Fermi National Accelerator Laboratory

FERMILAB Conf-84/60-A
June 1984

COSMOLOGY AND PARTICLE PHYSICS*

Michael S. Turner

Theoretical Astrophysics Group
Fermi National Accelerator Laboratory
Batavia, Illinois 60510

and

Enrico Fermi Institute
The University of Chicago
Chicago, Illinois 60637

*Lectures given at the NATO Advanced Study Institute on Quarks and Leptons, at the Max Planck Institute, Munich, W. Germany, September 1983 (to be published in the NATO ASI Series).

COSMOLOGY AND PARTICLE PHYSICS

Michael S. Turner

Theoretical Astrophysics Group
Fermi National Accelerator Laboratory
Batavia, Illinois 60510

and

Enrico Fermi Institute
The University of Chicago
Chicago, Illinois 60637

INTRODUCTION

In the past five years or so progress in both elementary particle physics and in cosmology has become increasingly dependent upon the interplay between the two disciplines. On the particle physics side, the $SU(3)_C \times SU(2)_L \times U(1)_Y$ model seems to very accurately describe the interactions of quarks and leptons at energies below, say, 10^3 GeV. At the very least, the so-called standard model is a satisfactory, effective low energy theory. The frontiers of particle physics now involve energies of much greater than 10^3 GeV--energies which are not now available in terrestrial accelerators, nor are ever likely to be available in terrestrial accelerators. For this reason particle physicists have turned both to the early Universe with its essentially unlimited energy budget (up to 10^{19} GeV) and high particle fluxes (up to 10^{107} cm^{-2} s^{-1}), and to various unique, contemporary astrophysical environments (centers of main sequence stars where temperatures reach 10^8 K, neutron stars where densities reach 10^{14}-10^{15} g cm^{-3}, our galaxy whose magnetic field can impart 10^{11} GeV to a Dirac magnetic charge, etc.) as non-traditional laboratories for studying physics at very high energies and very short distances.

On the cosmological side, the hot big bang model, the so called standard model of cosmology, seems to provide an accurate accounting of the history of the Universe from about 10^{-2} s after 'the bang' when the temperature was about 10 MeV, until today, some 10-20 billion years after 'the bang' and temperature of about 3 K ($\simeq 3 \times 10^{-13}$ GeV). Extending our understanding further back, to earlier times and higher temperatures, requires knowledge about the fundamental particles (presumably quarks and leptons) and their interactions at very high energies. For this reason, progress in cosmology has become linked to progress in elementary particle physics.

In these 4 lectures I will try to illustrate the two-way nature of the interplay between these fields by focusing on a few selected topics. In Lecture 1 I will review the standard cosmology, especially concentrating on primordial nucleosynthesis, and discuss how the standard cosmology has been used to place constraints on the properties of various particles. Grand Unification makes two striking predictions: (1) B non-conservation; (2) the existence of stable, superheavy magnetic monopoles. Both have had great cosmological impact. In Lecture 2 I will discuss baryogenesis, the very attractive scenario in which the B, C, CP violating interactions in GUTs provide a dynamical explanation for the predominance of matter over antimatter, and the present baryon-to-photon ratio. Baryogenesis is so cosmologically attractive, that in the absence of observed proton decay it has been called 'the best evidence for some kind of unification.' Monopoles are a cosmological disaster, and an astrophysicist's delight. In Lecture 3 I will discuss monopoles, cosmology, and astrophysics. To date, the most important 'cosmological payoff' of the Inner Space/Outer Space connection is the inflationary Universe scenario. In the final lecture I will discuss how a very early ($t \lesssim 10^{-34}$ sec), first-order phase transition associated with spontaneous symmetry breaking (SSB) has the potential to explain a handful of very fundamental cosmological facts--which can be accommodated by the standard cosmology, but are not elucidated by it. By selecting just a few topics I have left out some other very important and exciting ones--e.g., galaxy formation and the role of exotic debris from the early Universe (massive neutrinos, axions, other-inos, strings to mention a few types of interesting debris), supersymmetry/supergravity/Kaluza-Klein models and cosmology, and axions, astrophysics, and cosmology. I refer the interested reader to references 1-3.

LECTURE 1 -- THE STANDARD COSMOLOGY

The hot big bang model nicely accounts for the universal (Hubble) expansion, the 2.7 K cosmic microwave background radiation, and through primordial nucleosynthesis, the abundances of D, ^{4}He and perhaps also ^{3}He and ^{7}Li. Light received from the most distant objects observed (QSOs at redshifts $\simeq$ 3.5) left these objects when the Universe was only a few billion years old, and so observations of QSOs allow us to directly probe the history of the Universe to within a few billion years of 'the bang'. The surface of last scattering for the microwave background is the Universe about 100,000 yrs after the bang when the temperature was about 1/3 eV. The microwave background is a fossil record of the Universe at that very early epoch. In the standard cosmology an epoch of nucleosynthesis takes place from $t \simeq 10^{-2}$ s - 10^2 s when the temperature was $\simeq$ 10 MeV - 0.1 MeV. The light elements synthesized, primarily D, ^{3}He, ^{4}He, and ^{7}Li, are relics from this early epoch, and comparing their predicted big bang abundances with their inferred primordial abundances is the most stringent test of the standard cosmology we have at present. [Note that I must say inferred primordial abundance because contemporary astrophysical processes can affect the abundance of these light isotopes, e.g., stars very efficiently burn D, and produce ^4He.] At present the standard cosmology passes this test with flying colors (as we shall see shortly).

On the large scale (>> 100 Mpc), the Universe is isotropic and homogenous, and so it can accurately be described by the Robertson-Walker line element

$$ds^2 = -dt^2 + R(t)^2[dr^2/(1-kr^2) + r^2\, d\theta^2 + r^2 \sin^2\theta d\phi^2], \quad (1.1)$$

where ds^2 is the proper separation between two spacetime events, k = 1, 0, or -1 is the curvature signature, and R(t) is the cosmic scale factor. The expansion of the Universe is embodied in R(t)--as R(t) increases all proper (i.e., measured by meter sticks) distances scale with R(t), e.g., the distance between two galaxies comoving with the expansion (i.e., fixed r, θ, φ), or the wavelength of a freely-propagating photon ($\lambda \propto R(t)$). The k > 0 spacetime has positive spatial curvature and is finite in extent; the k < 0 spacetime has negative spatial curvature and is infinite in extent; the k = 0 spacetime is spatially flat and is also infinite in extent.

The evolution of the cosmic scale factor is determined by the Friedmann equations:

$$H^2 \equiv (\dot{R}/R)^2 = 8\pi G\rho/3 - k/R^2, \quad (1.2)$$

$$d(\rho R^3) = -p\, d(R^3), \quad (1.3)$$

where ρ is the total energy density and p is the pressure. The expansion rate H (also called the Hubble parameter) sets the characteristic time for the growth of R(t); H^{-1} ≃ e-folding time for R. The present value of H is 100 h $kms^{-1}\ Mpc^{-1} \simeq h\ (10^{10}\ yr)^{-1}$; the observational data strongly suggest that $1 \geq h \geq 1/2$ (ref. 4). As it is apparent from Eqn. 1.2 model Universes with $k \leq 0$ expand forever, while a model Universe with k > 0 must eventually recollapse. The sign of k (and hence the geometry of spacetime) can be determined from measurements of ρ and H:

$$k/H^2R^2 = \rho/(3H^2/8\pi G) - 1, \quad (1.4)$$

$$\equiv \Omega - 1,$$

where $\Omega = \rho/\rho_{crit}$ and $\rho_{crit} = 3H^2/8\pi G \simeq 1.88\ h^2 \times 10^{-29}\ gcm^{-3}$. The cosmic surveying required to directly determine ρ is far beyond our capabilities (i.e., weigh a cube of cosmic material 10^{25} cm on a side!). However, based upon the amount of luminous matter (i.e., baryons in stars) we can set a lower limit to Ω: $\Omega \geq \Omega_{lum} \simeq 0.01$. The best upper limit to Ω follows by considering the age of the Universe:

$$t_U = 10^{10}\ yr\ h^{-1}\ f(\Omega), \quad (1.5)$$

where $f(\Omega) \leq 1$ and is monotonically decreasing (e.g., f(0) = 1 and f(1) = 2/3). The ages of the oldest stars (in globular clusters) strongly suggest that $t_U \geq 10^{10}$ yr; combining this with Eqn. 1.5 implies that:

$\Omega f^2(\Omega) \geq \Omega h^2$. The function Ωf^2 is monotonically increasing and asymptotically approaches $(\pi/2)^2$, implying that independent of h, $\Omega h^2 \lesssim 2.5$. Restricting h to the interval (1/2, 1) it follows that: $\Omega h^2 \lesssim 0.8$ and $\Omega \lesssim 3.2$.

The energy density contributed by nonrelativistic matter varies as $R(t)^{-3}$--due to the fact that the number density of particles is diluted by the increase in the proper (or physical) volume of the Universe as it expands. For relativistic particles the energy density varies as $R(t)^{-4}$, the extra factor of R due to the redshifting of the particle's momentum (recall $\lambda \propto R(t)$). The energy density contributed by a relativistic species (T >> m) at temperature T is

$$\rho = g_{eff}\pi^2 T^4/30, \tag{1.6}$$

where g_{eff} is the number of degrees of freedom for a bosonic species, and 7/8 that number for a fermionic species. Note that $T \propto R(t)^{-1}$. Here and throughout I have taken $\hbar = c = k_B = 1$, so that 1 GeV = $(1.97 \times 10^{-14}$ cm$)^{-1}$ = $(1.16 \times 10^{13}$ K$)$ = $(6.57 \times 10^{-25}$ s$)^{-1}$, $G = m_{pl}^{-2}$ ($m_{pl} = 1.22 \times 10^{19}$ GeV), and 1 GeV^4 = 2.32×10^{17} g cm^{-3}. By the way, 1 light year $\simeq 10^{18}$ cm; 1 pc $\simeq$ 3 light year; and 1 Mpc $\simeq 3 \times 10^{24}$ cm $\simeq 1.6 \times 10^{38}$ GeV^{-1}.

Today, the energy density contributed by relativistic particles (photons and 3 neutrino species) is negligible: $\Omega_{rel} \simeq 4 \times 10^{-5}\, h^{-2}\, (T/2.7\text{ K})^4$. However, since $\rho_{rel} \propto R^{-4}$, while $\rho_{nonrel} \propto R^{-3}$, early on relativistic species dominated the energy density. For R/R(today) $< 4 \times 10^{-5}\, (\Omega h^2)^{-1}\, (T/2.7\text{ K})^4$, which corresponds to $t < 4 \times 10^{10}$ s $(\Omega h^2)^{-2}\, (T/2.7\text{ K})^6$ and $T > 6$ eV $(\Omega h^2)(2.7\text{ K}/T)^3$, the energy density of the Universe was dominated by relativistic particles. Since the curvature term varies as $R(t)^{-2}$, it too was small compared to the energy density contributed by relativistic particles early on, and so Eqn. 1.2 simplifies to:

$$H \equiv (\dot{R}/R) = (4\pi^3 g_*/45)^{1/2}\, T^2/m_{pl}, \tag{1.7}$$

$$= 1.66\, g_*^{1/2}\, T^2/m_{pl},$$

$$(\text{valid for } t \lesssim 10^{10}\text{ s},\ T \gtrsim 10\text{ eV}).$$

Here g_* counts the total number of effective degrees of freedom of all the relativistic particles (i.e., those species with mass << T):

$$g_* = \sum_{Bose} g_i(T_i/T)^4 + 7/8 \sum_{Fermi} g_i(T_i/T)^4 \quad , \tag{1.8}$$

where T_i is the temperature of species i, and T is the photon temperature. For example: $g_*(3\text{ K}) = 3.36$ $(\gamma, 3\ \bar{\nu}\nu)$; $g_*(\text{few MeV}) = 10.75$ $(\gamma, e^{\pm}, 3\ \bar{\nu}\nu)$; $g_*(\text{few 100 GeV}) = 110$ $(\gamma, W^{\pm}\, Z^0$, 8 gluons, 3 families of quarks and leptons, and 1 Higgs doublet).

If thermal equilibrium is maintained, then the second Friedmann equation, Eqn. 1.3 - conservation of energy, implies that the entropy per comoving volume (a volume with fixed r, θ, ϕ coordinates) $S \propto sR^3$ remains constant. Here s is the entropy density, which is dominated by the contribution from relativistic particles, and is given by:

$$s = (\rho + p)/T \simeq 2\pi^2 g_* T^3/45. \qquad (1.9)$$

The entropy density s itself is proportional to the number density of relativistic particles. So long as the expansion is adiabatic (i.e., in the absence of entropy production) S (and s) will prove to be useful fiducials. For example, at low energies ($E << 10^{14}$ GeV) baryon number is effectively conserved, and so the net baryon number per comoving volume $N_B \propto n_B(\equiv n_b - n_{\bar{b}}) R^3$ remains constant, implying that the ratio n_B/s is a constant of the expansion. Today $s \simeq 7n_\gamma$, so that $n_B/s \simeq \eta/7$, where $\eta \equiv n_b/n_\gamma$ is the baryon-to-photon ratio, which as we shall soon see, is known from primordial nucleosynthesis to be in the range: $4 \times 10^{-10} \lesssim \eta \lesssim 7 \times 10^{-10}$. The fraction of the critical density contributed by baryons (Ω_b) is related to η by:

$$\Omega_b \simeq 3.53 \times 10^{-3} (\eta/10^{-10}) h^{-2} (T/2.7\ K)^3. \qquad (1.10)$$

Whenever $g_* \simeq$ constant, the constancy of the entropy per comoving volume implies that $T \propto R^{-1}$; together with Eqn. 1.7 this gives

$$R(t) = R(t_o)(t/t_o)^{1/2}, \qquad (1.11)$$

$$t \simeq 0.3\ g_*^{-1/2}\ m_{pl}/T^2,$$

$$\simeq 2.4 \times 10^{-6}\ s\ g_*^{-1/2}\ (T/\mathrm{GeV})^{-2}, \qquad (1.12)$$

valid for $t \lesssim 10^{10}$ s and $T \gtrsim 10$ eV.

Finally, let me mention one more important feature of the standard cosmology, the existence of particle horizons. The distance that a light signal could have propagated since the bang is finite, and easy to compute. Photons travel on paths characterized by $ds^2 = 0$; for simplicity (and without loss of generality) consider a trajectory with $d\theta = d\phi = 0$. The coordinate distance covered by this photon since 'the bang' is just $\int_0^t dt'/R(t')$, corresponding to a physical distance (measured at time t) of

$$d_H(t) = R(t) \int_0^t dt'/R(t') \qquad (1.13)$$

$$= t/(1 - n) \qquad [\text{for } R \propto t^n,\ n < 1].$$

If $R \propto t^n$ ($n < 1$), then the horizon distance is finite and $\simeq t \simeq H^{-1}$. Note that even if $d_H(t)$ diverges (e.g., if $R \propto t^n$, $n \geq 1$), the Hubble radius H^{-1} still sets the scale for the 'physics horizon'. Since all physical lengths scale with R(t), they e-fold in a time of $O(H^{-1})$. Thus a coherent microphysical process can only operate over a time interval $\lesssim O(H^{-1})$, implying that a causally-coherent microphysical process can only operate over distances $\lesssim O(H^{-1})$.

During the radiation-dominated epoch $n = 1/2$ and $d_H = 2t$; the baryon number and entropy within the horizon at time t are easily computed:

$$S_{HOR} = (4\pi/3) t^3\ s,$$

$$\simeq 0.05\ g_*^{-1/2}\ (m_{pl}/T)^3; \tag{1.14}$$

$$N_{B-HOR} = (n_B/s) \times S_{HOR},$$

$$\simeq 10^{-12}\ (m_{pl}/T)^3; \tag{1.15a}$$

$$\simeq 10^{-2}\ M_\odot (T/\mathrm{MeV})^{-3}; \tag{1.15b}$$

where I have assumed that n_B/s has remained constant and has the value $\simeq 10^{-10}$. A solar mass ($M_\odot$) of baryons is $\simeq 1.2 \times 10^{57}$ baryons (or 2×10^{33} g).

Although our verifiable knowledge of the early history of the Universe only takes us back to $t \simeq 10^{-2}$ s and $T \simeq 10$ MeV (the epoch of primordial nucleosynthesis), nothing in our present understanding of the laws of physics suggests that it is unreasonable to extrapolate back to times as early as $\simeq 10^{-43}$ s and temperatures as high as $\simeq 10^{19}$ GeV. At high energies the interactions of quarks and leptons are asymptotically free (and/or weak) justifying the dilute gas approximation made in Eqn. 1.6. At energies below 10^{19} GeV quantum corrections to General Relativity are expected to be small. I hardly need to remind the reader that 'reasonable' does not necessarily mean 'correct'. Making this extrapolation, I have summarized 'The Complete History of the Universe' in Fig. 1.1. [For more complete reviews of the standard cosmology I refer the interested reader to refs. 5 and 6.]

Primordial Nucleosynthesis

At present the most stringent test of the standard cosmology is big bang nucleosynthesis. Here I will briefly review primordial nucleosynthesis, discuss the concordance of the predictions with the observations, and mention one example of how primordial nucleosynthesis has been used as a probe of particle physics--counting the number of light neutrino species.

The two fundamental assumptions which underlie big bang nucleosynthesis are: the validity of General Relativity and that the Universe was once hotter than a few MeV. An additional assumption (which, however, is not necessary) is that the lepton number, $n_L/n_\gamma = (n_{e^-} - n_{e^+})/n_\gamma + (n_\nu - n_{\bar\nu})/n_\gamma \simeq \eta + (n_\nu - n_{\bar\nu})/n_\gamma$, like the baryon number ($\equiv \eta$) is small. Having swallowed these assumptions, the rest follows like 1-2-3.

<u>Frame 1:</u> $t \simeq 10^{-2}$ sec, $T \simeq 10$ MeV. The energy density of the Universe is dominated by relativistic species: γ, e^+e^-, $\nu_i\bar\nu_i$ ($i = e, \mu, \tau, \ldots$); $g_* = 10.75$ (assuming 3 neutrino species). Thermal equilibrium is maintained by weak interactions ($e^+ + e^- \leftrightarrow \nu_i + \bar\nu_i$, $e^+ + n \leftrightarrow p + \bar\nu_e$, $e^- + p \leftrightarrow n + \nu_e$) as well as electromagnetic interactions ($e^+ + e^- \leftrightarrow \gamma + \gamma$, $\gamma + p \leftrightarrow \gamma + p$, etc.), both of which are occurring rapidly compared to the expansion rate $H = \dot R/R$. Thermal equilibrium implies that $T_\nu = T_\gamma$ and that $n/p = \exp(-\Delta m/T)$; here n/p is the neutron to proton ratio and $\Delta m = m_n - m_p$. No nucleosynthesis is occurring yet because of the tiny equilibrium abundance of D: $n_D/n_b \simeq \eta \exp(2.2\ \mathrm{MeV}/T) \simeq 10^{-10}$, where n_b, n_D, and n_γ are the baryon, deuterium, and photon number

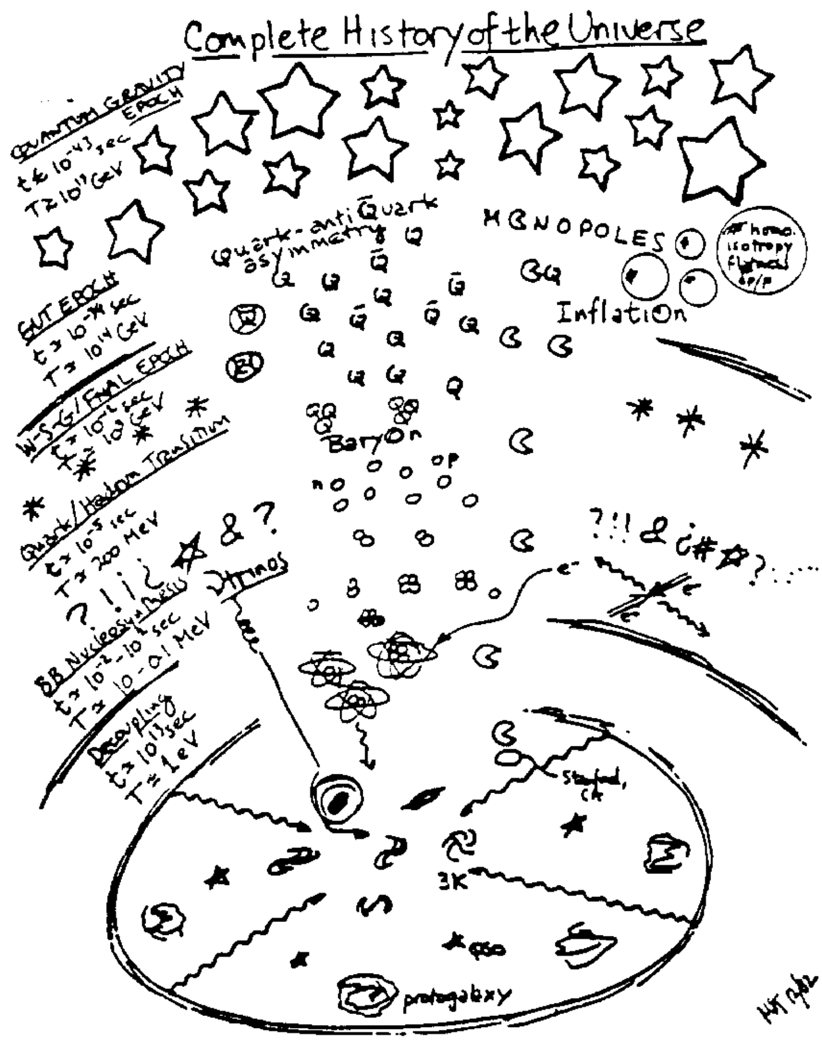

Fig. 1.1 'The Complete History of the Universe'. Highlights include: decoupling ($t \simeq 10^{13}$ s, $T \simeq 1/3$ eV) - the surface of last scattering for the cosmic microwave background, epoch after which matter and radiation cease to interact and matter 'recombines' into neutral atoms (D, ^{3}He, ^{4}He, ^{7}Li); also marks the beginning of the formation of structure; primordial nucleosynthesis ($t \simeq 10^{-2}$ s, $T \simeq 10$ MeV) - epoch during which all of the free neutrons and some of the free protons are synthesized into D, ^{3}He, ^{4}He, and ^{7}Li, and the surface of last scattering for the cosmic neutrino backgrounds; quark/hadron transition ($t \simeq 10^{-5}$ s, $T \simeq$ few 100 MeV) - epoch of 'quark enslavement' [confinement transition in SU(3)]; W-S-G epoch associated with electroweak breaking, SU(2) x U(1) $\rightarrow$ U(1); GUT epoch (?? $t \simeq 10^{-34}$ s, $T \simeq 10^{14}$ GeV??) - SSB of the GUT, during which the baryon asymmetry of the Universe evolves, monopoles are produced, and 'inflation' may occur; the Quantum Gravity Wall ($t \simeq 10^{-43}$ s, $T \simeq 10^{19}$ GeV).

densities, and 2.2 MeV is the binding energy of the deuteron. This is the so-called deuterium bottleneck.

Frame 2: $t \simeq 1$ sec, $T \simeq 1$ MeV. At about this temperature the weak interaction rates become slower than the expansion rate and thus weak interactions effectively cease occurring. The neutrinos decouple and thereafter expand adiabatically ($T_\nu \propto R^{-1}$). This epoch is the surface of last scattering for the neutrinos; detection of the cosmic neutrino seas would allow us to directly view the Universe as it was 1 sec after 'the bang'. From this time forward the neutron to proton ratio no longer 'tracks' its equilibrium value, but instead 'freezes out' a value $\simeq 1/6$, very slowly decreasing, due to occasional free neutron decays. A little bit later ($T \simeq m_e/3$), the $e^\pm$ pairs annihilate and transfer their entropy to the photons, heating the photons relative to the neutrinos, so that from this point on $T_\nu \simeq (4/11)^{1/3} T_\gamma$. The 'deuterium bottleneck' continues to operate, preventing nucleosynthesis.

