 137 (2002).
[142] Yu. B. Zeldovich, A. A. Starobinskii, M. Yu. Khlopov, and V. M. Chechetkin, Soviet Astronomy Letters **3**, 110 (1977).
[143] B. V. Vainer, O. V. Dryzhakova, and P.D. Naselskii, Soviet Astronomy Letters **4**, 4 (1978).
[144] V.M. Chechetkin, M. Yu. Khlopov, M.G. Sapozhnikov, and Y. B. Zeldovich, Physics Letters B **118**, 329 (1982); V.M. Chechetkin, M. Yu. Khlopv, and M.G. Shapozhnikov, Rivista del Nuovo Cimento **5**, 1 (1982).
[145] Yu. A. Batusov et al., Lettere del Nuovo Cimento della Societa Italiana di Fisica **41**, 223 (1984).

[146] M. Yu. Khlopov and A. D. Linde, Phys. Lett. 138B, 265
[147] R. Dominguez-Tenreiro, Astrophysical Journal **313**, 523 (1987).
[148] I. Halm, Physics Letters B **188**, 403 (1987).
[149] R.J. Protheroe, T. Stanev, V.S. Berezinsky, Physical Review D **51**, 4134 (1995).
[150] R.H. Cyburt, J. Ellis, B.D. Fields, and K.A. Olive, Physical Review D **67**, 103521 (2003).
[151] M. Kawasaki, K. Kohri, T. Moroi, Physical Review D **63** 103502 (2001).
[152] M. Kawasaki and T. Moroi, Astrophysical Journal **452**, 506 (1995).
[153] D. Lindley, Monthly Notices of the Royal Astronomical Society **193**, 593 (1980).
[154] J. Ellis, G.B. Gelmini, J.P. Lopez, D.V. Nanopoulos, and S. Sarkar, Nuclear Physics B **373**, 399 (1992).
[155] V.S. Berezinsky, Nuclear Physics B **380**, 478 (1992).
[156] J. Ellis, D.V. Nanopoulos, and S. Sarkar, Nuclear Physics B **529**, 478 (1985).
[157] R. Dominguez-Tenreiro and G. Yepes, Astrophysical Journal Letters **317**, L1 (1987).
[158] G. Yepes and R. Dominguez-Tenreiro, Astrophysical Journal 335, 3 (1988).
[159] J. McDonald, Physical Review Letters **84**, 4798 (2000).
[160] A.D. Dolgov, Physics Reports **370**, 333 (2002).
[161] M. Orito, T. Kajino, G.J. Mathews, Y. Wang, Physical Review D **65**, 123504 (2002).
[162] V. Barger, J.P. Kneller, P. Langacker, D. Marfatia, G. Steigman, Physics Letters B **569**, 123 (2003).
[163] T. Moroi and H. Murayama, Physics Letters B **553**, 126 (2003).
[164] L. Amendola, C. Gordon, D. Wands, M. Sasaki, Physical Review Letters **88**, 1302 (2002).
[165] G. Lazarides, R. Ruiz de Austri, and R. Trotta, Physical Review D **70**, 123527 (2004).
[166] K. Enqvist and J. McDonald, Physical Review D **62**, 043502 (2000).