Frame 3: $t \simeq 200$ sec, $T \simeq 0.1$ MeV. At about this temperature the 'deuterium bottleneck' breaks [$n_D/n_b \simeq \eta \exp(2.2\ \mathrm{MeV}/T) \simeq 1$], and nucleosynthesis begins in earnest. Essentially all the neutrons present ($n/p \simeq 1/7$) are quickly incorporated first into D, and then into ^{4}He nuclei. Trace amounts of D and ^{3}He remain unburned; substantial nucleosynthesis beyond ^{4}He is prevented by the lack of stable isotopes with A = 5 and 8, and by coulomb barriers. A small amount of ^{7}Li is synthesized by ^{4}He(t, γ)^{7}Li (for $\eta \lesssim 3 \times 10^{-10}$) and by ^{4}He(^{3}He, γ)^{7}Be followed by the eventual β-decay of ^{7}Be to ^{7}Li (for $\eta \gtrsim 3 \times 10^{-10}$).

The nucleosynthetic yields depend upon η, N_ν (which I will use to parameterize the number of light ($\lesssim 1$ MeV) species present, other than γ and $e^\pm$), and in principle all the nuclear reaction rates which go into the reaction network. In practice, most of the rates are known to sufficient precision that the yields only depend upon a few rates. ^{4}He production depends only upon η, N_ν, and $\tau_{1/2}$, the neutron half-life, which determines the rates for all the weak processes which interconvert neutrons and protons. The mass fraction Y_p of ^{4}He produced increases monotonically with increasing values of η, N_ν, and $\tau_{1/2}$ - a fact which is simple to understand. Larger η means that the 'deuterium bottleneck' breaks earlier, when the value of n/p is larger. More light species (i.e., larger value of N_ν) increases the expansion rate (since $H \propto (G\rho)^{1/2}$), while a larger value of $\tau_{1/2}$ means slower weak interaction rates ($\propto \tau_{1/2}^{-1}$) - both effects cause the weak interactions to freeze out earlier, when n/p is larger. The yield of ^{4}He is determined by the n/p ratio when nucleosynthesis commences, $Y_p \simeq 2(n/p)/(1 + n/p)$, so that a higher n/p ratio means more ^{4}He is synthesized. At present the value of the neutron half-life is only known to an accuracy of about 2%: $\tau_{1/2}$ = 10.6 min ± 0.2 min. Since ν_e and ν_μ are known (from laboratory measurements) to be light, $N_\nu \geq 2$. Based upon the luminous matter in galaxies, η is known to be $\gtrsim 0.3 \times 10^{-10}$. If all the mass in binary galaxies and small groups of galaxies (as inferred by dynamical measurements) is baryonic, then η must be $\gtrsim 2 \times 10^{-10}$.

To an accuracy of about 10%, the yields of D and ^{3}He only depend upon η, and decrease rapidly with increasing η. Larger η corresponds to a higher nucleon density and earlier nucleosynthesis, which in turn

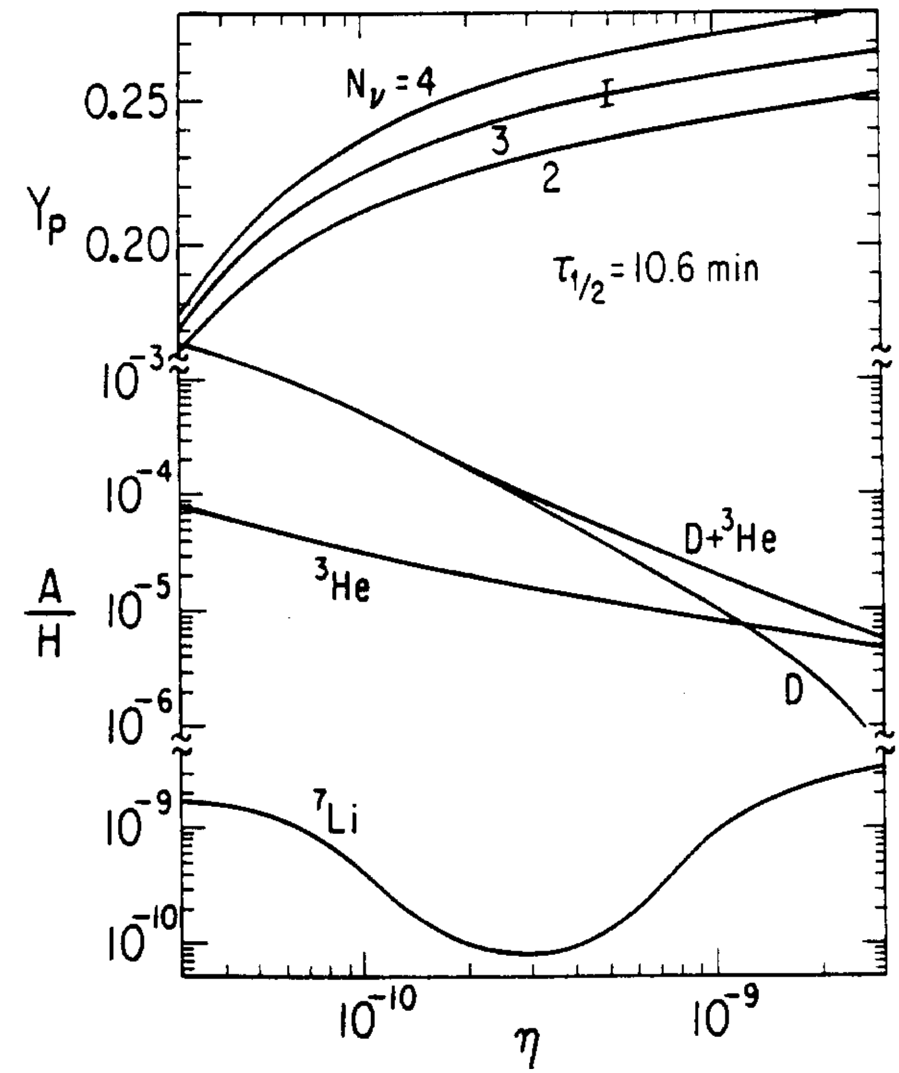

Fig. 1.2 The predicted primordial abundances of D, ^3He, ^4He, and ^7Li. [Note $\tau_{1/2}$ = 10.6 min was used; error bar shows $\Delta\tau_{1/2}$ = ± 0.2 min; Y_p = mass of ^4He; N_ν = equivalent number of light neutrino species.] Inferred primordial abundances: Y = 0.23–0.25; $(D/H) \geq 1 \times 10^{-5}$; $(D + {}^3He)/H \leq 10^{-4}$; $^7Li/H = (1.1 \pm 0.4) \times 10^{-10}$. Concordance requires: $\eta = (4\text{-}7) \times 10^{-10}$ and $N_\nu \leq 4$.

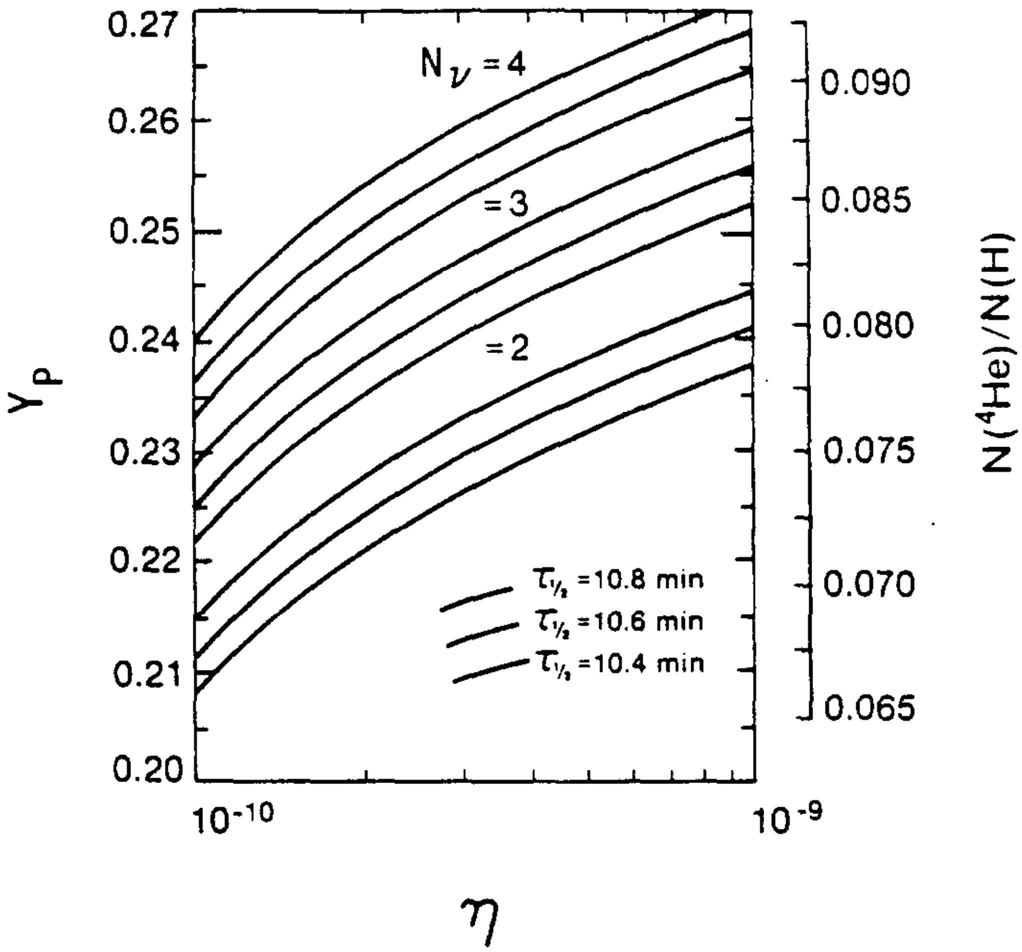

Fig. 1.3 The predicted primordial abundance of ^4He. Note that Y_p increases with increasing values of $\tau_{1/2}$, η, and N_ν. Hence lower bounds to η and $\tau_{1/2}$ and an upper bound to Y_p imply an upper bound to N_ν. Taking $\tau_{1/2}$ = 10.4 min, $\eta \geq 4 \times 10^{10}$ (based on D + ^3He production), and $Y_p \leq 0.25$, it follows that N_ν must be ≤ 4.

results in less D and ^3He remaining unprocessed. Because of large uncertainties in the rates of some reactions which create and destroy ^7Li, the predicted primordial abundance of ^7Li is only accurate to within about a factor of 2.

In 1946 Gamow[7] suggested the idea of primordial nucleosynthesis. In 1953, Alpher, Follin, and Herman[8] all but wrote a code to determine the primordial production of ^4He. Peebles[9] (in 1966) and Wagoner, Fowler, and Hoyle[10] (in 1967) wrote codes to calculate the primordial abundances. Yahil and Beaudet[11] (in 1976) independently developed a nucleosynthesis code and also extensively explored the effect of large lepton number ($n_\nu - n_{\bar{\nu}} = O(n_\gamma)$) on primordial nucleosynthesis. Wagoner's 1973 code[12] has become the 'standard code' for the standard model. In 1981 the reaction rates were updated by Olive et al.[13], the only significant change which resulted was an increase in the predicted ^7Li abundance by a factor of O(3). In 1982 Dicus et al.[14] corrected the weak rates in Wagoner's 1973 code for finite temperature effects and radiative/coulomb corrections, which led to a systematic decrease in Y_p of about 0.003. Figs. 1.2, 1.3 show the predicted abundances of D, ^3He, ^4He, and ^7Li, as calculated by the most up to date version of Wagoner's 1973 code.[15] The numerical accuracy of the predicted abundances is about 1%. Now let me discuss how the predicted abundances compare with the observational data. [This discussion is a summary of the collaborative work in ref. 15.]

The abundance of D has been determined in solar system studies and in UV absorption studies of the local interstellar medium (ISM). The solar system determinations are based upon measuring the abundances of deuterated molecules in the atmosphere of Jupiter and inferring the pre-solar (i.e., at the time of the formation of the solar system) D/H ratio from meteoritic and solar data on the abundance of ^3He. These determinations are consistent with a pre-solar value of $(D/H) = (2 \pm 1/2) \times 10^{-5}$. An average ISM value for $(D/H) = 2 \times 10^{-5}$ has been derived from UV absorption studies of the local ISM ($\lesssim$ few 100 pc), with individual measurements spanning the range $(1 - 4) \times 10^{-5}$. Note that these measurements are consistent with the solar system determinations of D/H.

The deuteron being very weakly-bound is easily destroyed and hard to produce, and to date, it has been difficult to find an astrophysical site where D can be produced in its observed abundance.[16] Thus, it is generally accepted that the presently-observed deuterium abundance provides a lower bound to the primordial abundance. Using $(D/H)_p \geq 1 \times 10^{-5}$ it follows that η must be less than about 10^{-9} in order for the predictions of primordial nucleosynthesis to be concordant with the observed abundance of D. [Note: because of the rapid variation of $(D/H)_p$ with η, this upper bound to η is rather insensitive to the precise lower bound to $(D/H)_p$ used.] Using Eqn. 1.10 to relate η to Ω_b, this implies an upper bound to Ω_b: $\Omega_b \lesssim 0.035h^{-2}(T/2.7K)^3 \lesssim 0.19$ -- baryons alone cannot close the Universe. One would like to also exploit the sensitive dependence of $(D/H)_p$ upon η to derive a lower bound to η for concordance; this is not possible because D is so easily destroyed. However, as we shall soon see, this end can be accomplished instead by using both D and ^3He.

The abundance of ^{3}He has been measured in solar system studies and by observations of the $^3He^+$ hyperfine line in galactic HII regions (the analog of the 21 cm line of H). The abundance of ^{3}He in the solar wind has been determined by analyzing gas-rich meteorites, lunar soil, and the foil placed upon the surface of the moon by the Apollo astronauts. Since D is burned to ^{3}He during the sun's approach to the main sequence, these measurements represent the pre-solar sum of D and ^{3}He. These determinations of D + ^{3}He are all consistent with a pre-solar $[(D + {}^3He)/H] \approx (4.0 \pm 0.3) \times 10^{-5}$. Earlier measurements of the $^3He^+$ hyperfine line in galactic HII regions and very recent measurements lead to derived present abundances of ^{3}He: $^3He/H \approx (3\text{-}20) \times 10^{-5}$. The fact that these values are higher than the pre-solar abundance is consistent with the idea that the abundance of ^{3}He should increase with time due to the stellar production of ^{3}He by low mass stars.

^{3}He is much more difficult to destroy than D. It is very hard to efficiently dispose of ^{3}He without also producing heavy elements or large amounts of ^{4}He (environments hot enough to burn ^{3}He are usually hot enough to burn protons to ^{4}He). In ref. 15 we have argued that in the absence of a Pop III generation of very exotic stars which process essentially all the material in the Universe and in so doing destroy most of the ^{3}He without overproducing ^{4}He or heavy elements, ^{3}He can have been astrated (i.e. reduced by stellar burning) by a factor of no more than $f_a \approx 2$. [The youngest stars, e.g. our sun, are called Pop I; the oldest observed stars are called Pop II. Pop III refers to a yet to be discovered, hypothetical first generation of stars.] Using this argument and the inequality

$$[(D+{}^3He)/H]_p \leq \text{pre-solar}(D/H) + f_a\ \text{pre-solar}({}^3He/H) \qquad (1.16)$$
$$\leq (1-f_a)\text{pre-solar}(D/H) + f_a \text{pre-solar}(D+{}^3He)/H;$$

the presolar abundances of D and D + ^{3}He can be used to derive an upper bound to the primordial abundance of D + ^{3}He: $[(D + {}^3He)/H]_p \lesssim 8 \times 10^{-5}$. [For a very conservative astration factor, $f_a \approx 4$, the upper limit becomes 13×10^{-5}.] Using 8×10^{-5} as an upper bound on the primordial D + ^{3}He production implies that for concordance, η must be greater than 4×10^{-10} (for the upper bound of 13×10^{-5}, η must be greater than 3×10^{-10}). To summarize, consistency between the predicted big bang abundances of D and ^{3}He, and the derived abundances observed today requires η to lie in the range $\approx (4 - 10) \times 10^{-10}$.

Until very recently, our knowledge of the ^{7}Li abundance was limited to observations of meteorites, the local ISM, and Pop I stars, with a derived present abundance of $^7Li/H \approx 10^{-9}$ (to within a factor of 2). Given that ^{7}Li is produced by cosmic ray spallation and some stellar processes, and is easily destroyed (in environments where $T \gtrsim 2 \times 10^6$K), there is not the slightest reason to suspect (or even hope!) that this value accurately reflects the primordial abundance. Recently, Spite and Spite[17] have observed ^{7}Li lines in the atmospheres of 13 unevolved halo and old disk stars with very low metal abundances ($Z_\odot/12$ - $Z_\odot/250$), whose masses span the range of $\approx (0.6 - 1.1)M_\odot$. Stars less massive than about 0.7 $M_\odot$ are expected to astrate (by factors $\gtrsim$ O(10)) their ^{7}Li abundance during their approach to the MS, while stars more massive than about 1 $M_\odot$ are not expected to significantly astrate ^7Li in their outer layers. Indeed, they see this trend in their data, and deduce a primordial ^7Li abundance of: $^7Li/H \approx (1.12 \pm 0.38) \times 10^{-10}$. Remarkably, this is the predicted big bang production for η in the range $(2 - 5) \times 10^{-10}$. If we take this to be the primordial ^7Li abundance, and allow for a possible factor of 2 uncertainty in the predicted abundance of Li (due to estimated uncertainties in the reaction rates which affect ^7Li), then concordance for ^7Li restricts η to the range $(1 - 7) \times 10^{-10}$. Note, of course, that their derived ^7Li abundance is the pre-Pop II abundance, and may not necessarily reflect the true primordial abundance (e.g., if a Pop III generation of stars processed significant amounts of material).

In sum, the concordance of big bang nucleosynthesis predictions with the derived abundances of D and ^3He requires $\eta \approx (4 - 10) \times 10^{-10}$; moreover, concordance for D, ^3He, and ^7Li further restricts η: $\eta \approx (4 - 7) \times 10^{-10}$.

In the past few years the quality and quantity of ^4He observations has increased markedly. In Fig. 1.4 all the ^4He abundance determinations derived from observations of recombination lines in HII regions (galactic and extragalactic) are shown as a function of metalicity Z (more precisely, 2.2 times the mass fraction of ^{16}O).

Since ^4He is also synthesized in stars, some of the observed ^4He is <u>not</u> primordial. Since stars also produce metals, one would expect some correlation between Y and Z, or at least a trend: lower Y where Z is lower. Such a trend is apparent in Fig. 1.4. From Fig. 1.4 it is also clear that there is a large primordial component to ^4He: $Y_p \approx 0.22 - 0.26$. Is it possible to pin down the value of Y_p more precisely?

There are many steps in going from the line strengths (what the observer actually measures), to a mass fraction of ^4He (e.g., corrections for neutral ^4He, reddening, etc.). In galactic HII regions, where abundances can be determined for various positions within a given HII region, variations are seen within a given HII region. Observations of extragalactic HII regions are actually observations of a superposition of <u>several</u> HII regions. Although observers have quoted statistical uncertainties of $\Delta Y \approx \pm 0.01$ (or lower), from the scatter in Fig. 1.4 it is clear that the systematic uncertainties must be larger. For example, different observers have derived ^4He abundances of between 0.22 and 0.25 for I Zw18, an extremely metal-poor dwarf emission line galaxy.

Perhaps the safest way to estimate Y_p is to concentrate on the ^4He determinations for metal-poor objects. From Fig. 1.4 $Y_p \approx 0.23 - 0.25$ appears to be consistent with all the data (although Y_p as low as 0.22 or high as 0.26 could not be ruled out). Recently Künth and Sargent[18] have studied 13 metal-poor ($Z \lesssim Z_\odot/5$) Blue Compact galaxies. From a weighted average for their sample they derive a <u>primordial abundance</u> $Y_p \approx 0.245 \pm 0.003$; allowing for a 3σ variation this suggests $0.236 \lesssim Y_p \lesssim 0.254$.

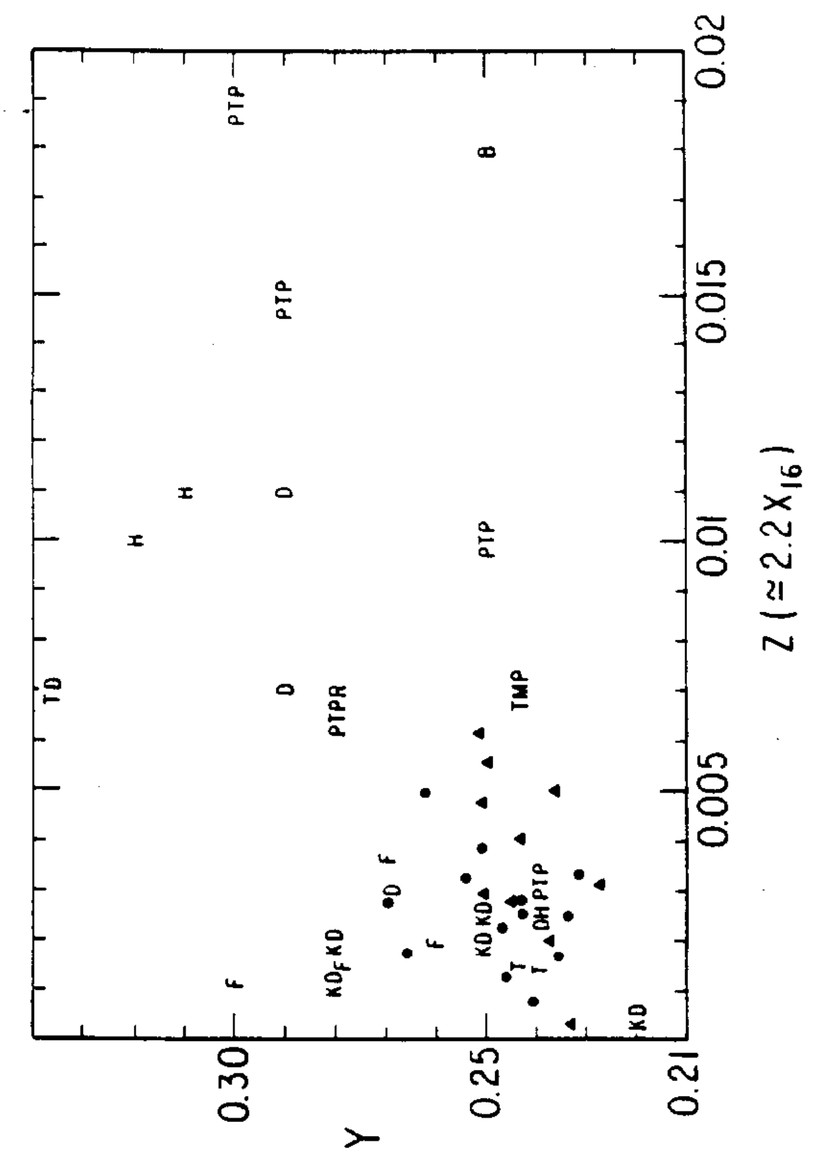

Fig. 1.4 Summary of ^{4}He abundance determinations (galactic and extragalactic) from recombination lines in HII regions vs. mass fraction of heavy ($A \geq 12$) elements Z ($\approx$ 2.2 mass fraction of ^{16}O). Note, observers do not usually quote errors for individual objects--scatter is probably indicative of the uncertainties. The triangles and filled circles represent two data sets of note: circles - 13 very metal poor emission line galaxies (Kunth and Sargent[18]); triangles - 9 metal poor, compact galaxies (Lequeux etal.[18]).

For the concordance range deduced from D, ^{3}He, and ^{7}Li ($\eta \geq 4 \times 10^{-10}$) and $\tau_{1/2} \gtrsim 10.4$ min, the predicted ^{4}He abundance is

$$Y_p \gtrsim \begin{array}{ll} 0.230 & N_\nu = 2. \\ 0.244 & \quad = 3. \\ 0.256 & \quad = 4. \end{array}$$

[Note, that $N_\nu = 2$ is permitted only if the τ-neutrino is heavy ($\gtrsim$ few MeV) and unstable; the present experimental upper limit on its mass is 160 MeV.] Thus, since $Y_p = 0.23 - 0.25$ ($0.22 - 0.26$?) there are values of η, N_ν, and $\tau_{1/2}$ for which there is agreement between the abundances predicted by big bang nucleosynthesis and the primordial abundances of D, ^{3}He, ^{4}He, and ^{7}Li derived from observational data.

To summarize, the only isotopes which are predicted to be produced in significant amounts during the epoch of primordial nucleosynthesis are: D, ^{3}He, ^{4}He, and ^{7}Li. At present there is concordance between the predicted primordial abundances of all 4 of these elements and their observed abundances for values of N_ν, $\tau_{1/2}$, and η in the following intervals: $2 \leq N_\nu \leq 4$; 10.4 min $\lesssim \tau_{1/2} \lesssim$ 10.8 min; and $4 \times 10^{-10} \lesssim \eta \lesssim 7 \times 10^{-10}$ (or 10×10^{-10} if the ^{7}Li abundance is not used). This is a truly remarkable achievement, and strong evidence that the standard model is valid back as early as 10^{-2} sec after 'the bang'.

The standard model will be in serious straights if the primordial mass fraction of ^{4}He is unambiguously determined to be less than 0.22. What alternatives exist if $Y_p \lesssim 0.22$? If a generation of Pop III stars which efficiently destroyed ^{3}He and ^{7}Li existed, then the lower bound to η based upon D, ^{3}He, (and ^{7}Li) no longer exists. The only solid lower bound to η would then be that based upon the amount of luminous matter in galaxies (i.e., the matter inside the Holmberg radius): $\eta \gtrsim 0.3 \times 10^{-10}$. In this case the predicted Y_p could be as low as 0.15 or 0.16. Although small amounts of anisotropy increase[19] the primordial production of ^{4}He, recent work[20] suggests that larger amounts could decrease the primordial production of ^{4}He. Another possibility is neutrino degeneracy; a large lepton number ($n_\nu - n_{\bar\nu} = O(n_\gamma)$) drastically modifies the predictions of big bang nucleosynthesis.[21] Finally, one might have to discard the standard cosmology altogether.

Primordial Nucleosynthesis as a Probe

If, based upon its apparent success, we accept the validity of the standard model, we can use primordial nucleosynthesis as a probe of cosmology and particle physics. For example, concordance requires: $4 \times 10^{-10} \lesssim \eta \lesssim 7 \times 10^{-10}$ and $N_\nu \lesssim 4$. This is the most precise determination we have of η and implies that

$$0.014h^{-2}(T/2.7K)^3 \lesssim \Omega_b \lesssim 0.024h^{-2}(T/2.7K)^3 \qquad (1.17)$$

$$0.014 \lesssim \Omega_b \lesssim 0.14,$$

$$n_B/s \simeq \eta/7 \simeq (6 - 10) \times 10^{-11}. \qquad (1.18)$$

If, as some dynamical studies suggest, $\Omega > 0.14$, then some other non-baryonic form of matter must account for the difference between Ω and Ω_b. [For a recent review of the measurements of Ω, see refs. 22, 23.] Numerous candidates have been proposed for the dark matter, including primordial black holes, axions, quark nuggets, photinos, gravitinos, relativistic debris, massive neutrinos, sneutrinos, monopoles, pyrgons, maximons, etc. [A discussion of some of these candidates is given in refs. 3, 24.]

With regard to the limit on N_ν, Schvartsman[25] first emphasized the dependence of the yield of ^4He on the expansion rate of the Universe during nucleosynthesis, which in turn is determined by g_*, the effective number of massless degrees of freedom. As mentioned above the crucial temperature for ^4He synthesis is $\simeq 1$ MeV -- the freeze out temperature for the n/p ratio. At this epoch the massless degrees of freedom include: γ, $\nu\bar{\nu}$, $e^\pm$ pairs, and any other light particles present, and so

$$g_* = g_\gamma + 7/8(g_{e^\pm} + N_\nu g_{\nu\bar{\nu}}) + \sum_{Bose} g_i\, (T_i/T)^4 + 7/8 \sum_{Fermi} g_i (T_i/T)^4$$
$$= 5.5 + 1.75N_\nu + \sum_{Bose} g_i (T_i/T)^4 + 7/8 \sum_{Fermi} g_i (T_i/T)^4. \qquad (1.19)$$

Here T_i is the temperature of species i, T is the photon temperature, and the total energy density of relativistic species is: $\rho = g_* \pi^2 T^4/30$. The limit $N_\nu \lesssim 4$ is obtained by assuming that the only species present are: γ, $e^\pm$, and N_ν neutrinos species, and follows because for $\eta \gtrsim 4 \times 10^{-10}$, $\tau_{1/2} \gtrsim 10.4$ min, and $N_\nu \gtrsim 4$, the mass fraction of ^4He produced is $\gtrsim 0.25$ (which is greater than the observed abundance). More precisely, $N_\nu \lesssim 4$ implies

$$g_* \lesssim 12.5 \qquad (1.20)$$

or

$$1.75 \gtrsim 1.75(N_\nu - 3) + \sum_{Bose} g_i (T_i/T)^4 + \sum_{Fermi} g_i (T_i/T)^4. \qquad (1.21)$$

At most 1 additional light ($\lesssim$ MeV) neutrino species can be tolerated; many more additional species can be tolerated if their temperatures T_i are < T. [Big bang nucleosynthesis limits on the number of light ($\lesssim$ MeV) species have been derived and/or discussed in refs. 26.]

The number of neutrino species can also be determined by measuring the width of the Z^0 boson: each neutrino flavor less massive than $O(m_Z/2)$ contributes ≈ 190 MeV to the width of the Z^0. Preliminary results on the width of the Z^0 imply that $N_\nu \lesssim O(20)$[27]. Note that while big bang nucleosynthesis and the width of the Z^0 both provide information about the number of neutrino flavors, they 'measure' slightly different quantities. Big bang nucleosynthesis is sensitive to the number of light ($\lesssim$ MeV) neutrino species, and all other light degrees of freedom, while the width of the Z^0 is determined by the number of particles less massive than about 50 GeV which couple to the Z^0 (neutrinos among them). This issue has been recently discussed in ref. 28.

Given the important role occupied by big bang nucleosynthesis, it is clear that continued scrutiny is in order. The importance of new observational data cannot be overemphasized: extragalactic D abundance determinations (Is the D abundance universal? What is its value?); more measurements of the ^{3}He abundance (What is its primordial value?); continued improvement in the accuracy of ^{4}He abundances in very metal poor HII regions (Recall, the difference between $Y_p = 0.22$ and $Y_p = 0.23$ is crucial); and further study of the ^{7}Li abundance in very old stellar populations (Has the primordial abundance of ^{7}Li already been measured?). Data from particle physics will prove useful too: a high precision determination of $\tau_{1/2}$ (i.e., $\Delta\tau_{1/2} \lesssim \pm 0.05$ min) will all but eliminate the uncertainty in the predicted ^{4}He primordial abundance; an accurate measurement of the width of the recently-found Z^0 vector boson will determine the total number of neutrino species (less massive than about 50 GeV) and thereby bound the total number of light neutrino species. All these data will not only make primordial nucleosynthesis a more stringent test of the standard cosmology, but they will also make primordial nucleosynthesis a more powerful probe of the early Universe.

'Freeze-out' and the Making of a Relic Species

In Eqns. 1.19, 1.21 I allowed for a species to have a temperature T_i which is less than the photon temperature. What could lead to this happening? As the Universe expands it cools ($T \propto R^{-1}$), and a particle species can only remain in 'good thermal contact' if the reactions which are important for keeping it in thermal equilibrium are occurring rapidly compared to the rate at which T is decreasing (which is set by the expansion rate $-\dot{T}/T = \dot{R}/R = H$). Roughly-speaking the criterion is

$$\Gamma \gtrsim H, \tag{1.22}$$

where $\Gamma = n\langle\sigma v\rangle$ is the interaction rate per particle, n is the number density of target particles and $\langle\sigma v\rangle$ is the thermally-averaged cross section. When Γ drops below H, that reaction is said to 'freeze-out' or 'decouple'. The temperature T_f (or T_d) at which $H \approx \Gamma$ is called the freeze-out or decoupling temperature. [Note that if $\Gamma = aT^n$ and the Universe is radiation-dominated so that $H = (2t)^{-1} = 1.67\, g_*^{1/2}T^2/m_{pl}$, then the number of interactions which occur for $T \lesssim T_f$ is just: $\int_{T_f}^{0} \Gamma dt \simeq (\Gamma/H)|_{T_f}/(n-2) \simeq (n-2)^{-1}$]. If the species in question is relativistic

($T_f >> m_i$) when it decouples, then its phase space distribution (in momentum space) remains thermal (i.e., Bose-Einstein or Fermi-Dirac) with a temperature $T_i \propto R^{-1}$. [It is a simple exercise to show this.] So long as the photon temperature also decreases as R^{-1}, $T_i = T$, as if the species were still in good thermal contact.

However, due to the entropy release when various massive species annihilate (e.g., $e^{\pm}$ pairs when $T \simeq 0.1$ MeV), the photon temperature does not always decrease as R^{-1}. Entropy conservation ($S \propto g_* T^3$=constant) can, however, be used to calculate its evolution; if g_* is decreasing, then T will decrease less rapidly than R^{-1}. As an example consider neutrino freeze-out. The cross section for processes like $e^+e^- \leftrightarrow \nu\bar{\nu}$ is: $\langle\sigma v\rangle \simeq 0.2 G_F^2 T^2$, and the number density of targets $n \simeq T^3$, so that $\Gamma \simeq 0.2\, G_F^2 T^5$. Equating this to H it follows that

$$T_f \simeq (30\, m_{pl}^{-1} G_F^{-2})^{1/3} \qquad (1.23)$$
$$\simeq \text{few MeV},$$

i.e., neutrinos freeze out before $e^{\pm}$ annihilations and do not share in subsequent entropy transfer. For $T \lesssim$ few MeV, neutrinos are decoupled and $T_\nu \propto R^{-1}$, while the entropy density in $e^{\pm}$ pairs and γs $s \propto R^{-3}$. Using the fact that before $e^{\pm}$ annihilation the entropy density of the $e^{\pm}$ pairs and γs is: $s \propto (7/8 g_{e^\pm} + g_\gamma)T^3 = 5.5\ T^3$ and that after $e^{\pm}$ annihilation $s \propto g_\gamma T^3 = 2T^3$, it follows that after the $e^{\pm}$ annihilations

$$T_\nu/T = [g_\gamma/(g_\gamma + 7/8\ g_{e^\pm})]^{1/3}$$
$$= (4/11)^{1/3}. \qquad (1.24)$$

Similarly, the temperature at the time of primordial nucleosynthesis T_i of a species which decouples at an arbitrary temperature T_d can be calculated:

$$T_i/T = [(g_\gamma + 7/8(g_{e^\pm} + N_\nu g_{\nu\bar{\nu}}))/g_{*d}]^{1/3}$$
$$\simeq (10.75/g_{*d})^{1/3} \qquad (\text{for } N_\nu = 3). \qquad (1.25)$$

Here $g_{*d} = g_*(T_d)$ is the number of species in equilibrium when the species in question decouples. Species which decouple at a temperature 30 MeV $\simeq m_\mu/3 \lesssim T \lesssim$ few 100 MeV do not share in the entropy release from $\mu^{\pm}$ annihilations, and $T_i/T \simeq 0.91$; the important factor for limits based upon primordial nucleosynthesis $(T_i/T)^4 \simeq 0.69$. Species which decouple at temperatures $T_d \gtrsim$ the temperature of the quark/hadron transition $\simeq$ few 100 MeV, do not share in the entropy transfer when the quark-gluon plasma [$g_* = g_\gamma + g_{Gluon} + 7/8(g_{e^\pm} + g_{\mu^\pm} + g_{\nu\bar{\nu}} + g_{u\bar{u}} + g_{d\bar{d}} + g_{s\bar{s}} + ..) \gtrsim 62$] hadronizes, and $T_i/T \simeq 0.56$; $(T_i/T)^4 \simeq 0.10$.

'Hot' relics- Consider a stable particle species X which decouples at a temperature $T_f >> m_X$. For $T < T_f$ the number density of Xs n_X just decreases as R^{-3} as the Universe expands. In the absence of entropy production the entropy density s also decreases as R^{-3}, and hence the ratio n_X/s remains constant. At freeze-out

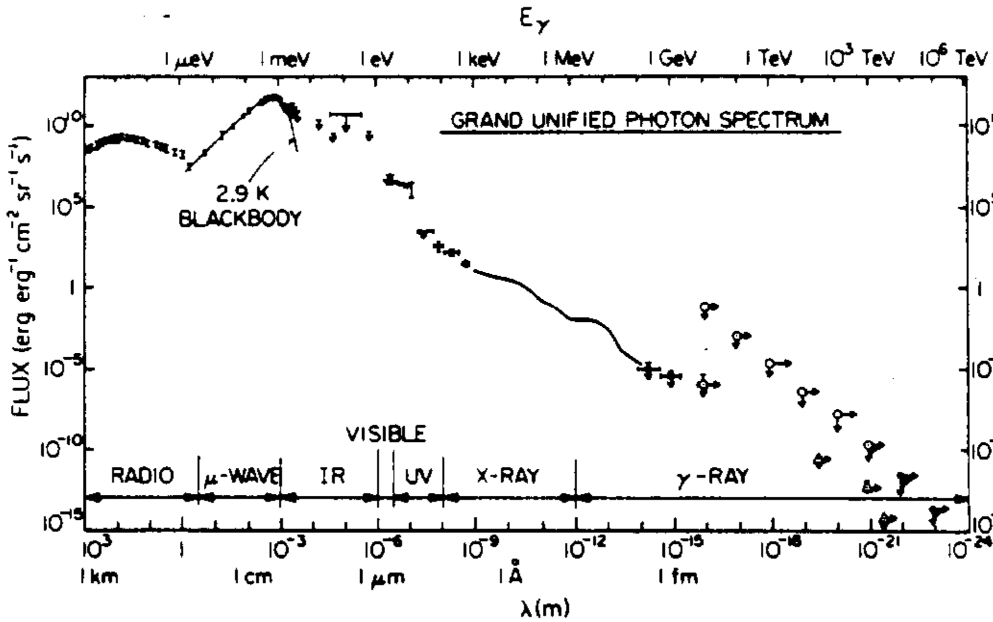

Fig. 1.5 The diffuse photon spectrum of the Universe from λ = 1 km to 10^{-24} m. Vertical arrows indicate upper limits.

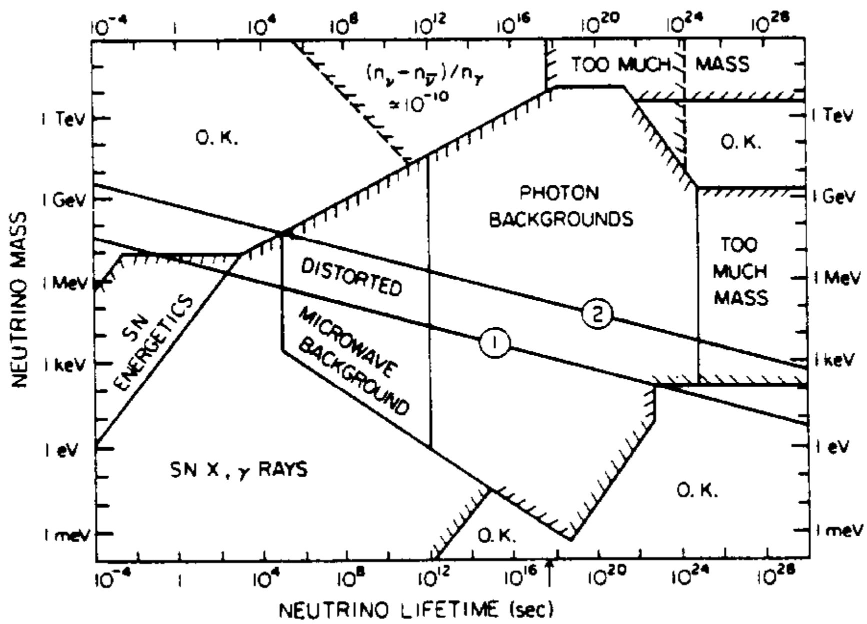

Fig. 1.6 Summary of astrophysical/cosmological constraints on neutrino masses/lifetimes. Lines 1 and 2 represent mass/lifetime relationships: $\tau = a \times 10^{-6}$ sec $(m_\mu/m_\nu)^5$, for a = 1, 10^{12}.

$$n_X/s = (g_{Xeff}\zeta(3)/\pi^2)/(2\pi^2 g_{*d}/45),$$
$$\simeq 0.278 g_{Xeff}/g_{*d}, \qquad (1.26)$$

where $g_{Xeff} = g_X$ for a boson or $3/4\ g_X$ for a fermion, $g_{*d} = g_*(T_d)$, and $\zeta(3) = 1.20206\ldots$. Today $s \simeq 7.1 n_\gamma$, so that the number density and mass density of Xs are

$$n_X \simeq (2g_{Xeff}/g_{*d})n_\gamma, \qquad (1.27)$$

$$\Omega_X = \rho_X/\rho_c \simeq 7.6(m_X/100\text{eV})(g_{Xeff}/g_{*d})h^{-2}(T/2.7\text{K})^3. \qquad (1.28)$$

Note, that if the entropy per comoving volume S has increased since the X decoupled, e.g., due to entropy production in a phase transition, then these values are decreased by the same factor that the entropy increased. As discussed earlier, Ωh^2 must be $\lesssim O(1)$, implying that for a stable particle species

$$m_X/100\ \text{eV} \lesssim 0.13\ g_{*d}/g_{Xeff}; \qquad (1.29)$$

for a neutrino species: $T_d \simeq$ few MeV, $g_{*d} \simeq 10.75$, $g_{Xeff} = 2 \times (3/4)$, so that $n_{\nu\bar{\nu}}/n_\gamma \simeq 3/11$ and m_ν must be $\lesssim 96$ eV. Note that for a species which decouples very early (say $g_{*d} = 200$), the mass limit (1.7 keV for $g_{Xeff} = 1.5$) which $\propto g_{*d}$ is much less stringent.

Constraint (1.29) obviously does not apply to an unstable particle with $\tau < 10$-15 billion yrs. However, any species which decays radiatively is subject to other very stringent constraints, as the photons from its decays can have various unpleasant astrophysical consequences, e.g., dissociating D, distorting the microwave background, 'polluting' various diffuse photon backgrounds, etc. The astrophysical/cosmological constraints on the mass/lifetime of an unstable neutrino species and the photon spectrum of the Universe are shown in Figs. 1.5, 1.6.

'Cold' relics- Consider a stable particle species which is still coupled to the primordial plasma ($\Gamma > H$) when $T \simeq m_X$. As the temperature falls below m_X, its equilibrium abundance is given by

$$n_X/n_\gamma = (g_{Xeff}/2)(\pi/8)^{1/2}(m_X/T)^{3/2}\exp(-m_X/T), \qquad (1.30)$$

$$n_X/s \simeq 0.17(g_{Xeff}/g_*)(m_X/T)^{3/2}\exp(-m_X/T), \qquad (1.31)$$

and in order to maintain an equilibrium abundance Xs must diminish in number (by annihilations since by assumption the X is stable). So long as $\Gamma_{ann} = n_X(\sigma v)_{ann} \gtrsim H$ the equilibrium abundance of Xs is maintained. When $\Gamma_{ann} = H$, when $T = T_f$, the Xs 'freeze-out' and their number density n_X decreases only due to the volume increase of the Universe, so that for $T \lesssim T_f$:

$$n_X/s \simeq (n_X/s)|_{T_f}. \qquad (1.32)$$

The equation for freeze-out ($\Gamma_{ann} \simeq H$) can be solved approximately, giving

Evidence for a Baryon Asymmetry

Within the solar system we can be very confident that there are no concentrations of antimatter (e.g., antiplanets). If there were, solar wind particles striking such objects would be the strongest γ-ray sources in the sky. Also, NASA has yet to lose a space probe because it annihilated with antimatter in the solar system.

Cosmic rays more energetic than O(0.1 GeV) are generally believed to be of "extrasolar" origin, and thereby provide us with samples of material from throughout the galaxy (and possibly beyond). The ratio of antiprotons to protons in the cosmic rays is about 3×10^{-4}, and the ratio of anti-^{4}He to ^{4}He is less than 10^{-5} (ref. 35). Antiprotons are expected to be produced as cosmic-ray secondaries (e.g. $p + p \rightarrow 3p + \bar{p}$) at about the 10^{-4} level. At present both the spectrum and total flux of cosmic-ray antiprotons are at variance with the simplest model of their production as secondaries. A number of alternative scenarios for their origin have been proposed including the possibility that the detected $\bar{p}$s are cosmic rays from distant antimatter galaxies. Although the origin of these $\bar{p}$s remains to be resolved, it is clear that they do not provide evidence for an appreciable quantity of antimatter in our galaxy. [For a recent review of antimatter in the cosmic rays we refer the reader to ref. 35.]

The existence of both matter and antimatter galaxies in a cluster of galaxies containing intracluster gas would lead to a significant γ-ray flux from decays of π^0s produced by nucleon-antinucleon annihilations. Using the observed γ-ray background flux as a constraint, Steigman[33] argues that clusters like Virgo, which is at a distance ≈ 20 Mpc ($\approx 10^{26}$ cm) and contains several hundred galaxies, must not contain both matter and antimatter galaxies.

Based upon the above-mentioned arguments, we can say that if there exist equal quantities of matter and antimatter in the Universe, then we can be absolutely certain they are separated on mass scales greater than 1 $M_\odot$, and reasonably certain they are separated on scales greater than (1-100) $M_{galaxy} \approx 10^{12}$-$10^{14} M_\odot$. As discussed below, this fact is virtually impossible to reconcile with a symmetric cosmology.

It has often been pointed out that we drive most of our direct knowledge of the large-scale Universe from photons, and since the photon is a self-conjugate particle we obtain no clue as to whether the source is made of matter or antimatter. Neutrinos, on the other hand, can in principle reveal information about the matter-antimatter composition of their source. Large neutrino detectors such as DUMAND may someday provide direct information about the matter-antimatter composition of the Universe on the largest scales.

Baryons account for only a tiny fraction of the particles in the Universe, the 3K-microwave photons being the most abundant species (yet detected). The number density of 3K photons is $n_\gamma = 399(T/2.7K)^3$ cm^{-3}. The baryon density is not nearly as well determined. Luminous matter (baryons in stars) contribute at least 0.01 of closure density ($\Omega_{lum} > 0.01$), and as discussed in Lecture 1 the age of the Universe requires that Ω_{tot} (and Ω_b) must be < 0(2). These direct determinations place the baryon-to-photon ratio $\eta \equiv n_b/n_\gamma$ in the range 3×10^{-11} to 6×10^{-8}. As I also discussed in Lecture 1 the yields of big-bang nucleosynthesis depend directly on η, and the production of amounts of D, ^{3}He, ^{4}He, and ^{7}Li that are consistent with their present measured abundances restricts η to the narrow range $(4-7) \times 10^{-10}$.

Since today it appears that $n_b >> n_{\bar{b}}$, η is also the ratio of net baryon number to photons. The number of photons in the Universe has not remained constant, but has increased at various epochs when particle species have annihilated (e.g. $e^\pm$ pairs at T ≈ 0.5 MeV). Assuming the expansion has been isentropic (i.e. no significant entropy production), the entropy per comoving volume ($\propto sR^3$) has remained constant. The "known entropy" is presently about equally divided between the 3K photons and the three cosmic neutrino backgrounds (e, μ, τ). Taking this to be the present entropy, the ratio of baryon number to entropy is

$$n_B/s \approx (1/7)\eta \approx (6-10) \times 10^{-11}, \qquad (2.1)$$

where $n_B \equiv n_b - n_{\bar{b}}$ and η is taken to be in the range $(4-7) \times 10^{-10}$. So long as the expansion is isentropic and baryon number is at least effectively conserved this ratio remains constant and is what I will refer to as the baryon number of the Universe.

Although the matter-antimatter asymmetry appears to be "large" today (in the sense that $n_B \approx n_b >> n_{\bar{b}}$), the fact that $n_B/s \simeq 10^{-10}$ implies that at very early times the asymmetry was "tiny" ($n_B << n_b$). To see this, let us assume for simplicity that nucleons are the fundamental baryons. Earlier than 10^{-6} s after the bang the temperature was greater than the mass of a nucleon. Thus nucleons and antinucleons should have been about as abundant as photons, $n_N \approx n_{\bar{N}} \approx n_\gamma$. The entropy density s is $\approx g_* n_\gamma \approx g_* n_N \approx 0(10^2)n_N$. The constancy of $n_B/s \approx 0(10^{-10})$ requires that for $t < 10^{-6}$ s, $(n_N - n_{\bar{N}})/n_N (\approx 10^2 n_B/s) \approx 0(10^{-8})$. During its earliest epoch, the Universe was nearly (but not quite) baryon symmetric.

The Tragedy of a Symmetric Cosmology

Suppose that the Universe were initially locally baryon symmetric. Earlier than 10^{-6} s after the bang nucleons and antinucleons were about as abundant as photons. For T < 1 GeV the equilibrium abundance of nucleons and antinucleons is $(n_N/n_\gamma)_{EQ} \approx (m_N/T)^{3/2} \exp(-m_N/T)$, and as the Universe cooled the number of nucleons and antinucleons would decrease tracking the equilibrium abundance as long as the annihilation rate $\Gamma_{ann} \approx n_N(\sigma v)_{ann} \approx n_N m_\pi^{-2}$ was greater than the expansion rate H. At a temperature T_f annihilations freeze out ($\Gamma_{ann} \approx H$), nucleons and antinucleons being so rare they can no longer find each other to annihilate. Using Eqn. 1.33 we can compute T_f: $T_f \simeq 0(20$ MeV). Because of the incompleteness of the annihilations, residual nucleon and antinucleon to photon ratios (given by Eqn. 1.34) $n_{\bar{N}}/n_\gamma \approx n_N/n_\gamma \simeq 10^{-18}$ are "frozen in." Even if the matter and antimatter could subsequently be separated, n_N/n_γ is a factor of 10^8 too small. To avoid 'the annihilation catastrophe', matter and antimatter must be separated on large scales before $t \approx 3 \times 10^{-3}$ s(T ≈ 20 MeV).

Statistical fluctuations: One possible mechanism for doing this is statistical (Poisson) fluctuations. The co-moving volume that encompasses our galaxy today contains $\simeq 10^{12}$ $M_\odot$ $\approx 10^{69}$ baryons and $\simeq 10^{79}$ photons. Earlier than 10^{-6} s after the bang this same comoving volume contained $\simeq 10^{79}$ photons and $\simeq 10^{79}$ baryons and antibaryons. In order to avoid the annihilation catastrophe, this volume would need an excess of baryons over antibaryons of $\simeq 10^{69}$, but from statistical fluctuations one would expect $N_b - N_{\bar{b}} \approx O(N_b^{1/2}) \simeq 3 \times 10^{39}$ - a mere 29 1/2 orders of magnitude too small!

Causality constraints: Clearly, statistical fluctuations are of no help, so consider a hypothetical interaction that separates matter and antimatter. In the standard cosmology the distance over which light signals (and hence causal effects) could have propagated since the bang (the horizon distance) is finite and $\simeq 2t$. When $T \approx 20$ MeV ($t \approx 3 \times 10^{-3}$ s) causally coherent regions contained only about 10^{-5} $M_\odot$. Thus, in the standard cosmology causal processes could have only separated matter and antimatter into lumps of mass $\lesssim 10^{-5}$ $M_\odot$ $<< M_{galaxy} \approx 10^{12}$ $M_\odot$. [In Lecture 4 I will discuss inflationary scenarios; in these scenarios it is possible that the Universe is globally symmetric, while asymmetric locally (within our observable region of the Universe). This is possible because inflation removes the causality constraint.]

It should be clear that the two observations, $n_b >> n_{\bar{b}}$ on scales at least as large as 10^{12} $M_\odot$ and $n_b/n_\gamma \approx (4-7) \times 10^{-10}$, effectively render all baryon-symmetric cosmologies untenable. A viable pre-GUT cosmology needed to have as an initial condition a tiny baryon number, $n_B/s \approx (6-10) \times 10^{-11}$--a very curious initial condition at that!

The Ingredients Necessary for Baryogenesis

More than a decade ago Sakharov[36] suggested that an initially baryon-symmetric Universe might dynamically evolve a baryon excess of $O(10^{-10})$, which after baryon-antibaryon annihilations destroyed essentially all of the antibaryons, would leave the one baryon per 10^{10} photons that we observe today. In his 1967 paper Sakharov outlined the three ingredients necessary for baryogenesis: (a) B-nonconserving interactions; (b) a violation of both C and CP; (c) a departure from thermal equilibrium.

It is clear that B(baryon number) must be violated if the Universe begins baryon symmetric and then evolves a net B. In 1967 there was no motivation for B nonconservation. After all, the proton lifetime is more than 35 orders of magnitude longer than that of any unstable elementary particle--pretty good evidence for B conservation. Of course, grand unification provides just such motivation, and proton decay experiments are likely to detect B nonconservation in the next decade if the proton lifetime is $\lesssim 10^{33}$ years.

Under C (charge conjugation) and CP (charge conjugation combined with parity), the B of a state changes sign. Thus a state that is either C or CP invariant must have B = 0. If the Universe begins with equal amounts of matter and antimatter, and without a preferred direction (as in the standard cosmology), then its initial state is both C and CP invariant. Unless both C and CP are violated, the Universe will remain C and CP invariant as it evolves, and thus cannot develop a net baryon number even if B is not conserved. Both C and CP violations are needed to provide an arrow to specify that an excess of matter be produced. C is maximally violated in the weak interactions, and both C and CP are violated in the K^0-$\bar{K}^0$ system. Although a fundamental understanding of CP violation is still lacking at present, GUTs can accommodate CP violation. It would be very surprising if CP violation only occurred in the K^0-$\bar{K}^0$ system and not elsewhere in the theory also (including the B-nonconserving sector). In fact, without miraculous cancellations the CP violation in the neutral kaon system will give rise to CP violation in the B-nonconserving sector at some level.

The necessity of a departure from thermal equilibrium is a bit more subtle. It has been shown that CPT and unitary alone are sufficient to guarantee that equilibrium particle phase space distributions are given by: $f(p) = [\exp(\mu/T+E/T)\pm 1]^{-1}$. In equilibrium, processes like $\gamma + \gamma \leftrightarrow b + \bar{b}$ imply that $\mu_b = -\mu_{\bar{b}}$, while processes like (but not literally) $\gamma + \gamma \leftrightarrow b + b$ require that $\mu_b = 0$. Since $E^2 = p^2 + m^2$ and $m_b = m_{\bar{b}}$ by CPT, it follows that in thermal equilibrium, $n_b \equiv n_{\bar{b}}$. [Note, $n = \int d^3p f(p)/(2\pi)^3$.]

Because the temperature of the Universe is changing on a characteristic timescale H^{-1}, thermal equilibrium can only be maintained if the rates for reactions that drive the Universe to equilibrium are much greater than H. Departures from equilibrium have occurred often during the history of the Universe. For example, because the rate for γ + matter $\rightarrow \gamma'$ + matter' is << H today, matter and radiation are not in equilibrium, and nucleons do not all reside in ^{56}Fe nuclei (thank God!).

The Standard Scenario: Out-of-Equilibrium Decay

The basic idea of baryogenesis has been discussed by many authors.[37-42] The model that incorporates the three ingredients discussed above and that has become the "standard scenario" is the so-called out-of-equilibrium decay scenario. I now describe the scenario in some detail.

Denote by "X" a superheavy ($\gtrsim 10^{14}$ GeV) boson whose interactions violate B conservation. X might be a gauge or a Higgs boson (e.g., the XY gauge bosons in SU(5), or the color triplet component of the 5 dimensional Higgs). [Scenarios in which the X particle is a superheavy fermion have also been suggested.] Let its coupling strength to fermions be $\alpha^{1/2}$, and its mass be M. From dimensional considerations its decay rate $\Gamma_D = \tau^{-1}$ should be

$$\Gamma_D \simeq \alpha M. \tag{2.2}$$

At the Planck time ($\simeq 10^{-43}$ s) assume that the Universe is baryon symmetric ($n_B/s = 0$), with all fundamental particle species (fermions, gauge and Higgs bosons) present with equilibrium distributions. At this epoch $T \simeq g_*^{-1/4} m_{pl} \simeq 3 \times 10^{18}$ GeV >> M. (Here I have taken $g_* \simeq O(100)$; in minimal SU(5) $g_* \simeq 160$.) So at the Planck time X, $\bar{X}$ bosons are very

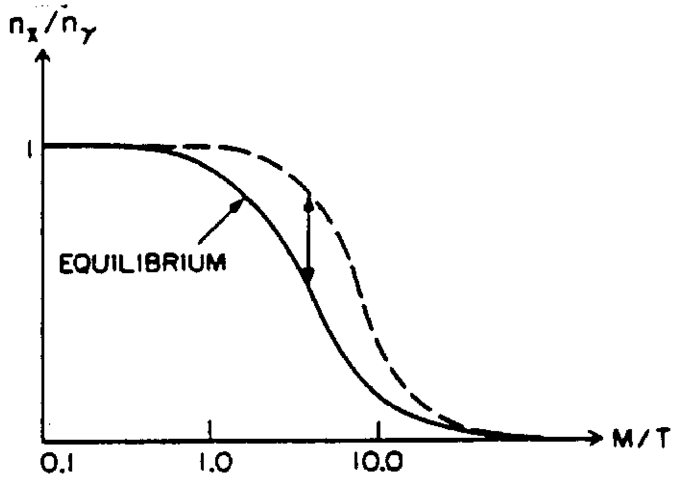

Fig. 2.1 The abundance of X bosons relative to photons. The broken curve shows the actual abundance, while the solid curve shows the equilibrium abundance.

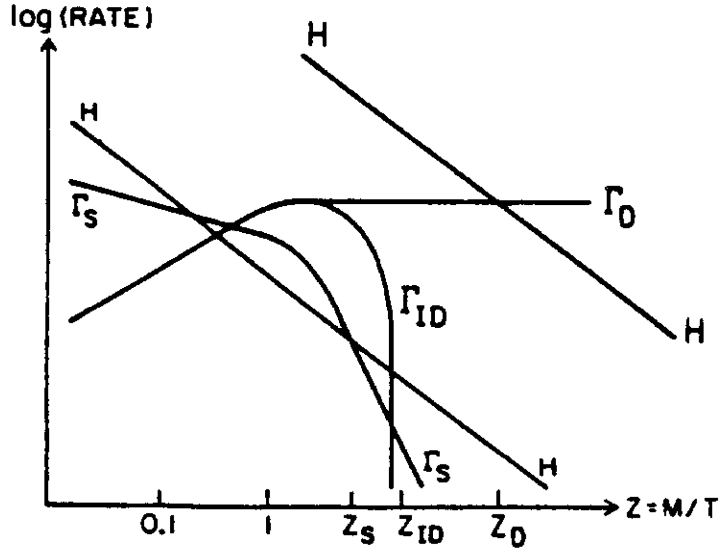

Fig. 2.2 Important rates as a function of $z = M/T$. H is the expansion rate, Γ_D the decay rate, Γ_{ID} the inverse decay rate, and Γ_s the $2 \leftrightarrow 2$ $\not{B}$ scattering rate. Upper line marked H corresponds to case where $K \ll 1$; lower line the case where $K > 1$. For $K \ll 1$, Xs decay when $z = z_D$; for $K > 1$, freeze out of IDs and S occur at $z = z_{ID}$ and z_s.

relativistic and up to statistical factors as abundant as photons: $n_X = n_{\bar{X}} \approx n_\gamma$. Nothing of importance occurs until $T \approx M$.

For $T < M$ the equilibrium abundance of X, $\bar{X}$ bosons relative to photons is

$$X_{EQ} \approx (M/T)^{3/2} \exp(-M/T),$$

where $X \equiv n_X/n_\gamma$ is just the number of X, $\bar{X}$ bosons per comoving volume. In order for X, $\bar{X}$ bosons to maintain an equilibrium abundance as T falls below M, they must be able to diminish in number rapidly compared to $H \approx |\dot{T}/T|$. The most important process in this regard is decay; other processes (e.g. annihilation) are higher order in α. If $\Gamma_D >> H$ for $T \approx M$, then X, $\bar{X}$ bosons can adjust their abundance (by decay) rapidly enough so that X "tracks" the equilibrium value. In this case thermal equilibrium is maintained and no asymmetry is expected to evolve.

More interesting is the case where $\Gamma_D < H \approx 1.66\, g_*^{1/2} T^2/m_{pl}$ when $T \approx M$, or equivalently $M > g_*^{-1/2} \alpha 10^{19}$ GeV. In this case, X, $\bar{X}$ bosons are not decaying on the expansion timescale ($\tau > t$) and so remain as abundant as photons ($X \approx 1$) for $T \lesssim M$; hence they are overabundant relative to their equilibrium number. This overabundance (indicated with an arrow in Fig. 2.1) is the departure from thermal equilibrium. Much later, when $T << M$, $\Gamma_D \approx H$ (i.e. $t \approx \tau$), and X, $\bar{X}$ bosons begin to decrease in number as a result of decays. To a good approximation they decay freely since the fraction of fermion pairs with sufficient center-of-mass energy to produce an X or $\bar{X}$ is $\simeq \exp(-M/T) << 1$, which greatly suppresses inverse decay processes ($\Gamma_{ID} \approx \exp(-M/T)\Gamma_D << H$). Fig. 2.1 summarizes the time evolution of X; Fig. 2.2 shows the relationship of the various rates (Γ_D, Γ_{ID}, and H) as a function of $M/T(\propto t^{1/2})$.

Now consider the decay of X and $\bar{X}$ bosons: suppose X decays to channels 1 and 2 with baryon numbers B_1 and B_2, and branching ratios r and (1−r). Denote the corresponding quantities for $\bar{X}$ by $-B_1$, $-B_2$, $\bar{r}$, and $(1-\bar{r})$ [e.g. $1 = (\bar{q}\bar{q})$, $2 = (q\ell)$, $B_1 = -2/3$, and $B_2 = 1/3$]. The mean net baryon number of the decay products of the X and $\bar{X}$ are, respectively, $B_X = rB_1 + (1-r)B_2$ and $B_{\bar{X}} = -\bar{r}B_1-(1-\bar{r})B_2$. Hence the decay of an X, $\bar{X}$ pair on average produces a baryon number ε,

$$\epsilon \equiv B_X + B_{\bar{X}} = (r-\bar{r})(B_1-B_2). \tag{2.3}$$

If $B_1 = B_2$, $\epsilon = 0$. In this case X could have been assigned a baryon number B_1, and B would not be violated by X, $\bar{X}$ bosons.

It is simple to show that $r = \bar{r}$ unless both C and CP are violated. Let $\bar{X}$ = the charge conjugate of X, and $r_\uparrow$, $r_\downarrow$, $\bar{r}_\uparrow$, $\bar{r}_\downarrow$ denote the respective branching ratios in the upward and downward directions. [For simplicity, I have reduced the angular degree of freedom to up and down.] The quantities r and $\bar{r}$ are branching ratios averaged over angle: $r = (r_\uparrow+r_\downarrow)/2$, $\bar{r} = (\bar{r}_\uparrow+\bar{r}_\downarrow)/2$ and $\epsilon = (r_\uparrow-\bar{r}_\uparrow+r_\downarrow-\bar{r}_\downarrow)/2$. If C is conserved,

$r_\uparrow = \bar{r}_\uparrow$ and $r_\downarrow = \bar{r}_\downarrow$, and $\varepsilon = 0$. If CP is conserved $r_\uparrow = \bar{r}_\downarrow$ and $r_\downarrow = \bar{r}_\uparrow$, and once again $\varepsilon = 0$.

When the X, $\bar{X}$ bosons decay ($T \ll M$, $t \approx \tau$) $n_X = n_{\bar{X}} \approx n_\gamma$. Therefore, the net baryon number density produced is $n_B \approx \varepsilon n_\gamma$. The entropy density $s \approx g_* n_\gamma$, and so the baryon asymmetry produced is $n_B/s \approx \varepsilon/g_* \approx 10^{-2}\ \varepsilon$.

Recall that the condition for a departure from equilibrium to occur is $K \doteq (\Gamma_D/H)|_{T=M} < 1$ or $M > g_*^{-1/2}\ \alpha m_{pl}$. If X is a gauge boson then $\alpha \approx 1/45$, and so M must be $\gtrsim 10^{16}$ GeV. If X is a Higgs boson, then α is essentially arbitrary, although $\alpha \approx (m_f/M_W)^2\ \alpha_{gauge} \approx 10^{-3} - 10^{-6}$ if the X is in the same representation as the light Higgs bosons responsible for giving mass to the fermions (here m_f = fermion mass, M_W = mass of the W boson ≈ 83 GeV). It is apparently easier for Higgs bosons to satisfy this mass condition than it is for gauge bosons. If $M > g_*^{-1/2}\ \alpha m_{pl}$, then only a modest C, CP-violation ($\varepsilon \approx 10^{-8}$) is necessary to explain $n_B/s \approx (6-10) \times 10^{-11}$. As I will discuss below ε is expected to be larger for a Higgs boson than for a gauge boson. For both these reasons a Higgs boson is the more likely candidate for producing the baryon asymmetry.

Numerical Results

Boltzmann equations for the evolution of n_B/s have been derived and solved numerically in refs. 43, 44. They basically confirm the correctness of the qualitative picture discussed above, albeit, with some important differences. The results can best be discussed in terms of

$$K \doteq \Gamma_D/2H(M) \doteq \alpha m_{pl}/3 g_*^{1/2} M, \qquad (2.4)$$

$$\doteq 3 \times 10^{17}\ \alpha\ \text{GeV}/M.$$

K measures the effectiveness of decays, i.e., rate relative to the expansion rate. K measures the effectiveness of B-nonconserving processes in general because the decay rate characterizes the rates in general for B nonconserving processes, for $T \lesssim M$ (when all the action happens):

$$\Gamma_{ID} \doteq (M/T)^{3/2}\ \exp(-M/T)\ \Gamma_D, \qquad (2.5)$$

$$\Gamma_S \doteq A\alpha(T/M)^5\ \Gamma_D, \qquad (2.6)$$

where Γ_{ID} is the rate for inverse decays (ID), and Γ_S is the rate for 2 ↔ 2 B nonconserving scatterings (S) mediated by X. [A is a numerical factor which depends upon the number of scattering channels, etc, and is typically O(100-1000).]

[It is simple to see why $\Gamma_S \propto \alpha(T/M)^5\Gamma_D \propto \alpha^2 T^5/M^4$. $\Gamma_S \doteq n(\sigma v)$; $n \doteq T^3$ and for $T < M$, $(\sigma v) \propto \alpha^2 T^2/M^4$. Note, in some supersymmetric GUTs, there exist fermionic partners of superheavy Higgs which mediate $\not{B}$ (and also lead to dim-5 $\not{B}$ operators). In this case $(\sigma v) \propto \alpha^2/M^2$ and $\Gamma_S \doteq A\alpha(T/M)^3\Gamma_D$, and 2 ↔ 2 $\not{B}$ scatterings are much more important.]

The time evolution of the baryon asymmetry (n_B/s vs $z = M/T \propto t^{1/2}$) and the final value of the asymmetry which evolves are shown in Figs. 2.3 and 2.4 respectively. For $K < 1$ all B nonconserving processes are ineffective (rate < H) and the asymmetry which evolves is just ε/g_* (as predicted in the qualitative picture). For $K_c > K > 1$, where K_c is determined by

$$K_c \ (\ell n\ K_c)^{-2.4} = 300/A\alpha, \tag{2.7}$$

S 'freeze out' before IDs and can be ignored. Equilibrium is maintained to some degree (by Ds and IDs), however a sizeable asymmetry still evolves

$$n_B/s \simeq (\varepsilon/g_*)\ 0.3\ K^{-1}(\ell n K)^{-0.6}. \tag{2.8}$$

This is the surprising result: for $K_c > K >> 1$, equilibrium is not well maintained and a significant n_B/s evolves, whereas the qualitative picture would suggest that for $K >> 1$ no asymmetry should evolve. For $K > K_c$, S are very important, and the n_B/s which evolves becomes exponentially small:

$$n_B/s \simeq (\varepsilon/g_*)(AK\alpha)^{1/2}\ \exp[-4/3\ (AK\alpha)^{1/4}]. \tag{2.9}$$

[In supersymmetric models which have dim-5 $\not{B}$ operators, $K_c(\ell n K_c)^{-1.2} \simeq 18/A\alpha$ and the analog of Eqn. 2.9 for $K > K_c$ is: $n_B/s \simeq (\varepsilon/g_*)\ A\alpha K \exp[-2(A\alpha K)^{1/2}]$.]

For the XY gauge bosons of SU(5) $\alpha \simeq 1/45$, $A \simeq$ few x 10^3, and $M \simeq$ few x 10^{14} GeV, so that $K_{XY} \simeq 0(30)$ and $K_c \simeq 100$. The asymmetry which could evolve due to these bosons is $\simeq 10^{-2}\ (\varepsilon_{XY}/g_*)$. For a color triplet Higgs $\alpha_H \simeq 10^{-3}$ (for a top quark mass of 40 GeV) and $A \simeq$ few x 10^3, leading to $K_H \simeq 3 \times 10^{14}$ GeV/M_H and $K_c \simeq$ few x 10^3. For $M_H \lesssim 3 \times 10^{14}$ GeV, $K_H < 1$ and the asymmetry which could evolve is $\simeq \varepsilon_H/g_*$.

Very Out-of-Equilibrium Decay

If the X boson decays very late, when $M >> T$ and $\rho_X > \rho_{rad}$, the additional entropy released in its decays must be taken into account. This is very easy to do. Before the Xs decay, $\rho = \rho_X + \rho_{rad} \simeq \rho_X = Mn_X$. After they decay $\rho_X \simeq \rho_{rad} = (\pi^2/30)g_*\ T_{RH}^4 = (3/4)sT_{RH}$ (s, T_{RH} = entropy density and temperature after the X decays). As usual assume that on average each decay produces a mean net baryon number ε. Then the resulting n_B/s produced is

$$n_B/s = \varepsilon n_X/s,$$

$$= (3/4)\varepsilon\ T_{RH}/M \tag{2.10}$$

[Note, I have assumed that when the Xs decay $\rho_X >> \rho_{rad}$ so that the initial entropy can be ignored compared to entropy produced by the decays; this assumption guarantees that $T_{RH} \lesssim M$. I have also assumed that $T << M$ so that IDs and S processes can be ignored. Finally, note that how the Xs produce a baryon number of ε per X is irrelevant; it could be by X → q's ℓ's, or equally well by X → ϕs → q's ℓ's (ϕ = any other particle species).]

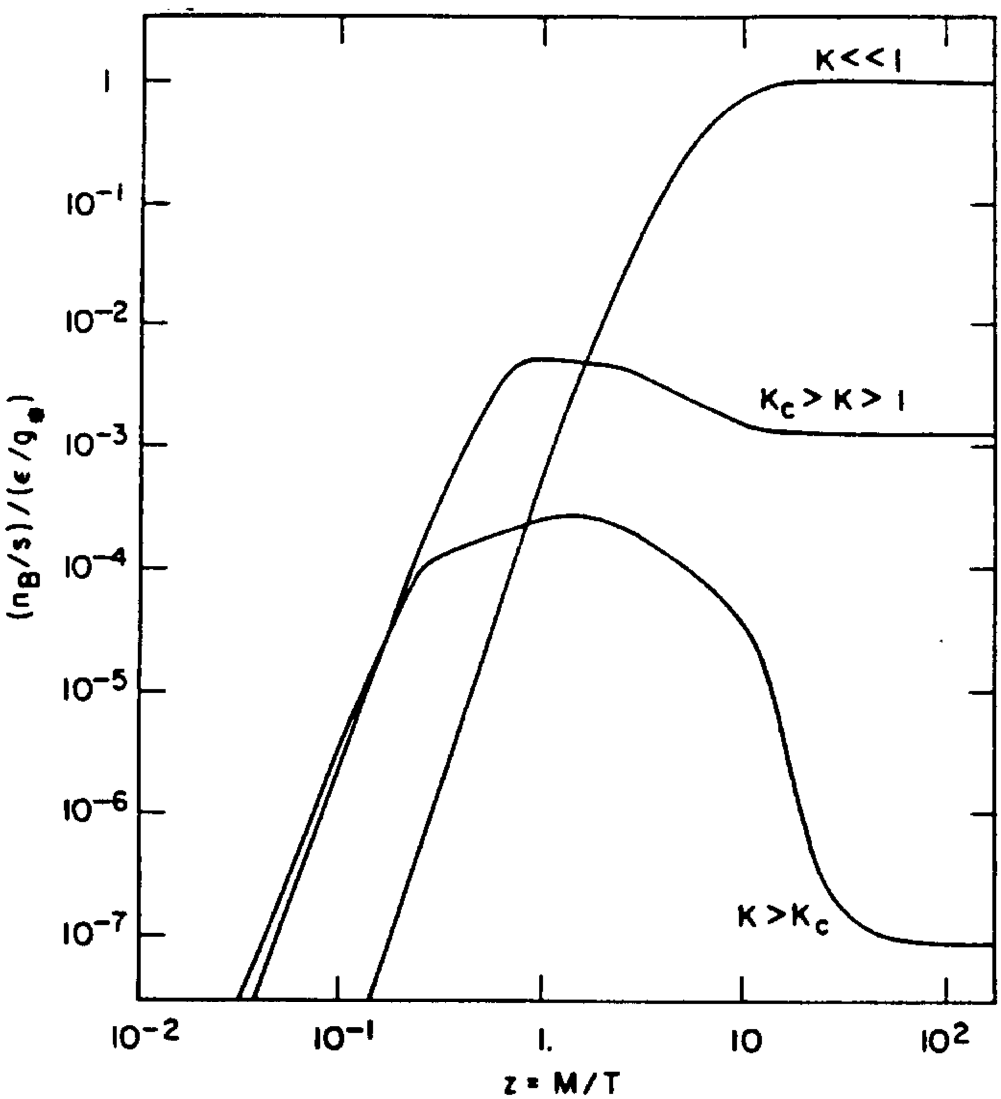

Fig. 2.3 Evolution of n_B/s as a function of $z = M/T$ ($\sim t^{1/2}$). For $K \ll 1$, n_B/s is produced when Xs decay out-of-equilibrium ($z \gg 1$). For $K_c > K > 1$, $n_B/s \propto z^{-1}$ (due to IDs) until the IDs freeze out ($z \simeq 10$). For $K > K_c$ 2 ↔ 2 scatterings are important, and n_B/s decreases very rapidly until they freeze out.

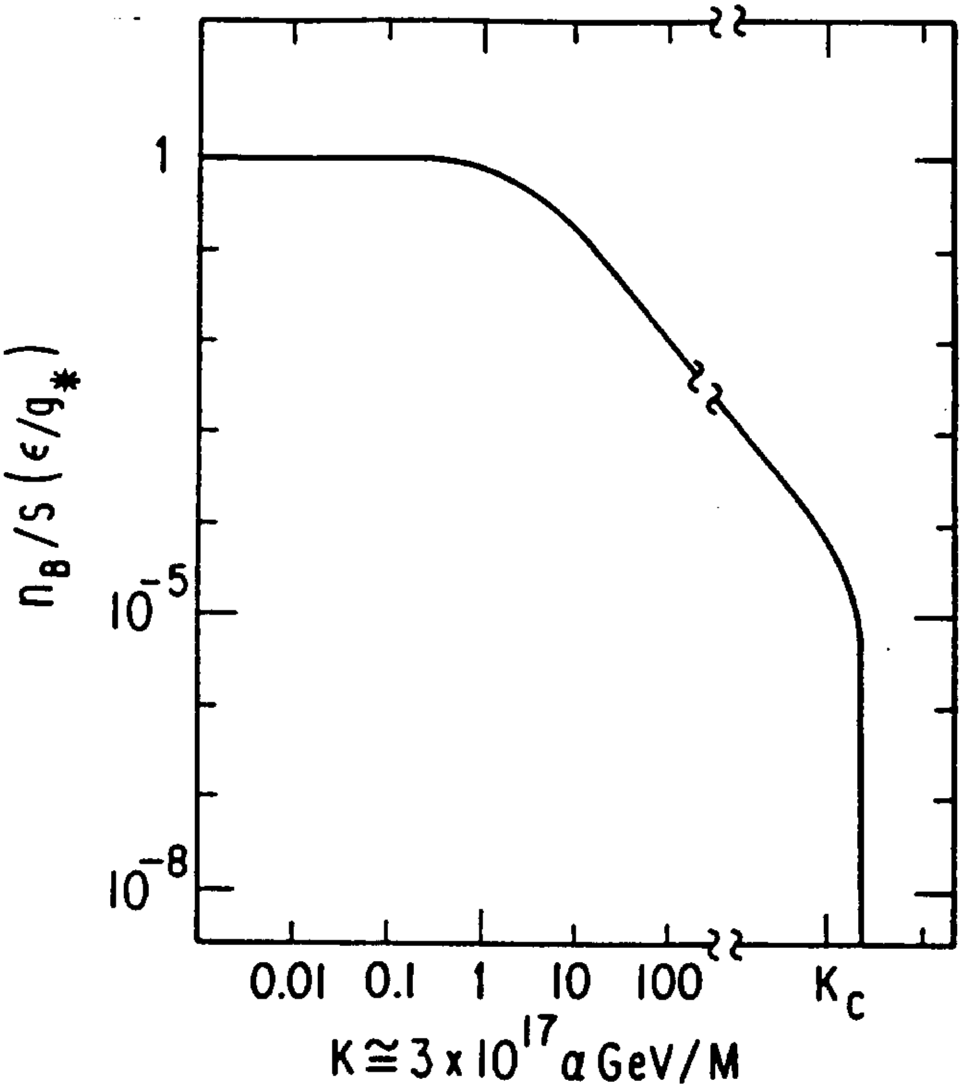

Fig. 2.4 The final baryon asymmetry (in units of ε/g_*) as a function of $K \simeq 3 \times 10^{17}\, \alpha$ GeV/M. For $K \lesssim 1$, n_B/s is independent of K and $\simeq \varepsilon/g_*$. For $K_c > K > 1$, n_B/s decreases slowly, $\propto 1/(K(\ell nK)^{0.6})$. For $K > K_c$ (when $2 \leftrightarrow 2$ scatterings are important), n_B/s decreases exponentially with $K^{1/4}$.

Note that the asymmetry produced depends upon the ratio T_{RH}/M and not T_{RH} itself--this is of some interest in inflationary scenarios in which the Universe does not reheat to a high enough temperature for baryogenesis to proceed in the standard way (out-of- equilibrium decays). For reference T_{RH} can be calculated in terms of $\tau_X = \Gamma^{-1}$; when the Xs decay ($t \simeq \tau_X$, $H \simeq t^{-1} \simeq \Gamma$): $\Gamma^2 \simeq H^2 \simeq 8\pi\rho_X/3m_{pl}^2$. Using the fact that $\rho_X \simeq g_*(\pi^2/30)T_{RH}^4$ it follows that

$$T_{RH} \simeq g_*^{-1/4} (\Gamma m_{pl})^{1/2} \qquad (2.11)$$

The C,CP Violation ε

The crucial quantity for determining n_B/s is ε--the C, CP violation in the superheavy boson system. Lacking 'The GUT', ε cannot be calculated precisely, and hence n_B/s cannot be predicted, as, for example, the ^{4}He abundance can be.

The quantity $\varepsilon \propto (r-\bar{r})$; at the tree graph (i.e., Born approximation) level $r-\bar{r}$ must vanish. Non-zero contributions to $(r-\bar{r})$ arise from higher order loop corrections due to Higgs couplings which are complex.[41,45,46] For these reasons, it is generally true that:

$$\varepsilon_{Higgs} \lesssim O(\alpha^N) \sin\delta, \qquad (2.12)$$

$$\varepsilon_{gauge} \lesssim O(\alpha^{N+1}) \sin\delta, \qquad (2.13)$$

where α is the coupling of the particle exchanged in loop (i.e., $\alpha \simeq g^2/4\pi$), $N \geq 1$ is the number of loops in the diagrams which make the lowest order, non-zero contributions to $(r-\bar{r})$, and δ is the phase of some complex coupling. The C, CP violation in the gauge boson system occurs at 1 loop higher order than in the Higgs because gauge couplings are necessarily real. Since $\alpha \lesssim \alpha_{gauge}$, ε is at most $O(10^{-2})$--which is plenty large enough to explain $n_B/s \simeq 10^{-10}$. Because K for a Higgs is likely to be smaller, and because C, CP violation occurs at lower order in the Higgs boson system, the out-of-equilibrium decay of a Higgs is the more likely mechanism for producing n_B/s. [No additional cancellations occur when calculating $(r-\bar{r})$ in supersymmetric theories, so these generalities also hold for supersymmetric GUTs.]

In minimal SU(5)--one $\underline{5}$ and one $\underline{24}$ of Higgs, and three families of fermions, $N \simeq 3$. This together with the smallness of the relevant Higgs couplings implies that $\varepsilon_H \lesssim 10^{-15}$ which is not nearly enough.[41,45,46] With 4 families the relevant couplings can be large enough to obtain $\varepsilon_H \simeq 10^{-8}$--if the top quark and fourth generation quark/lepton masses are $O(m_W)$ (ref. 47). By enlarging the Higgs sector (e.g., by adding a second $\underline{5}$ or a $\underline{45}$), $(r-\bar{r})$ can be made non-zero at the 1-loop level, making $\varepsilon_H \simeq 10^{-8}$ easy to achieve.

In more complicated theories, e.g., E6, S(10), etc., $\varepsilon \simeq 10^{-8}$ can also easily be achieved. However, to do so restricts the possible symmetry breaking patterns. Both E6 and SO(10) are C-symmetric, and of course C-symmetry must be broken before ε can be non-zero. In general, in these models ε is suppressed by powers of M_C/M_G where M_C (M_G) is the scale of C(GUT) symmetry breaking, and so M_C cannot be significantly smaller than M_G.

It seems very unlikely that ε can be related to the parameters of the K^0-$\bar{K}^0$ system, the difficulty being that not enough C, CP violation can be fed up to the superheavy boson system. It has been suggested that ε could be related to the electric dipole moment of the neutron.[48]

Although baryogenesis is nowhere near being on the same firm footing as primordial nucleosynthesis, we now at least have for the first time a very attractive framework for understanding the origin of $n_B/s \simeq 10^{-10}$. A framework which is so attractive, that in the absence of observed proton decay, the baryon asymmetry of the Universe is probably the best evidence for some kind of quark/lepton unification. [In writing up this lecture I have borrowed freely and heavily from the review on baryogenesis written by myself and E. W. Kolb (ref.49) and refer the interested reader there for a more thorough discussion of the details of baryogenesis.]

LECTURE 3: MONOPOLES, COSMOLOGY, AND ASTROPHYSICS

Birth: Glut or Famine

In 1931 Dirac[50] showed that if magnetic monopoles exist, then the single-valuedness of quantum mechanical wavefunctions require the magnetic charge of a monopole to satisfy the quantization condition

$$g = ng_D, \; n = 0, \pm 1, \pm 2 \ldots$$

$$g_D = 1/2e \simeq 69e.$$

However, one is not required to have Dirac monopoles in the theory--you can take 'em or leave 'em! In 1974 't Hooft[51] and Polyakov[52] independently made a remarkable discovery. They showed that monopoles are obligatory in the low-energy theory whenever a semi-simple group G, e.g., SU(5), breaks down to a group G' x U(1) which contains a U(1) factor [e.g., SU(3) x SU(2) x U(1)]; this, of course, is the goal of unification. These monopoles are associated with nontrivial topology in the Higgs field responsible for SSB, topological knots if you will, have a mass $m_M \simeq O(M/\alpha)$ [$\simeq 10^{16}$ GeV in SU(5); M = scale of SSB], and have a magnetic charge which is a multiple of the Dirac charge.

Since there exist no contemporary sites for producing particles of mass even approaching 10^{16} GeV, the only plausible production site is the early Universe, about 10^{-34} s after 'the bang' when the temperature was $\simeq O(10^{14}$ GeV). There are two ways in which monopoles can be produced: (1) as topological defects during the SSB of the unified group G; (2) in monopole-antimonopole pairs by energetic particle collisions. The first process has been studied by Kibble[53], Preskill[54], and Zel'dovich and Khlopov[55], and I will review their important conclusions here.

The magnitude of the Higgs field responsible for the SSB of the unified group G is determined by the minimization of the free energy. However, this does not uniquely specify the direction of the Higgs field in group space. A monopole corresponds to a configuration in which the direction of the Higgs field in group space at different points in physical space is topologically distinct from the configuration in which the Higgs field points in the same direction (in group space) everywhere in physical space (which corresponds to no monopole):

→ = direction of Higgs field in group space

↑ ↑ ↑ ↑ ↑ ↑	↑ ← → ↓
<u>no monopole</u>	<u>monopole</u>

Clearly monopole configurations cannot exist until the SSB [$G \rightarrow G' \times U(1)$] transition takes place. When spontaneous symmetry breaking occurs, the Higgs field can only be smoothly oriented (i.e., the no monopole configuration) on scales smaller than some characteristic correlation length ξ. On the microphysical side, the inverse Higgs mass at the Ginzburg temperature (T_G) sets such a scale: $\xi \simeq m_H^{-1}(T_G)$ (in a second-order phase transition)[56]. [The Ginzburg temperature is the temperature below which it becomes improbable for the Higgs field to fluctuate between the SSB minimum and $\phi = 0$.] Cosmological considerations set an absolute upper bound: $\xi \lesssim d_H (\simeq t$ in the standard cosmology). [Note, even if the horizon distance $d_H(t)$ diverges, e.g., because $R \propto t^n$ ($n > 1$) for $t \lesssim t_{pl}$, the physics horizon H^{-1} sets an absolute upper bound on ξ, which is numerically identical.] On scales larger than ξ the Higgs field must be uncorrelated, and thus we expect of order 1 monopole per correlation volume ($\simeq \xi^3$) to be produced as a topological defect when the Higgs field freezes out.

Let's focus on the case where the phase transition is either second order or weakly-first order. Denote the critical temperature for the transition by T_c ($\simeq O(M)$), and as before the monopole mass by $m_M \simeq O(M/\alpha)$. The age of the Universe when $T \simeq T_c$ is given in the standard cosmology by: $t_c \simeq 0.3\, g_*^{-1/2} m_{pl}/T_c^2$, cf. Eqn. 1.12. For SU(5): $T_c \simeq 10^{14}$ GeV, $m_M \simeq 10^{16}$ GeV and $t_c \simeq 10^{-34}$ s. Due to the fact that the freezing of the Higgs field <u>must</u> be uncorrelated on scales $\gtrsim \xi$, we expect an initial monopole abundance of O(1) per correlation volume; using $d_H(t_c)$ as an <u>absolute</u> upper bound on ξ this leads to: $(n_M)_i \simeq O(1)\, t_c^{-3}$. Comparing this to our fiducials S_{HOR} and N_{B-HOR}, we find that the initial monopole-to-entropy and monopole-to-baryon number ratios are:

$$n_M/s \gtrsim 10^2\, (T_c/m_{pl})^3, \tag{3.1a}$$

$$n_M/n_B \gtrsim 10^{12}\, (T_c/m_{pl})^3. \tag{3.1b}$$

[Note: $\langle F_M \rangle$, the average monopole flux in the Universe, and Ω_M, the fraction of critical density contributed by monopoles, are related to n_M/s and n_M/n_B by:

$$\langle F_M \rangle \simeq 10^{10}(n_M/s)\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{sec}^{-1}\ , \tag{3.2a}$$

$$\simeq (n_M/n_B)\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{sec}^{-1}\ , \tag{3.2b}$$

$$\Omega_M h^2 \simeq 10^{24}(n_M/s)(m_M/10^{16}\mathrm{GeV})\ , \tag{3.3a}$$

$$\simeq 10^{14} \langle F_M \rangle (m_M/10^{16}\,\mathrm{GeV}) \ , \qquad (3.3b)$$

where the monopole velocity has been assumed to be $\simeq 10^{-3}c$ (this assumption will be discussed in detail later).

Preskill[54] has shown that unless n_M/s is $> 10^{-10}$ monopole-antimonopole annihilations do not significantly reduce the initial monopole abundance. If $n_M/s > 10^{-10}$, he finds that n_M/s is reduced to $\simeq 10^{-10}$ by annihilations. For $T_c < 10^{15}$ GeV our estimate for n_M/s is $< 10^{-10}$, and we will find that in the standard cosmology T_c must be $\ll 10^{15}$ GeV to have an acceptable monopole abundance, so for our purposes we can ignore annihilations. Assuming that the expansion has been adiabatic since $T \simeq T_c$, this estimate for n_M/s translates into:

$$\langle F_M \rangle \simeq 10^{-3}\ (T_c/10^{14}\ \mathrm{GeV})^3\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1}, \qquad (3.4a)$$

$$\Omega_M \simeq 10^{11}\ (T_c/10^{14}\ \mathrm{GeV})^3 (m_M/10^{16}\ \mathrm{GeV}) \qquad (3.4b)$$

--a flux that would make any monopole hunter/huntress ecstatic, and an Ω_M that is unacceptably large (except for $T_c \ll 10^{14}$ GeV). As was discussed previously, Ω can be at most O(few), so we have a very big problem with the simplest GUTs (in which $T_c \simeq 10^{14}$ GeV). This is the so-called 'Monopole Problem'. The statement that $\Omega_M \simeq 10^{11}$ for $T_c \simeq 10^{14}$ GeV is a bit imprecise; clearly if $k < 0$ (corresponding to $\Omega < 1$) monopole production cannot close the Universe (and in the process change the geometry from being infinite in extent and negatively-curved, to being finite in extent and positively-curved). More precisely, a large monopole abundance would result in the Universe becoming matter-dominated much earlier, at $T \simeq 10^3$ GeV $(T_c/10^{14}\ \mathrm{GeV})^3\ (m_M/10^{16}\ \mathrm{GeV})$, and eventually reaching a temperature of 3 K at the young age of $t \simeq 10^4\ \mathrm{yrs}(T_c/10^{14}\ \mathrm{GeV})^{-3/2}\ (m_M/10^{16}\ \mathrm{GeV})^{-1/2}$. The requirement that $\Omega_M \lesssim$ O(few) implies that

$$T_c \lesssim 10^{11}\ \mathrm{GeV} \qquad (\Omega_M \lesssim \mathrm{few})$$

where I have taken m_M to be O(100 T_c). Note, given the generous estimate for ξ, even this is probably not safe; if one had a GUT in which $T_c \simeq 10^{11}$ GeV a more careful estimate for ξ would be called for.

The Parker bound (to be discussed below) on the average monopole flux in the galaxy, $\langle F_M \rangle \lesssim 10^{-15}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1}$, results in a slightly more stringent constraint:

$$T_c \lesssim 10^{10}\ \mathrm{GeV} \qquad (\text{Parker bound})$$

The most restrictive constraints on T_c follow from the neutron star catalysis bounds on the monopole flux (also to be discussed below) and the most restrictive of those, $\langle F_M \rangle \lesssim 10^{-27}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1}$, implies that

$$T_c \lesssim 10^{6}\ \mathrm{GeV} \qquad (\text{Neutron star catalysis bound})$$

Note, to obtain these bounds I have compared my estimate for the average monopole flux in the Universe, Eqn. 3.4a, with the astrophysical bounds on the average flux of monopoles in our galaxy. If monopoles cluster in galaxies (which I will later argue is <u>unlikely</u>), then the average galactic flux of monopoles is greater than the average flux of monopoles in the Universe, making the above bounds on T_c <u>more</u> restrictive.

If the GUT transition is strongly first order (I am excluding inflationary Universe scenarios for the moment), then the transition will proceed by bubble nucleation at a temperature T_n ($\ll T_c$), when the nucleation rate becomes comparable to the expansion rate H. Within each bubble the Higgs field is correlated; however, the Higgs field in different bubbles should be uncorrelated. Thus one would expect O(1) monopole per bubble to be produced. When the Universe supercools to a temperature T_n, bubbles nucleate, expand, and rapidly fill all of space; if r_b is the typical size of a bubble when this occurs, then one expects n_M to be $\simeq r_b^{3}$. After the bubbles coalesce, and the Universe reheats, the entropy density is once again $s \simeq g_* T_c^{3}$, so that the resulting monopole to entropy ratio is: $n_M/s \simeq (g_* r_b^{3} T_c^{3})^{-1}$. Guth and Weinberg[57] have calculated r_b and find that $r_b \simeq (m_{pl}/T_c^{2})/\ln(m_{pl}^{4}/T_c^{4})$, leading to a relatively accurate estimate for the monopole abundance:

$$n_M/s \simeq [\ln(m_{pl}^{4}/T_c^{4})(T_c/m_{pl})]^{3} \quad , \tag{3.5}$$

which is even more disasterous than the estimate for a second order phase transition [recall, however, estimate 3.1 was an absolute <u>lower</u> bound].

The bottom line is that we have a serious problem here--the standard cosmology extrapolated back to $T \simeq T_c$ and the simplest GUTs are incompatible (to say the least). One (or both) must be modified. Although this result is discouraging (especially when viewed in the light of the great success of baryogenesis), it does provide a valuable piece of information about physics at very high energies and/or the earliest moments of the Universe, in that regard a 'window' to energies $\gtrsim 10^{14}$ GeV and times $\lesssim 10^{-34}$ sec.

A number of possible solutions have been suggested. To date the most attractive is the new inflationary Universe scenario (which will be the subject of Lecture 4). In this scenario, a small region (size $\lesssim$ the horizon) within which the Higgs field could be correlated, grows to a size which encompasses all of the presently observed Universe, due to the exponential expansion which occurs during the phase transition. This results in less than one monopole in the entire observable Universe (due to Kibble production).

Let me very briefly review some of the other attempts to solve the monopole problem. Several people have pointed out that if there is no complete unification [e.g., if $G = H \times U(1)$], or if the full symmetry of the GUT is not restored in the very early Universe (e.g., if the maximum temperature the Universe reached was $< T_c$, or if a large lepton number[58], $n_L/n_\gamma > 1$, prevented symmetry restoration at high temperature), then there would be no monopole problem. However, none of these possibilities seems particularly attractive.

Several authors[59-62] have studied the possibility that monopole-antimonopole annihilation could be enhanced over Preskill's estimate, due to 3-body annihilations or the gravitational clumping of monopoles (or both). Thus far, this approach has not solved the problem.

Bais and Rudaz[63] have suggested that large fluctuations in the Higgs field at temperatures near T_c could allow the monopole density to relax to an acceptably small value. They do not explain how this mechanism can produce the acausal correlations needed to do this.

Scenarios have been suggested in which monopoles and antimonopoles form bound pairs connected by flux tubes, leading to rapid monopole-antimonopole annihilation. For example, Linde[64] proposed that at high temperatures color magnetic charge is confined, and Lazarides and Shafi[65] proposed that monopoles and antimonopoles become connected by Z^0 flux tubes after the SU(2) x U(1) SSB phase transition. In both cases, however, the proposed flux tubes are not topologically stable, nor has their existence even been demonstrated.

Langacker and Pi[66] have suggested a solution which does seem to work. It is based upon an unusual (although perhaps contrived) symmetry breaking pattern for SU(5):

$$\mathrm{SU}(5) \xrightarrow[T_c \simeq 10^{14}\,\mathrm{GeV}]{} \mathrm{SU}(3)\times \mathrm{SU}(2)\times \mathrm{U}(1) \xrightarrow[T_1]{} \mathrm{SU}(3) \xrightarrow[T_2]{} \mathrm{SU}(3)\times \mathrm{U}(1)$$

$$\underbrace{T_1 \longleftrightarrow T_2}_{\text{superconducting phase}}$$

(note T_1 could be equal to T_c). The key feature of their scenario is the existence of the epoch ($T = T_1 \rightarrow T_2$) in which the U(1) of electromagnetism is spontaneously broken (a superconducting phase); during this epoch magnetic flux must be confined to flux tubes, leading to the annihilation of the monopoles and antimonopoles which were produced earlier on, at the GUT transition. Although somewhat contrived, their scenario appears to be viable (however, I'll have more to say about it shortly).

Finally, one could invoke the Tooth Fairy (in the guise of a perfect annihilation scheme). E. Weinberg[67] has recently made a very interesting point regarding 'perfect annihilation schemes', which applies to the Langacker-Pi scenario[66], and even to a Tooth Fairy which operates causally. Although the Kibble mechanism results in equal numbers of monopoles and antimonopoles being produced, E. Weinberg points out that in a finite volume there can be magnetic charge fluctuations. He shows that if the Higgs field 'freezes out' at $T \simeq T_c$ and is uncorrelated on scales larger than the horizon at that time, then the expected net RMS magnetic charge in a volume V which is much bigger than the horizon is

$$\Delta n_M \simeq (V/t_c{}^3)^{1/3}. \qquad (3.6)$$

He then considers a perfect, causal annihilation mechanism which operates from $T = T_1 \rightarrow T_2$ (e.g., formation of flux tubes between monopoles and antimonopoles). At <u>best</u>, this mechanism could reduce the monopole abundance down to the net RMS magnetic charge contained in the horizon at $T = T_2$, leaving a final monopole abundance of

$$n_M/s = 10^2\ T_c T_2{}^2/m_{pl}{}^3, \tag{3.7}$$

resulting in

$$\Omega_M \gtrsim 0.1(T_c/10^{14}\ \mathrm{GeV})(m_M/10^{16}\ \mathrm{GeV})(T_2/10^8\ \mathrm{GeV})^2, \tag{3.8a}$$

$$\langle F_M \rangle \gtrsim 10^{-15}(T_c/10^{14}\mathrm{GeV})(T_2/10^8\mathrm{GeV})^2\mathrm{cm}^{-2}\mathrm{sr}^{-1}\mathrm{s}^{-1}. \tag{3.8b}$$

It is difficult to imagine a perfect annihilation mechanism which could operate at temperatures $\lesssim 10^3$ GeV, without having to modify the standard SU(2) x U(1) electroweak theory; for $T_c = 10^{14}$ GeV and $T_2 = 10^3$ GeV, E. Weinberg's argument[67] implies that $\langle F_M \rangle$ must be $\gtrsim 10^{-25}$ cm^{-2} sr^{-1} s^{-1}, which would be in conflict with the most stringent neutron star catalysis bound, $F_M < 10^{-27}$ cm^{-2} sr^{-1} s^{-1}.

Finally, I should emphasize that the estimate of n_M/s based upon $\xi \lesssim d_H(t)$ is an absolute (and very generous) lower bound to n_M/s. Should a model be found which succeeds in suppressing the monopole abundance to an acceptable level (e.g., by having $T_c << 10^{14}$ GeV or by a perfect annihilation epoch), then the estimate for ξ must be refined and scrutinized.

If the glut of monopoles produced as topological defects in the standard cosmology can be avoided, then the only production mechanism is pair production in very energetic particle collisions, e.g., particle(s) + antiparticle(s) → monopole + antimonopole. [Of course, the 'Kibble production' of monopoles might be consistent with the standard cosmology (and other limits to the monopole flux) if the SSB transition occurred at a low enough temperature, say << $O(10^{10}$ GeV).] The numbers produced are intrinsically small because monopole configurations do not exist in the theory until SSB occurs (T_c = M = scale of SSB), and have a mass $O(M/\alpha) = 100\ M = 100\ T_c$. For this reason they are never present in equilibrium numbers; however, some are produced due to the rare collisions of particles with sufficient energy. This results in a present monopole abundance of [68-70]

$$n_M/s = 10^2\ (m_M/T_{max})^3\ \exp(-2m_M/T_{max}), \tag{3.9a}$$

$$\Omega_M = 10^{26}(m_M/10^{16}\mathrm{GeV})(m_M/T_{max})^3\exp(-2m_M/T_{max}), \tag{3.9b}$$

$$\langle F_M \rangle = 10^{12}\mathrm{cm}^{-2}\mathrm{sr}^{-1}\mathrm{s}^{-1}(m_M/T_{max})^3\ \exp(-2m_M/T_{max}), \tag{3.9c}$$

where T_{max} is the highest temperature reached after SSB.

In general, $m_M/T_{max} = O(100)$ so that $\Omega_M = O(10^{-40})$ and $\langle F_M \rangle = O(10^{-32}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1})$--a negligible number of monopoles. However, the number produced is exponentially sensitive to m_M/T_{max}, so that a factor of 3-5 uncertainty in m_M/T_{max} introduces an enormous uncertainty in the predicted production. For example, in the new inflationary Universe, the monopole mass can be ∝ the Higgs field responsible for SSB, and as that field oscillates about the SSB minimum during the reheating process m_M

also oscillates, leading to enhanced monopole production [m_M/T_{max} in Eqns. 3.9a,b,c is replaced by fm_M/T_{max}, where f < 1 depends upon the details of reheating; see refs. 71, 72].

Cosmology seems to leave the poor monopole hunter/huntress with two firm predictions: that there should be equal numbers of north and south poles; and that either far too few to detect, or far too many to be consistent with the standard cosmology should have been produced. The detection of any superheavy monopoles would necessarily send theorists back to their chalkboards!

From Birth Through Adolescence ($t \simeq 10^{-34}$sec to $t \simeq 3 \times 10^{17}$sec)

As mentioned in the previous section, monopoles and antimonopoles do not annihilate in significant numbers; however, they do interact with the ambient charged particles (e.g., monopole + e^- ↔ monopole + e^-) and thereby stay in kinetic equilibrium (KE = 3T/2) until the epoch of $e^\pm$ annihilations (T ≃ 1/2 MeV, t ≃ 10 s). At the time of $e^\pm$ annihilations monopoles and antimonopoles should have internal velocity dispersions of:

$$\langle v_M^2 \rangle^{1/2} \simeq 30 \text{ cm s}^{-1} (10^{16} \text{ GeV}/m_M)^{1/2}.$$

After this monopoles are effectively collisionless, and their velocity dispersion decays $\propto R(t)^{-1}$, so that if we neglect gravitational and magnetic effects, today they should have an internal velocity dispersion of

$$\langle v_M^2 \rangle^{1/2} \simeq 10^{-8} \text{ cm s}^{-1} (10^{16} \text{ GeV}/m_M)^{1/2}.$$

Since they are collisionless, only their velocity dispersion can support them against gravitational collapse. With such a small velocity dispersion to support them they are gravitationally unstable on all scales of astrophysical interest ($\lambda_{Jeans} \simeq 10^{-10}$ LY).

After decoupling (T ≃ 1/3 eV, t ≃ 10^{13} s) [or the epoch of matter domination in scenarios where the mass of the Universe is dominated by a nonbaryonic component], matter can begin to clump, and structure can start to form. Monopoles, too, should clump and participate in the formation of structure. However, since they cannot dissipate their gravitational energy, they cannot collapse into the more condensed objects (such as stars, planets, the disk of the galaxy, etc.) whose formation clearly must have involved the dissipation of gravitational energy. Thus, one would only expect to find monopoles in structures whose formation did not require dissipation (such as clusters of galaxies, and galactic haloes). However, galactic haloes are not likely to be a safe haven for monopoles in galaxies with magnetic fields; monopoles less massive than about 10^{20} GeV will, in less than 10^{10} yrs, gain sufficient KE from a magnetic field of strength a few x 10^{-6} G to reach escape velocity[73]. We are led to the conclusion that initially monopoles should either be uniformly distributed through the cosmos, or clumped in clusters of galaxies or in the haloes of galaxies with weak or non-existent magnetic fields. Since our own galaxy has a magnetic field of strength $\approx$ few x 10^{-6} G, and is not a member of a cluster of galaxies, we would expect the local flux of monopoles to be not too different from the average monopole flux in the Universe.

Although monopoles initially have a very small internal velocity dispersion, there are many mechanisms for increasing their velocities. First, typical peculiar velocities (i.e., velocities relative to the Hubble flux) are $O(10^{-3}$ c), leading to a typical monopole-galaxy velocity of 10^{-3}c. Monopoles will be accelerated by the gravitational fields of galaxies (to $\approx 10^{-3}$ c $\approx$ orbital velocity in the galaxy), and if they encounter them, clusters of galaxies (to $\approx 3 \times 10^{-3}$ c). A typical monopole, however, will never encounter a galaxy or a cluster of galaxies, the respective mean free paths being: L_{gal} ($\approx 10^{26}$ cm $\approx 10^{-2}$ c x age of the Universe) and $L_{cluster} \approx 3 \times 10^{28}$ cm.

Monopoles will also be accelerated by magnetic fields. The intragalactic magnetic field strength is $\lesssim 3 \times 10^{-11}$ G (ref. 74), and results in a monopole velocity of

$$v_M \approx 3 \times 10^{-4}\ c\ (B/10^{-11}\ G)(10^{16}\ GeV/m_M).$$

The galactic magnetic field will accelerate monopoles in our galaxy to velocities of[73]

$$v_M \approx 3 \times 10^{-3}\ c\ (10^{16}\ GeV/m_M)^{1/2}.$$

Taking all of these 'sources of velocity' into account, we can make an educated estimate of the typical monopole-detector relative velocity (see Table 3.1). From Table 3.1 below it should be clear that the typical monopole should be moving with a velocity of at least a few x 10^{-3} c with respect to an earth-based detector. It goes without saying that 'this fact' is an important consideration for detector design.

Although planets, stars, etc. should be monopole-free at the time of their formation, they will accumulate monopoles during their lifetimes. The number captured by an object is

Table 3.1 Typical Monopole-Detector Relative Velocities

DETECTOR VELOCITY		MONOPOLE VELOCITY	
orbit in galaxy	$2/3 \times 10^{-3}$ c	galactic B-field	3×10^{-3} c $(10^{16}\text{GeV}/m_M)^{1/2}$
orbit in solar system	10^{-4} c	grav. acceleration by galaxy	10^{-3} c
		grav. acceleration by sun	10^{-4} c
		monopole-galaxy relative velocity	10^{-3} c

$$N_M = (4\pi R^2)(\pi\text{-sr})(1 + 2GM/Rv_M^2)\langle F_M\rangle\varepsilon\tau, \qquad (3.10)$$

where M, R and τ are the mass, radius and age of the object, v_M is the monopole velocity, and ε is the efficiency with which the object stops monopoles which strikes its surface. The efficiency of capture ε depends upon the mass and velocity of the monopole, and its rate of energy loss in the object. The quantity $(1 + 2GM/R\, v_M^2)$ is just the ratio of the capture cross section to the geometric cross section. Main sequence stars of mass $(0.6 - 30)M_\odot$ will capture monopoles less massive than about 10^{18} GeV with velocities $\lesssim 10^{-3}$ c with good efficiency ($\varepsilon \simeq 1$); in its main sequence lifetime a star will capture approximately 10^{24} F_{-16} monopoles[75] (essentially independent of its mass). Here $\langle F_M\rangle = F_{-16}$ 10^{-16} cm^{-2} sr^{-1} s^{-1}. Neutron stars will capture monopoles less massive than about 10^{20} GeV with velocities $\lesssim 10^{-3}$ c with unit efficiency, capturing about 10^{21} F_{-16} monopoles in 10^{10} yrs. Planets like Jupiter can stop monopoles less massive than about 10^{16} GeV with velocities < 10^{-3}, accumulating about 10^{22} F_{-16} monopoles in 10^{10} yrs.[76]. A planet like the earth can only stop light or slowly-moving monopoles[76] (for $m_M = 10^{16}$ GeV, v_M must be $\lesssim 3 \times 10^{-5}$ c). Once inside, monopoles can do interesting things, like catalyze nucleon decay (to be discussed below), which keeps the object hot (and leads to a potentially observable photon flux), and eventually depletes the object of all its nucleons. A monopole flux of F_{-21} 10^{-21} cm^{-2} sr^{-1} s^{-1} will cause a neutron star to evaporate in 10^{11} $F_{-21}{}^{-1/2}$ yrs, a Jupiter-like planet to evaporate in 5×10^{15} $F_{-21}{}^{-1/2}$ yrs, and an Earth-like planet to evaporate in 10^{18} $F_{-21}{}^{-1/2}$ yrs[77]. Accretion of monopoles by astrophysical objects, however, does not significantly reduce the monopole flux; the mean free path of a monopole in the galaxy is $\simeq 10^{42}$ cm.

What are Monopoles Doing Today?--Astrophysical Constraints

The three most conspicuous properties of a GUT monopole are: (i) macroscopic mass ($\simeq M/\alpha$--10^{16} GeV $\simeq 10^{-8}$ g for SU(5)); (ii) hefty magnetic charge $h \simeq n\, 69e$ ($n = \pm 1, \pm 2, \ldots$); (iii) the ability to catalyze nucleon decay. Because of these properties, monopoles, if present, should be doing very astrophysically interesting things today--so interesting and so conspicuous that very stringent astrophysical bounds can be placed upon their flux (summarized in Fig. 3.1).

Theoretical prejudice strongly favors the flat cosmological model (i.e., $\Omega = 1$). As I discussed in Lecture 1 big bang nucleosynthesis strongly suggests that baryons contribute $\Omega_b \lesssim 0.15$. In addition, the flat rotation curves of galaxies provide strong evidence that most of the mass associated with a galaxy is dark and exists in an extended structure (most likely a spherical halo). Monopoles are certainly a candidate for the dark matter in galaxies and for providing the closure density.

As I discussed in the first lecture the age of the Universe implies that $\Omega h^2 \lesssim O(1)$; if monopoles are uniformly distributed in the cosmos, then this constrains their average flux to be

$$\langle F_M \rangle \lesssim 10^{-14}\ \mathrm{cm}^{-2}\mathrm{sr}^{-1}\mathrm{s}^{-1}(m_M/10^{16}\ \mathrm{GeV})^{-1}, \qquad (3.11)$$

cf. Eqn. 3.3b. For comparison $10^{-14}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1} \simeq 30$ monopoles $(\mathrm{soccerfield})^{-1}\ \mathrm{yr.}^{-1}$

If monopoles are clustered in galaxies the local galactic flux can be significantly higher. The mass density in the neighborhood of the sun is about $10^{-23}\ \mathrm{gcm}^{-3}$; of this about 1/2 is accounted for (stars, gas, dust, etc.). Monopoles can at most provide the other 1/2, resulting in the flux bound

$$F_M \lesssim 5 \times 10^{-10}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1}\ (m_M/10^{16}\ \mathrm{GeV})^{-1}. \qquad (3.12)$$

Actually the bound is probably at least a factor of 10-30 more stringent. The unseen material has a column density ($= \int \rho dz$) of no more than about $(30\ \mathrm{kpc})(10^{-25}\ \mathrm{gcm}^{-3})$ (as determined by studying the motions of stars in the stellar neighborhood[78]). Since monopoles are effectively collisionless, if present, they would be distributed in an extended spherical halo. Flat rotation curves indicate that the scale of galactic halos is O(30 kpc), so that the local column density of halo material is $\rho_{halo} \times 30$ kpc. Comparing this to the bound on the local column density of unseen material it follows that locally $\rho_{halo} \lesssim 10^{-25}\ \mathrm{g\ cm}^{-3}$. Using this as the limit to the density contributed by monopoles the flux bound 3.12 becomes

$$F_M \lesssim 10^{-11}\mathrm{cm}^{-2}\mathrm{sr}^{-1}\mathrm{s}^{-1}(m_M/10^{16}\mathrm{GeV})^{-1}. \qquad (3.13)$$

A monopole by virtue of its magnetic charge will be accelerated by magnetic fields, and in the process can gain KE. Of course, any KE gained must come from somewhere. Any gain in KE is exactly compensated for by a loss in field energy: $\Delta KE = -\Delta[(B^2/8\pi) \times \mathrm{Vol}]$. Consider a monopole which is initially at rest in a region of uniform magnetic field. It will be accelerated along the field and after moving a distance ℓ the monopole will have

$$KE = hB\ell \simeq 10^{11}\mathrm{GeV}(B/3\times10^{-6}\mathrm{G})(\ell/300\mathrm{pc}), \qquad (3.14)$$

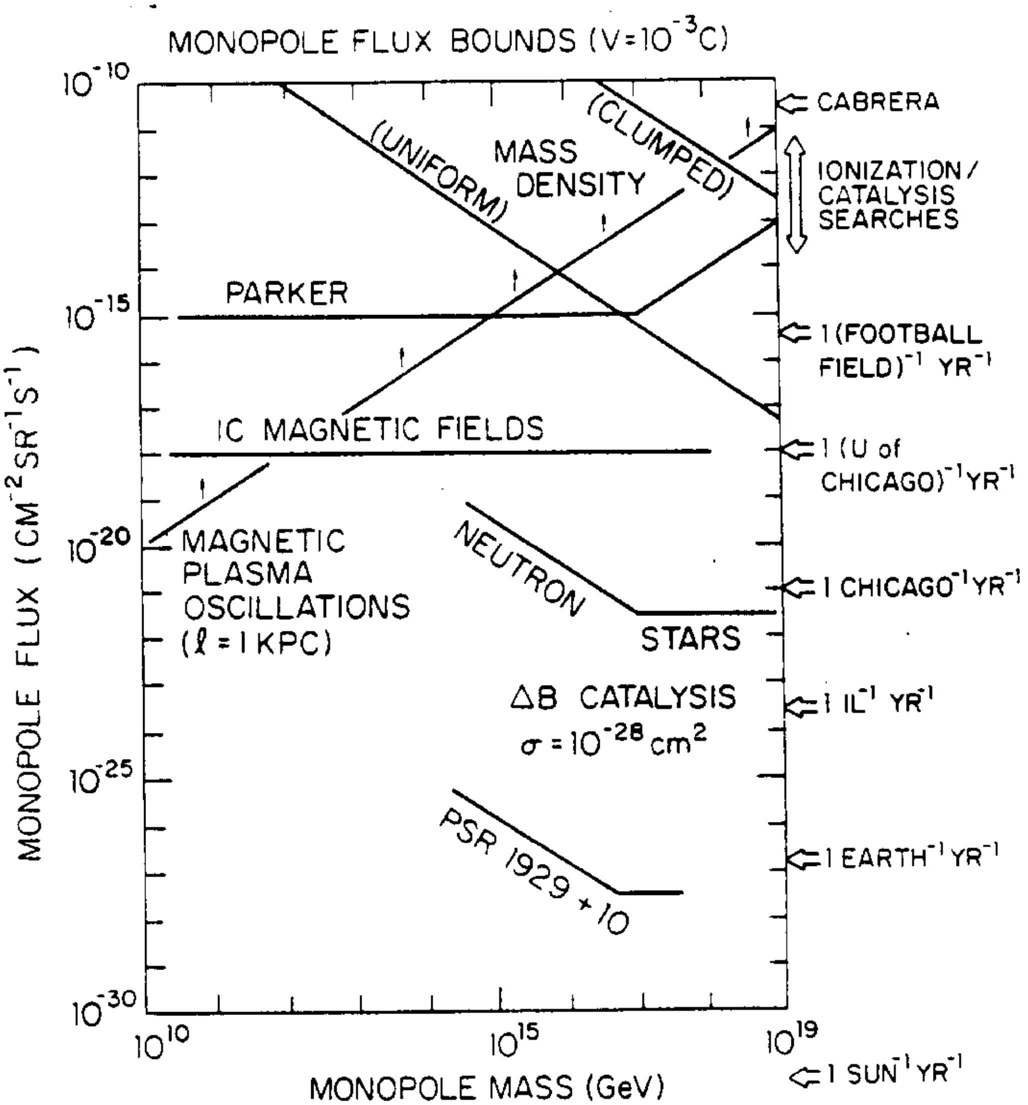

Fig. 3.1 Summary of the astrophysical/cosmological limits to the monopole flux as a function of monopole mass. Wherever necessary the monopole velocity is taken to be 10^{-3} c. The monopole catalysis bound based upon white dwarfs (ref. 93) is: $F_M \lesssim 2 \times 10^{-18} (\sigma v)^{-1}_{-28}$ cm^{-2} sr^{-1} s^{-1} (not shown here). The line labeled 'magnetic plasma oscillations' is the <u>lower</u> bound to the flux predicted in scenarios which evade the 'Parker bound' by having monopoles participate in the maintenance of the galactic B field.

$$v_{mag} = (2hB\ell/m_M)^{1/2}$$
$$= 3\times10^{-3}c(B/3\times10^{-6}G)^{1/2}(\ell/300pc)^{1/2}(10^{16}GeV/m_M)^{1/2}. \qquad (3.15)$$

If the monopole is not initially at rest the story is a bit different. There are two limiting situations, and they are characterized by the relative sizes of the initial velocity of the monopole, v_o, and the velocity just calculated above, v_{mag}. First, if the monopole is moving slowly compared to v_{mag}, $v_o << v_{mag} \simeq (2hB\ell/m)^{1/2}$, then it will undergo a large deflection due to the magnetic field and its change in KE will be given by 3.14. On the other hand, if $v_o >> v_{mag}$, then the monopole will only be slightly deflected by the magnetic field, and its change in KE will depend upon the direction of its motion relative to the magnetic field. In this situation the energy gained by a spatially isotropic distribution of monopoles, or a flux of equal numbers of north and south poles will vanish at first order in B-some poles will lose KE and some poles will gain KE. However, there is a net gain in KE at second order in B by the distribution of monopoles as a whole:

$$\langle\Delta KE\rangle \simeq (hB\ell)\ (v_o/v_{mag})^2/4 \qquad \text{(per monopole)}. \qquad (3.16)$$

For the galactic magnetic field $B \simeq 3\times10^{-6}$ G, $\ell \simeq 300$ pc, and $v_{mag} \simeq 3\times10^{-3}c\ (10^{16}\ GeV/m)^{1/2}$. Since $v_o \simeq 10^{-3}$ c, monopoles less massive than about 10^{17} GeV will undergo large deflections when moving through the galactic field and their gain in KE is given by Eqn. 3.14. Because of this energy gain, monopoles less massive than 10^{17} GeV will be ejected from galaxies in a very short time, and thus are unlikely to cluster in the haloes of galaxies. In fact the second order gain in KE will "evaporate" monopoles as massive as $O(10^{20}GeV)$ in a time less than the age of the galaxy[73]. Although consideration of galaxy formation would suggest that monopoles should cluster in galactic haloes, galactic magnetic fields should prevent monopoles less massive than $O(10^{20}$ GeV) from clumping in galactic haloes. [These conclusions are not valid if the magnetic field of the galaxy is in part produced by monopoles, a point to which I will return.]

The "no free-lunch principle" ($\Delta KE = -\Delta$ Magnetic Field Energy) and formulae 3.15 and 3.16 can be used to place a limit on the average flux of monopoles in the galaxy.[73,79-80] If, as it is commonly believed, the origin of the galactic magnetic field is due to dynamo action, then the time required to generate/regenerate the field is of the order of a galactic rotation time $\simeq O(10^8$ yr). Demanding that monopoles not drain the field energy in a time shorter than this results in the following constraints:

$m_M \lesssim 10^{17}$ GeV:

$$F \lesssim 10^{-15}cm^{-2}sr^{-1}s^{-1}(B/3\times10^{-6}\ G)\ (3\times10^7\ yr/\tau)\times$$
$$(r/30\ kpc)^{1/2}\ (300\ pc/\ell)^{1/2}, \qquad (3.17)$$

$m_M \gtrsim 10^{17}$ GeV:

$$F \lesssim 10^{-16}cm^{-2}sr^{-1}s^{-1}(m_M/10^{16}\ GeV)(3\times10^7yr/\tau)(300pc/\ell), \qquad (3.18)$$

where v_o has been assumed to be 10^{-3} c, τ is the regeneration time of the field, ℓ is the coherence length of the field, and r is the size of the magnetic field region in the galaxy. Constraint 3.17 which applies to 10^{16} GeV monopoles is very stringent (less than 3 monopoles soccer field^{-1} yr^{-1}) and is known as the "Parker bound." For more massive monopoles ($\gtrsim 10^{17}$ GeV) the "Parker bound" becomes less restrictive[70,73] (because the KE gain is a second order effect); however, the mass density constraint becomes more restrictive (cf. Fig. 3.1). These two bounds together restrict the flux to be $\lesssim 10^{-13}$ cm^{-2} sr^{-1} s^{-1} (which is allowed for monopoles of mass $\simeq 3 \times 10^{19}$ GeV).

Analogous arguments can be applied to other astrophysical magnetic fields. Rephaeli and Turner[81] have analyzed intracluster (IC) magnetic fields and derived a flux bound of $O(10^{-18}$ cm^{-2} sr^{-1} s$^{-1})$ for monopoles less massive than $O(10^{18}$ GeV). Although the presence of such fields has been inferred from diffuse radio observations for a number of clusters (including Coma), the existence of IC fields is not on the same firm footing as galactic fields. It is also interesting to note that the IC magnetic fields are sufficiently weak so that only monopoles lighter than $O(10^{16}$ GeV) should be ejected, and thus it is very likely that monopoles more massive than 10^{16} GeV will cluster in rich clusters of galaxies, where the local mass density is $O(10^2$-$10^3)$ higher than the mean density of the Universe. Unfortunately, our galaxy is not a member of a rich cluster.

Several groups have pointed out that the 'Parker bound' can be evaded if the monopoles themselves participate in the maintenance of the galactic magnetic field.[73,82-83] In such a scenario a monopole magnetic plasma mode is excited, and monopoles only 'borrow the KE' they gain from the magnetic field, returning it to the magnetic field a half cycle later. In order for this to work the monopole oscillations must maintain coherence; if they do not 'phase-mixing' (Landau damping) will cause the oscillations to rapidly damp. The criterion for coherence to be maintained is that the phase velocity of the oscillations $v_{ph} \simeq \omega_{pl}(\ell/2\pi)$ be greater than the gravitational velocity dispersion of the monopoles ($\simeq 10^{-3}$c); ℓ = wavelength of the relevant mode $\simeq$ coherence length of the galactic field $\lesssim$ 1 kpc. The monopole plasma frequency is given by

$$\omega_{pl} = (4\pi h^2 n_M/m_M)^{1/2}, \tag{3.19}$$

where n_M is the monopole number density. The condition that v_{ph} be $\gtrsim 10^{-3}$c implies a <u>lower</u> bound to flux of

$$\begin{aligned} F_M &\gtrsim 1/4\ m_M\ v_{grav}^3\ (h\ell)^{-2}, \\ &\gtrsim 10^{-14}(m_M/10^{16}\text{GeV})(1\text{kpc}/\ell)^2\text{cm}^{-2}\text{sr}^{-1}\text{s}^{-1}. \end{aligned} \tag{3.20}$$

Incidently, this also implies an upper bound to the oscillation period: $\tau = 2\pi/\omega_{pl} \lesssim \ell/v_{grav} \simeq 3 \times 10^6$ yr $(\ell/1\text{kpc})$--a very short time compared to other galactic timescales.

While it is <u>possible</u> that such scenarios could allow one to beat the 'Parker bound', a number of hurdles remain to be cleared before these scenarios can be called realistic or even viable. To mention a few, monopole oscillations can always be damped on sufficiently small scales (recall $v_{ph} = (\omega_{pl}/2\pi)\ell$), and nonlinear effects in this very complicated system--coupled electric and magnetic plasmas in a self-gravitating fluid, tend to feed power from large scales down to small scales. Can the coherence of the oscillations which is so crucial be maintained both spatially and temporally in the presence of inhomogeneities (after all the galaxy is not a homogeneous fluid)?

Finally, as the observational limits continue to improve, the large monopole flux predicted in these models will be the ultimate test. Already, the oscillation scenario for $m_M = 10^{16}$ GeV is probably observationally excluded.

Perhaps the most intriguing property of the monopole is its ability to catalyze nucleon decay with a strong interaction cross section: $(\sigma v) \simeq 10^{-28}$ cm^2. Since the symmetry of the GUT is restored at the monopole core, one would expect, on geometric grounds, that monopoles would catalyze nucleon decay with a cross section $\simeq M^{-2} \simeq 10^{-56}$ cm^2 ($M^{-1} \simeq$ size of monopole core)--which of course is utterly negligible. Rubakov[84] and independently Callan[85] showed that due to the singular nature of the potential between the s-wave of a fermion and a monopole, the fermion wave function is literally sucked into the core (technically, one might call this 's-wave sucking'), with the cross section saturating the unitarity bound: $(\sigma v) \simeq$ (fermion energy)$^{-2}$, or for low energies $(\sigma v) \simeq$ (fermion mass)$^{-2}$.

Needless to say, monopole catalysis has great astrophysical potential! For comparison, the nuclear reaction $4p \rightarrow {}^4He + 2e^+ + 2\nu_e$ which powers most stars proceeds at a weak interaction rate (first step: $p + p \rightarrow D$) and releases only about 0.7% of the rest mass involved, while monopole catalysis proceeds at a strong interaction rate and releases 100% of the rest mass of the nucleon (e.g., $M + n \rightarrow M + \pi^- + e^+$). The energy released by monopole catalysis is 3×10^3 erg s^{-1} $(\sigma v)_{-28}(\rho/1\,\mathrm{gcm}^{-3})$ per monopole; only about 10^{30} monopoles in the sun ($\simeq 10^{57}$ nucleons) are needed to produce the solar luminosity ($\simeq 4 \times 10^{33}$ erg s^{-1}). Here and throughout I will parameterize (σv) by:

$$(\sigma v) = (\sigma v)_{-28}\ c\ 10^{28}\ \mathrm{cm}^2.$$

Because of their awesome power to release energy via catalysis, there can't be too many monopoles in astrophysical objects like stars, planets, etc., otherwise the sky would be aglow in all wavebands from the energy released by monopoles. [This energy released in catalysis would be thermalized and radiated from the surface of the object.] The measured luminosities of neutron stars (some as low as 3×10^{30} erg s^{-1}); white dwarfs (some as low as 10^{29} erg s^{-1}); Jupiter (10^{25} erg s^{-1}); and the Earth (3×10^{20} erg s^{-1}) imply upper limits to the number of monopoles in these objects: some neutron stars ($\lesssim 10^{12}\ (\sigma v)_{-28}^{-1}$ monopoles); some white dwarfs ($\lesssim 10^{18}\ (\sigma v)_{-28}^{-1}$ monopoles); Jupiter ($\lesssim 10^{20}\ (\sigma v)_{-28}^{-1}$ monopoles); and the Earth ($\lesssim 3 \times 10^{15}\ (\sigma v)_{-28}^{-1}$ monopoles). In order to translate these limits into bounds on the monopole flux and abundance we need to know how many monopoles would be expected in each of these objects. As I discussed earlier, ab initio we would expect very few; those present must have been captured since the formation of the object. The number is $\propto F_M$ and is given by Eqn. 3.10; hence the limits above can be used to constrain the monopole flux.

The most stringent limit on F_M follows from considering neutron stars. A variety of techniques have been used to obtain limits to the luminosities of neutron stars [recall the limit to the number of monopoles is: $N_M \lesssim$ luminosity/$(10^{18}\text{erg s}^{-1}\ (\sigma v)_{-28}\ (\rho/3\times10^{14}\text{g cm}^{-3}))$]. I will just discuss one. The other techniques lead to similar bounds on F_M and are reviewed in ref.86.

PSR 1929 + 10 is an old ($\simeq 3\times10^6$ yr), radio pulsar whose distance from the earth is about 60 pc. The Einstein x-ray observatory was used to measure the luminosity of this pulsar, and it was determined to be $L \simeq 3 \times 10^{30}$ erg s^{-1} corresponding to a surface temperature of about 30 eV, making it the coolest neutron star yet observed. In its tenure as a neutron star it should have captured $10^{17}\ F_{-16}$ monopoles. The measured luminosity sets a limit to the number of monopoles in PSR 1929 + 10, $N_M \lesssim 10^{12}\ (\sigma v)^{-1}_{-28}$, which in turn can be used to bound $\langle F_M \rangle$:

$$\langle F_M \rangle \lesssim 10^{-21}(\sigma v)^{-1}_{-28}\text{cm}^{-2}\text{sr}^{-1}\text{s}^{-1} \qquad (3.21)$$

--which is less than one monopole Munich^{-1} yr^{-1}!

The progenitors of neutron stars are main sequence (MS) stars of mass $(1-30)M_\odot$ which were either too massive to become white dwarfs (WDs), or evolved to the WD state and were pushed over the Chandrasekhar limit by accretion from a companion star. Freese etal.[75] have calculated that MS stars in the mass range $(1-30)M_\odot$ will during their MS lifetime capture $(10^{23}-10^{25})F_{-16}$ monopoles (for $v_M = 10^{-3}c$ and $m_M \lesssim 10^{18}$ GeV, and depending on the star's mass). The progenitor of PSR 1929 + 10 should have captured at least 10^6 times more monopoles than the neutron star, and Freese etal.[75] argue that it is likely that a fair fraction of them should be retained in the neutron star. If we include these monopoles, the bound improves significantly, to

$$\langle F_M \rangle \lesssim 10^{-27}(\sigma v)^{-1}_{-28}\text{cm}^{-2}\text{sr}^{-1}\text{s}^{-1} \qquad (3.22)$$

--less than one monopole earth^{-1} yr^{-1}!

How reliable are these astrophysical bounds? The most stringent, Eqn. 3.22, relies upon an additional assumption, that the monopoles captured by the progenitor MS star make their way into the neutron star. Both bounds (and all catalysis bounds) are $\propto (\sigma v)^{-1}$. If the cross section for catalysis is not large, e.g., because the physics at the core of the monopole does not violate B conservation (such is the case for the Z_2 monopoles in SU(10))[87,88], or because the Callan-Rubakov calculation is incorrect, then the catalysis limits are not stringent.

In addition there are astrophysical uncertainties. Hot neutron stars radiate both γs and $\nu\bar{\nu}$s, but only the photons can be detected. The ratio of these luminosities has been calculated for various neutron star equations of state and was taken into account in deriving the catalysis

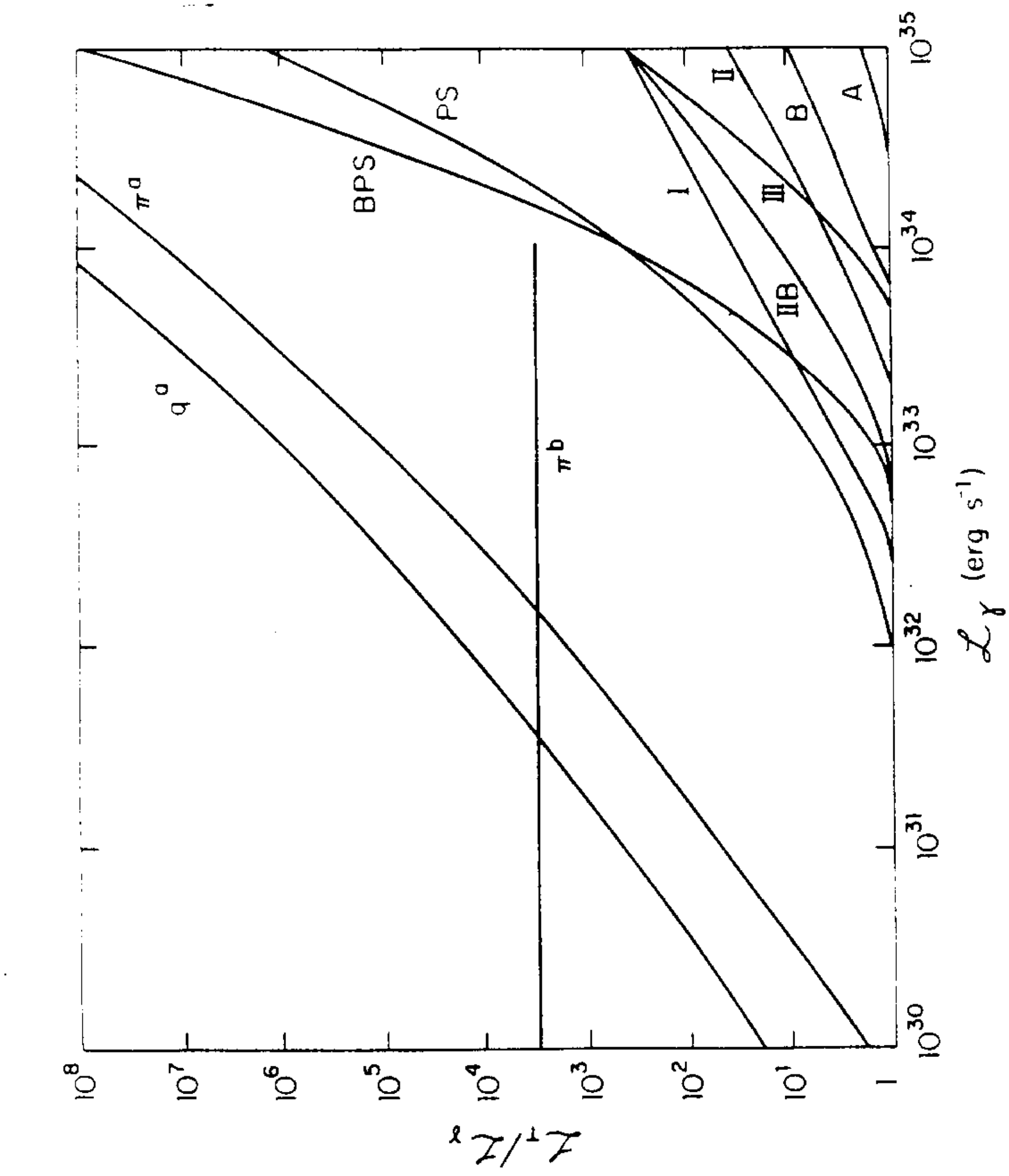

Fig. 3.2 The ratio of the total luminosity (= $L_\gamma + L_\nu$) of a hot neutron star to its photon luminosity as a function of L_γ. The different curves represent different neutron star equations of state: q (quark matter); π^a, π^b (pion condensate); the rest are more conventional equations of state (from ref. 86).

bounds. [For $L_\gamma \lesssim 10^{32}$ erg s^{-1}, L_ν is typically $\lesssim L_\gamma$; while for $L_\gamma \gtrsim 10^{32}$ erg s^{-1} L_ν can be (10^3-10^6) L_γ, see Fig. 3.2.] Monopoles less massive than about 10^{14} GeV may be deflected away from neutron stars with B fields $\gtrsim 10^{12}$ G; monopoles inside neutron stars which have pion condensates in their cores <u>may</u> be ejected by the so-called 'pion-slingshot effect'.[89]

The strength of the neutron star catalysis bounds lies in the number of different techniques which have been used. Individual objects have been studied[90] (PSR 1929 + 10 and 10 or so other old radio pulsars); searches for bright, nearby x-ray point sources have been made with negative results[91] [the number density of old ($\simeq 10^{10}$ yrs) neutron stars in our neighborhood should be $\gtrsim 10^{-4}$ pc^{-3}, implying that there should be O(100) or so within 100 pc of the solar system - if due to 'monopole heating' their luminosities were $\gtrsim 10^{31}$ erg s^{-1} they would surely have been detected]; the integrated contribution of old neutron stars to the diffuse soft x-ray background has been used to limit the average luminosity of an old neutron star ($\lesssim 10^{32}$ erg s^{-1}) and in turn the monopole flux.[86,91,92] The three techniques just mentioned involve different astrophysical assumptions and uncertainties, but all result in comparable bounds to $\langle F_M \rangle$: $\langle F_M \rangle \lesssim 10^{-21}\,(\sigma v)^{-1}_{-28}$ cm^{-2} sr^{-1} s^{-1}. Although I will not discuss it here, the same analysis has been applied to WDs,[93] and results in a less stringent bound, $\langle F_M \rangle \lesssim 2 \times 10^{-18}(\sigma v)^{-1}_{-28} cm^{-2} sr^{-1} s^{-1}$, but more importantly one which involves a different astrophysical system.

If monopoles catalyze nucleon decay with a large cross section, $(\sigma v)_{-28}$ not too much less than order unity, then, based upon the astrophysical arguments, it seems certain that the monopole flux must be small ($<< 10^{-18} cm^{-2} sr^{-1} sec^{-1}$). On the other hand, if the monopoles of interest do not catalyze nucleon decay at a significant rate (for whatever reason), then the 'Parker bound' is the relevant (and I believe reliable) constraint, with the outside possibility that it could be exceeded due to monopole plasma oscillations (--a scenario which is very astrophysically interesting!).

<u>Monopole Hunting</u>

There are two basic techniques for detecting a monopole: (1) inductive - a monopole which passes through a loop will induce a persistent current $\propto h/L$ (L = inductance of the loop $\propto$ radius, for a circular loop); (2) energy deposition - a monopole can deposit energy due to ionization $[dE/dx \simeq (10\ \mathrm{MeV/cm})(v/10^{-3}c)(\rho/1 gcm^{-3})]$, or indirectly by any nucleon decays it catalyzes. Method (1) has the advantage that the signal only depends upon the monopole's magnetic charge (and can be calculated by any first year graduate student who knows Maxwell's equations), and furthermore because of its unique signature (step function in the current) has the potential for clean identification. However, because the induced current $\propto L^{-1} \propto \mathrm{Area}^{-1/2}$, the simplest loop detectors are limited in size to $\lesssim 1m^2$ ($1m^2 \times 2\pi$ - sr x 1yr $\simeq 10^{12}$ cm^2 sr sec). In method (2) the detection signal depends upon other properties of the monopole (e.g. velocity, ability to catalyze nucleon decay), and the calculation of the energy loss is not so straightforward, as it involves the physics of the detector material.

However, it is very straightforward to fabricate very large detectors of this type.

On 14 February 1982 using a superconducting loop Blas Cabrera detected a jump in current of the correct amplitude for a Dirac magnetic charge.[94] His exposure at that time was about 2×10^9 cm^2 sr s--which naively corresponds to an enormous flux ($\approx 6 \times 10^{-10}$ cm^{-2} sr^{-1} s^{-1}), especially when compared to the astrophysical bounds discussed above. Sadly, since then his exposure has increased more than 100-fold with no additional candidates.[95] Ionization type searches with exposures upto 10^{14} cm sr s, sensitivities to monopole velocities 3×10^{-4} - 3×10^{-3} c, and no candidates have been reported. Searches which employ large proton decay detectors to search for multiple, colinear proton decays caused by a passing monopole with similar exposures (although these searches are only sensitive to specific windows in the (σv) - v_M space) have seen no candidate events. [There is a bit of a Catch 22 here; if σv is large enough so that a monopole would catalyze a string of proton decays in a proton decay detector ($(\sigma v)_{-28} \approx O(1)$), then the astrophysical bounds strongly suggest that $\langle F_M \rangle \lesssim 10^{-21}$ cm^{-2} sr^{-1} s^{-1}.] The most intriguing search done to date involves the etching of a 1/2 Byr old piece of mica of size a few cm^2 (exposure $\approx 10^{18}$ cm^2 sr s).[96] A monopole passing through mica leaves <u>no</u> etchable track; however, a monopole with a nucleus with $Z \gtrsim 10$ (e.g. Al) attached to it leaves an etchable track. Unfortunately, the negative results of searches of this type imply flux limits $\propto$ (probability of a monopole picking a nucleus and holding on to it)$^{-1}$. However exposures of up to 10^{22} cm^2 sr s can possibly be achieved, and if a track is seen, it would be a strong candidate for a monopole. [Very thorough and excellent reviews of monopole searches and searching techniques can be found in refs.97, 98.]

Concluding Remarks

What have we learned about GUT monopoles? (1) They are exceedingly interesting objects, which, if they exist, must be relics of the earliest moments of the Universe. (2) They are one of the very few predictions of GUTs that we can attempt to verify and study in our low energy environment. (3) Because of the glut of monopoles that should have been produced as topological defects in the very early Universe, the simplest GUTs and the standard cosmology (extrapolated back to times as early as $\approx 10^{-34}$ s) are not compatible. This is a very important piece of information about physics at very high energies and/or the earliest moments of the Universe. (4) There is no <u>believable</u> prediction for the flux of relic, superheavy magnetic monopoles. (5) Based upon astrophysical considerations, we can be reasonably certain that the flux of relic monopoles is small. Since it is not obligatory that monopoles catalyze nucleon decay at a prodigious rate, a <u>firm</u> upper limit to the flux is provided by the Parker bound[73], $\langle F_M \rangle \lesssim 10^{-15}$ cm^{-2} sr^{-1} s^{-1}. Note, this is not a predicted flux, it is only a firm <u>upper</u> bound to the flux. It is very likely that flux has to be even smaller, say $\lesssim 10^{-18}$ cm^{-2} sr^{-1} s^{-1} or even 10^{-21} cm^{-2} sr^{-1} s^{-1}. (6) There is every reason to believe that typical monopoles are moving with velocities (relative to us) of at least a few x 10^{-3} c. [Although it is <u>possible</u> that the largest contribution to the local monopole flux is due to a cloud of monopoles orbiting the sun with velocities $\approx (1 - 2) \times 10^{-4}$ c, I think that it is very <u>unlikely.</u>[99,100]]

LECTURE 4 - INFLATION

As I have discussed in Lecture 1 the hot big bang model seems to provide a reliable accounting of the Universe at least as far back as 10^{-2} sec after 'the bang' ($T \lesssim 10$ MeV). There are, however, a number of very fundamental 'cosmological facts' which the hot big bang model by itself does not elucidate (although it can easily accomodate them). The inflationary Universe paradigm, as originally proposed by Guth,[101] and modified by Linde,[102] and Albrecht and Steinhardt,[103] provides for the first time a framework for understanding the origin of these cosmological facts in terms of dynamics rather than just as particular initial data. As we shall see the underlying mechanism of their solution is rather generic--the temporary abolition of particle horizons and the production of entropy, and while inflation is the first realization of this mechanism which is based upon relatively well-known physics (spontaneous symmetry breaking (SSB) phase transitions), it may not prove to be the only such framework. I will begin by reviewing the cosmological puzzles, and then will go on to discuss the new inflationary Universe scenario.

Large-Scae Homogeneity and Isotropy

The observable Universe ($d \simeq H^{-1} \simeq 10^{28}$ cm $\simeq$ 3000 Mpc) is to a high degree of precision isotropic and homogenous on the largest scales (> 100 Mpc). The best evidence for this is provided by the uniformity of the cosmic background temperature: $\Delta T/T \lesssim 10^{-3}$ (10^{-4} if the dipole anisotropy is interpreted as being due to our peculiar motion through the cosmic rest frame; see Fig. 4.1). Large-scale density inhomogeneities or an anisotropic expansion would result in fluctuations in the microwave background temperature of a comparable size (see, e.g., refs. 104, 105). The smoothness of the observable Universe is puzzling if one wishes to understand it as a result of microphysical processes operating in the early Universe. As I mentioned in Lecture 1 the standard cosmology has particle horizons, and when matter and radiation last vigorously interacted (decoupling: $t \simeq 10^{13}$ s, $T \simeq 1/3$ eV) what was to become the presently observable Universe was comprised of $\simeq 10^{6}$ causally-distinct regions. Put slightly differently, the particle horizon at decoupling only subtends an angle of about 1/2° on the sky today; how is it that the microwave background temperature is so uniform on angular scales >> 1/2°?

Small-Scale Inhomogeneity

As any astronomer will gladly tell you on small scales ($\lesssim$ 100 Mpc) the Universe is very lumpy (stars, galaxies, clusters of galaxies, etc.). [Note, today $\delta\rho/\rho \simeq 10^{5}$ on the scale of a galaxy.] The uniformity of the microwave background on very small angular scales (<<1°) indicates that the Universe was smooth, even on these scales at the time of decoupling (see Fig. 4.1). [The relationship between angle subtended on the sky and mass contained within the corresponding length scale at decoupling is: $\theta \simeq 1' \, h(M/10^{12}M_{\odot})^{1/3}$.] Whence came the structure that is so conspicuous today? Once matter decouples from the radiation and is free of the pressure support provided by the radiation, small inhomogeneities will grow via the Jeans (gravitational) instability:

$\delta\rho/\rho \propto t^{2/3} \propto R$ (in the linear regime). [If the mass density of the Universe is dominated by a collisionless particle species, e.g., a light relic neutrino species, or axions, density perturbations in these particles can begin to grow when the Universe becomes matter-dominated, $R \simeq 3 \times 10^{-5} R_{today}$ for $\Omega h^2 = 1$.] Density perturbations of amplitude $\delta\rho/\rho \simeq 10^{-3}$ or so, on the scale of a galaxy ($\simeq 10^{12} M_\odot$) at the time of decoupling seem to be required to account for the small-scale structure observed today. Their origin, their spectrum (certainly perturbations should exist on scales other than $10^{12} M_\odot$), their nature (adiabatic or isothermal), and the composition of the dark matter (see ref. 3) are all crucial questions for understanding the formation of structure, which to date remain unanswered.

Flatness

The quantity $\Omega \equiv \rho/\rho_c$ measures the ratio of the energy density of the Universe to the critical energy density ($\rho_c \equiv 3H^2/8\pi G$). Although Ω is not known with great precision, from Lecture 1 we know that $0.01 \lesssim \Omega \lesssim$ few. Using Eqn. 1.5 Ω can be written as

$$\Omega = 1/(1 - x(t)), \tag{4.1a}$$

$$x(t) = (k/R^2)/(8\pi G\rho/3). \tag{4.1b}$$

Note that Ω is not constant, but varies with time since $x(t) \propto R(t)^n$ ($n = 1$ - matter-dominated, or 2 - radiation-dominated). Since $\Omega \simeq O(1)$ today, x_{today} must be at most $O(1)$. This implies that at the epoch of nucleosynthesis: $x_{BBN} \lesssim 10^{-16}$ and $\Omega_{BBN} = 1 \pm O(\lesssim 10^{-16})$, and that at the Planck epoch: $x_{pl} \lesssim 10^{-60}$ and $\Omega_{pl} = 1 \pm O(\lesssim 10^{-60})$. That is, very early on the ratio of the curvature term to the density term was extremely small, or equivalently, the expansion of the Universe proceeded at the critical rate ($H^2_{crit} = 8\pi G\rho/3$) to a very high degree of precision. Since $x(t)$ has apparently always been $\lesssim 1$, our Universe is today and has been in the past closely-described by the $k = 0$ flat model. Were the ratio x not exceedingly small early on, the Universe would have either recollapsed long ago ($k > 0$), or began its coasting phase ($k < 0$) where $R \propto t$. [If $k < 0$ and $x_{BBN} = 1$, then $T = 3K$ for $t = 300$ yrs!] The smallness of the ratio x required as an 'initial condition' for our Universe is puzzling. [The flatness puzzle has been emphasized in refs. 101,106.]

Predominance of Matter Over Antimatter

The puzzle involving the baryon number of the Universe, and its attractive explanation by B, C, CP violating interactions predicted by GUTs has been discussed at length in Lecture 2.

The Monopole Problem

The glut of monopoles predicted in the standard cosmology ('the monopole problem') and the lack of a compelling solution (other than inflation) has been discussed in Lecture 3.

The Smallness of the Cosmological Constant

With the possible exception of supersymmetry and supergravity theories, the absolute scale of the effective potential $V(\phi)$ is not determined in gauge theories (ϕ = one or more Higgs field). At low temperatures $V(\phi)$ is equivalent to a cosmological term (i.e., contributes $Vg_{\mu\nu}$ to the stress energy of the Universe). The observed expansion rate of the Universe today ($H \approx 50 - 100$ km s^{-1} Mpc^{-1}) limits the total energy density of the Universe to be $\lesssim O(10^{-29}$ g $cm^{-3}) \approx 10^{-46}$ GeV^4. Thus empirically the vacuum energy of our $T \approx 0$ SU(3) x U(1) vacuum (= $V(\phi)$ at the SSB minimum) must be $\lesssim 10^{-46}$ GeV^4. Compare this to the difference in energy density between the false ($\phi = 0$) and true vacua, which is $O(T_c^4)$ (T_c = symmetry restoration temperature): for $T_c \approx 10^{14}$ GeV, $V_{SSB}/V(\phi = 0) \lesssim 10^{-102}$! At present there is no satisfactory explanation for the vanishingly small value of the $T = 0$ vacuum energy density (equivalently, the cosmological term).

Today, the vacuum energy is apparently negligibly small and seems to play no significant role in the dynamics of the expansion of the Universe. If we accept this empirical determination of the absolute scale of $V(\phi)$, then it follows that the energy of the false ($\phi = 0$) vacuum is enormous ($\approx T_c^4$), and thus could have played a significant role in determining the dynamics of the expansion of the Universe. Accepting this very non-trivial assumption about the zero of the vacuum energy is the starting point for inflation (see Fig. 4.2).

Generic New Inflation

The basic idea of the inflationary Universe scenario is that there was an epoch when the vacuum energy density dominated the energy density of the Universe. During this epoch $\rho \approx V \approx$ constant, and thus $R(t)$ grows exponentially ($\propto \exp(Ht)$), allowing a small, causally-coherent region (initial size $\lesssim H^{-1}$) to grow to a size which encompasses the region which eventually becomes our presently-observable Universe. In Guth's original scenario[101], this epoch occurred while the Universe was trapped in the false ($\phi = 0$) vacuum during a strongly first-order phase transition. Unfortunately, in models which inflated enough (i.e., underwent sufficient exponential expansion) the Universe never made a 'graceful return' to the usual radiation-dominated FRW cosmology.[57,107] Rather than discussing the original model and its shortcomings in detail, I will instead focus on the variant, dubbed 'new inflation', proposed independently by Linde[102] and Albrecht and Steinhardt[103]. In this scenario, the vacuum-dominated epoch occurs while the region of the Universe in question is slowly, but inevitably, evolving toward the true, SSB vacuum. Rather than considering specific models in this section, I will discuss new inflation for a generic model.

Consider a SSB phase transition which occurs at an energy scale M_G. For $T \gtrsim T_c = M_G$ the symmetric ($\phi = 0$) vacuum is favored, i.e., $\phi = 0$ is the global minimum of the finite temperature effective potential $V_T(\phi)$ (= free energy density). As T approaches T_c a second minimum develops at $\phi \neq 0$, and at $T = T_c$ the two minima are degenerate. [I am assuming that this SSB transition is a first-order phase transition.] At temperatures below T_c the SSB ($\phi = \sigma$) minimum is the global minimum of $V_T(\phi)$ (see

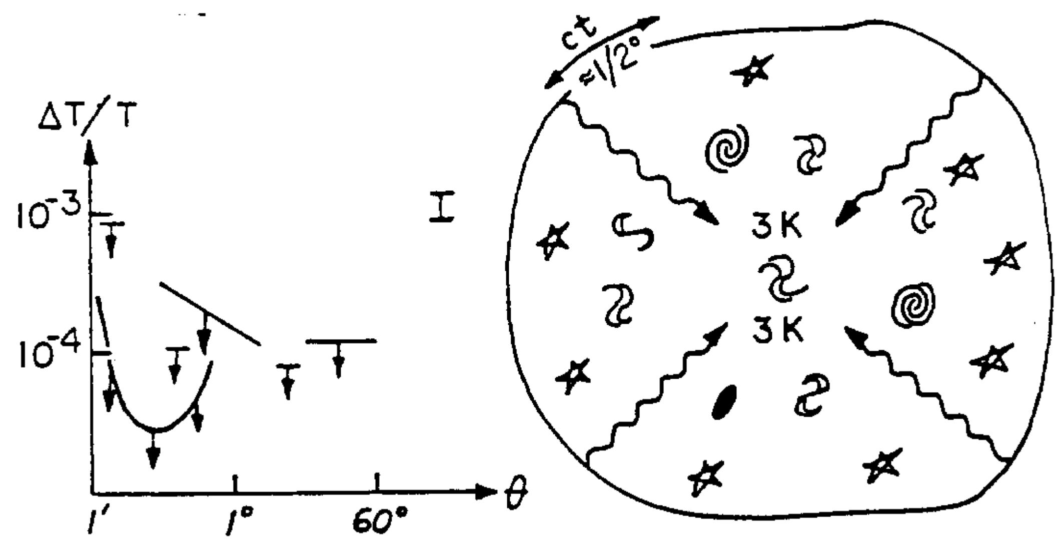

Fig. 4.1 Summary of measurements of the anisotropy of the 3K background on angular scales > 1' (from refs. 112, 113).

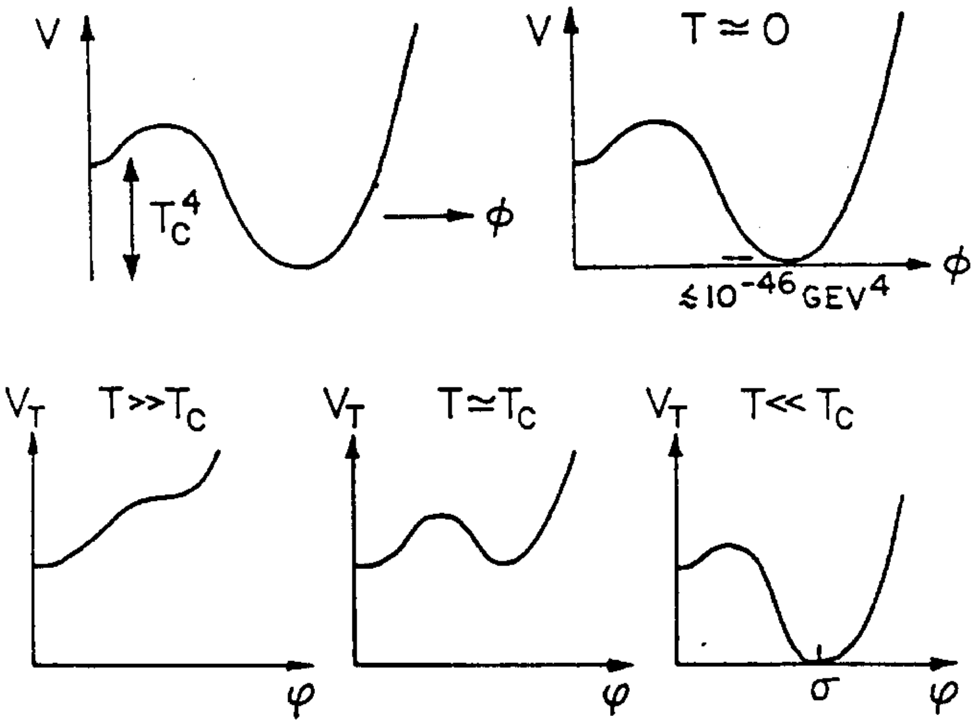

Fig. 4.2 The finite temperature effective potential V_T, for T > T_c; T = T_c; and T << T_c; here $\phi = \sigma$ is the SSB minimum of V.

Fig. 4.2). However, the Universe does not instantly make the transition from $\phi = 0$ to $\phi = \sigma$; the details and time required are a question of dynamics. [The scalar field ϕ is the order parameter for the SSB transition under discussion; in the spirit of generality ϕ might be a gauge singlet field or might have nontrivial transformation properties under the gauge group, possibly even responsible for the SSB of the GUT.]

Assuming a barrier exists between the false and true vacua, thermal fluctuations and/or quantum tunneling must be responsible for taking ϕ across the barrier. The dynamics of this process determine when and how the process occurs (bubble formation, spinodal decomposition, etc.) and the value of ϕ after the barrier is penetrated. For definiteness suppose that the barrier is overcome when the temperature is T_{MS} and the value of ϕ is ϕ_o. From this point the journey to the true vacuum is downhill (literally) and the evolution of ϕ should be adequately described by the semi-classical equations of motion for ϕ:

$$\ddot{\phi} + 3H\dot{\phi} + \Gamma\dot{\phi} + V' = 0, \tag{4.2}$$

where ϕ has been normalized so that its kinetic term in the Lagrangian is $1/2\ \partial_\mu\phi\partial^\mu\phi$, and prime indicates a derivative with respect to ϕ. The subscript T on V has been dropped; for $T \ll T_c$ the temperature dependence of V_T can be neglected and the zero temperature potential ($\equiv V$) can be used. The $3H\dot{\phi}$ term acts like a frictional force, and arises because the expansion of the Universe 'redshifts away' the kinetic energy of $\phi(\propto R^{-3})$. The $\Gamma\dot{\phi}$ term accounts for particle creation due to the time-variation of ϕ[refs. 108-110]. The quantity Γ is determined by the particles which couple to ϕ and the strength with which they couple (Γ^{-1} = lifetime of a ϕ particle). As usual, the expansion rate H is determined by the energy density of the Universe: ($H^2 = 8\pi G\rho/3$), with

$$\rho = 1/2\ \dot{\phi}^2 + V(\phi) + \rho_r, \tag{4.3}$$

where ρ_r represents the energy density in radiation produced by the time variation of ϕ. For $T_{MS} \ll T_c$ the original thermal component makes a negligible contribution to ρ. The evolution of ρ_r is given by

$$\dot{\rho}_r + 4H\rho_r = \Gamma\dot{\phi}^2, \tag{4.4}$$

where the $\Gamma\dot{\phi}^2$ term accounts for particle creation by ϕ.

In writing Eqns. 4.2-4.4 I have implicitly assumed that ϕ is spatially homogeneous. In some small region (inside a bubble or a fluctuation region) this will be a good approximation. The size of this smooth region will be unimportant; take it to be of order the 'physics horizon', H^{-1}. Now follow the evolution of ϕ within the small, smooth patch of size H^{-1}.

If V is sufficiently flat somewhere between $\phi = \phi_o$ and $\phi = \sigma$, then ϕ will evolve very slowly in that region, and the motion of ϕ will be 'friction-dominated' so that $3H\dot{\phi} = -V'$ (in the slow growth phase particle creation is not important[110]). If V is sufficiently flat, then the time required for ϕ to transverse the flat region can be long compared to the expansion timescale H^{-1}, say for definiteness, $\tau_\phi = 100\ H^{-1}$. During this slow growth phase $\rho \simeq V(\phi) \simeq V(\phi = 0)$; both ρ_r and $1/2\ \dot{\phi}^2$ are << $V(\phi)$. The expansion rate H is then just

$$H = (8\pi V(0)/3m_{pl}^2)^{1/2} \qquad (4.5)$$
$$\simeq M_G^2/m_{pl},$$

where V(0) is assumed to be of order M_G^4. While H = constant R grows exponentially: $R \propto \exp(Ht)$; for $\tau_\phi \approx 100\ H^{-1}$ R expands by a factor of e^{100} during the slow rolling period, and the physical size of the smooth region increases to $e^{100}H^{-1}$. This exponential growth phase is called a deSitter phase.

As the potential steepens, the evolution of ϕ quickens. Near $\phi = \sigma$, ϕ oscillates around the SSB minimum with frequency ω: $\omega^2 \simeq V''(\sigma) = M_G^2 >> H^2 = M_G^4/m_{pl}^2$. As ϕ oscillates about $\phi = \sigma$ its motion is damped by particle creation and the expansion of the Universe. If $\Gamma^{-1} << H^{-1}$, the coherent field energy density ($V + 1/2\ \dot{\phi}^2$) is converted into radiation in less than an expansion time ($\Delta t_{RH} \simeq \Gamma^{-1}$), and the patch is reheated to a temperature $T \simeq O(M_G)$ - the vacuum energy is efficiently converted into radiation ('good reheating'). On the other hand, if $\Gamma^{-1} >> H^{-1}$, then ϕ continues to oscillate and the coherent field energy redshifts away with the expansion: $(V + 1/2\ \dot{\phi}^2) \propto R^{-3}$. [The coherent field energy behaves like nonrelativistic matter; see ref. 111 for more details.] Eventually, when $t \simeq \Gamma^{-1}$ the energy in radiation begins to dominate that in coherent field oscillations, and the patch is reheated to a temperature $T \simeq (\Gamma/H)^{1/2}M_G \simeq (\Gamma m_{pl})^{1/2} << M_G$ ('poor reheating'). The evolution of ϕ is summarized in Fig. 4.3.

For the following discussion let us assume 'good reheating' ($\Gamma >> H$). After reheating the patch has a physical size $e^{100}H^{-1}$ ($\simeq 10^{17}$cm for $M_G \simeq 10^{14}$ GeV), is at a temperature of order M_G, and in the approximation that ϕ was initially constant throughout the patch, the patch is exactly smooth. From this point forward the region evolves like a radiation-dominated FRW model. How have the cosmological conundrums been 'explained'? First, <u>the homogeneity and isotropy</u>; our observable Universe today ($\simeq 10^{28}$cm) had a physical size of about 10 cm ($\approx 10^{28}$cm x $3K/10^{14}$ GeV) when T was 10^{14} GeV. Thus it lies well within one of the smooth regions produced by the inflationary epoch. At this point <u>the inhomogeneity puzzle</u> has not been solved, since the patch is precisely uniform. Due to deSitter space produced quantum fluctuations in ϕ, ϕ is not exactly uniform even in a small patch. Later, I will discuss the density inhomogeneities that result from the quantum fluctuations in ϕ. <u>The flatness puzzle</u> involves the smallness of the ratio of the curvature term to the energy density term. This ratio is exponentially smaller after inflation: $x_{after} \simeq e^{-200}\ x_{before}$ since the energy density before and after inflation is $O(M_G^4)$, while k/R^2 has decreased exponentially (by e^{200}). Since the ratio x is reset to an exponentially small value, the inflationary scenario predicts that today Ω should be $1 \pm O(10^{-BIG})$. If the Universe is reheated to a temperature of order M_G, a baryon asymmetry can evolve in the usual way, although the quantitative details may be slightly different[49,110]. If the Universe is not efficiently reheated ($T_{RH} << M_G$), it may be possible for n_B/s to be produced directly in the decay of the coherent field oscillations (which behave just like NR ϕ particles). This is an example of very out-of-equilibrium decay (discussed in Lecture 2), in which case the n_B/s produced is $\propto T_{RH}/(m_\phi = \omega)$ and does not depend upon T_{RH} being of order 10^{14} GeV or so. In any case, it is absolutely necessary to have baryogenesis occur after reheating since any baryon number (or any other quantum number) present before inflation is diluted by a factor $(M_G/T_{MS})^3 \exp(3H\tau_\phi)$ - the factor by which the total entropy increases. Note that if C, CP are violated spontaneously, then ε (and n_B/s) could have a different sign in different patches--leading to a Universe which on the very largest scales ($>> e^{100}H^{-1}$) is baryon symmetric.

Since the patch that our observable Universe lies within was once (at the beginning of inflation) causally-coherent, the Higgs field could have been aligned throughout the patch (indeed, this is the lowest energy configuration), and thus there is likely to be $\lesssim 1$ monopole within the entire patch which was produced as a topological defect. The glut of monopoles which occurs in the standard cosmology does not occur. [The production of other topological defects (such as domain walls, etc.) is avoided for similar reasons.] As discussed in Lecture 3, some monopoles will be produced after reheating in rare, very energetic particle collisions. The number produced is exponentially small and exponentially uncertain. [In discussing the resolution of the monopole problem I am tacitly assuming that the SSB of the GUT is occurring during the SSB transition in question, or that it has already occurred in an earlier SSB transition; if not then one has to worry about the monopoles produced in the subsequent GUT transition.]

The key point is that although monopole production is intrinsically small in inflationary models, the uncertainties in the number of monopoles produced are exponential. Of course, it is also possible that monopoles might be produced as topological defects in a subsequent phase transition[114], although it may be difficult to arrange that they not be overproduced.

Finally, the inflationary scenario sheds no light upon the cosmological constant puzzle. Although it can potentially successfully resolve all of the other puzzles in my list, inflation is, in some sense, a house of cards built upon the cosmological constant puzzle.

Density Inhomogeneities

Before I discuss the production of density inhomogeneities during the inflationary transition I will briefly review some of the 'Standard Lore'. [A more thorough and systematic treatment of the subject can be found in ref.105.]

A density perturbation is described by its wavelength λ or its wavenumber $k(= 2\pi/\lambda)$, and its amplitude $\delta\rho/\rho$ (ρ = average energy density). As the Universe expands the physical (or proper) wavelength of a given perturbation also expands; it is useful to scale out the

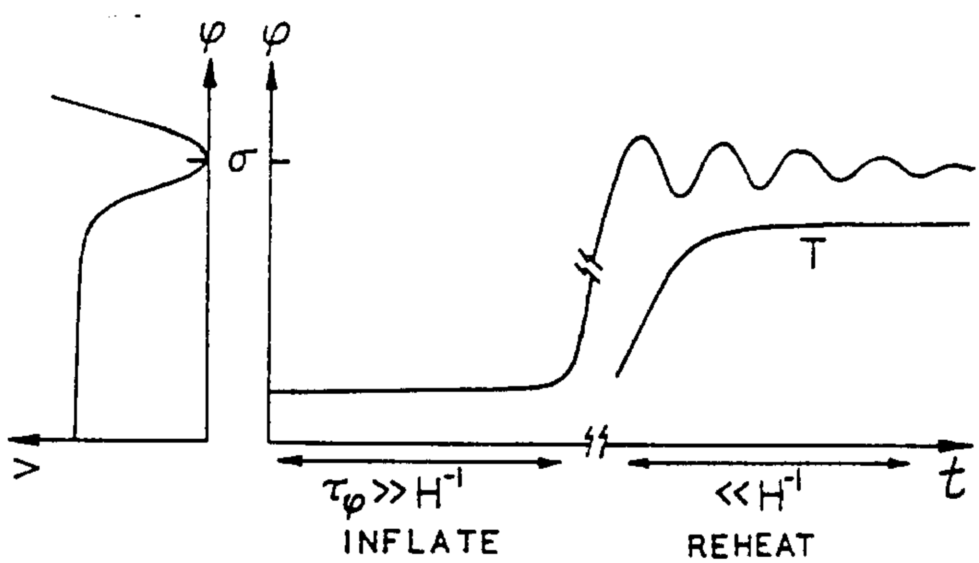

Fig. 4.3 The time evolution of ϕ. During the slow growth phase the time required for ϕ to change appreciably is $>> H^{-1}$. As the potential steepens ϕ evolves more rapidly (timescale $<< H^{-1}$), eventually oscillating about the SSB minimum. Particle creation damps the oscillations in a time $\simeq \Gamma^{-1}$ ($<<H^{-1}$, if $\Gamma>>H$, as shown here) reheating the patch to $T \simeq \min[M_G, (\Gamma m_{pl})^{1/2}]$.

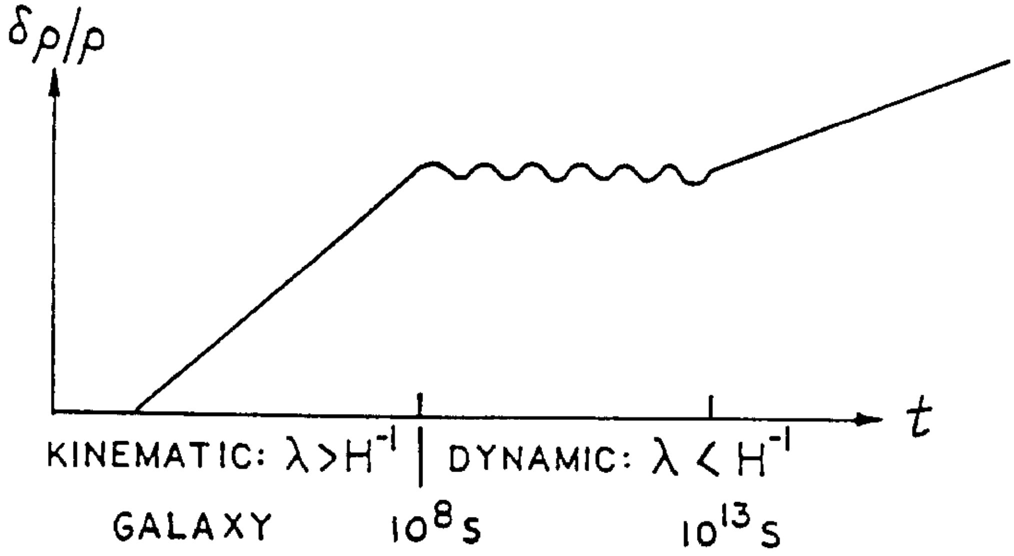

Fig. 4.4 Evolution of a galactic mass adiabatic density perturbation.

expansion so that a particular perturbation is always labeled by the same comoving wavelength $\lambda_c \equiv \lambda/R(t)$ or comoving wavenumber $k_c \equiv kR(t)$. [$R(t)$ is often normalized so that $R_{today} = 1$.] Even more common is to label a perturbation by the comoving baryon mass (or total mass in nonrelativistic particles if $\Omega_b \neq \Omega_{TOT}$) within a half wavelength $M = \pi\lambda^3 n_B m_N/6$ (n_B = net baryon number density, m_N = nucleon mass).

The relative sizes of λ and H^{-1} (= 'physics horizon' and particle horizon also in the standard cosmology) are crucial for determining the evolution of $\delta\rho/\rho$. When $\lambda \lesssim H^{-1}$ (the perturbation is said to be inside the horizon) microphysics can affect the perturbation. If $\lambda > \lambda_J = v_s H^{-1}$ (physically λ_J, the Jeans length, is the distance a pressure wave can propagate in an expansion time; v_s = sound speed) and the Universe is matter-dominated, then $\delta\rho/\rho$ grows $\propto t^{2/3} \propto R$. Perturbations with $\lambda < \lambda_J$ oscillate as pressure-supported sound waves (and may even damp).

When a perturbation is outside the horizon ($\lambda > H^{-1}$) the situation is a bit more complicated. The quantity $\delta\rho/\rho$ is not gauge-invariant; when $\lambda < H^{-1}$ this fact creates no great difficulties. However when $\lambda > H^{-1}$ the gauge-noninvariance is a bit of a nightmare. Although Bardeen[115] has developed an elegant gauge-invariant formalism to handle density perturbations in a gauge-invariant way, his gauge invariant quantities are not intuitively easy to understand. I will try to give a brief, intuitive description in terms of the gauge dependent, but more intuitive quantity $\delta\rho/\rho$. Physically, only real, honest-to-God wrinkles in the geometry (called curvature fluctuations or adiabatic fluctuations) can 'grow'. In the synchronous gauge ($g_{00} = -1$, $g_{0i} = 0$) $\delta\rho/\rho$ for these perturbations grows $\propto t^n$ (n = 1 - radiation dominated, = 2/3 - matter dominated). Geometrically, when $\lambda > H^{-1}$ these perturbations are just wrinkles in the space time which are evolving kinematically (since microphysical processes cannot affect their evolution). Adiabatic perturbations are characterized by $\delta\rho/\rho \neq 0$ and $\delta(n_B/s) = 0$; while isothermal perturbations (which do not grow outside the horizon) are characterized by $\delta\rho/\rho = 0$ and $\delta(n_B/s) \neq 0$. [With greater generality $\delta(n_B/s)$ can be replaced by any spatial perturbation in the equation of state $\delta p/p$, where $p = p(\rho, \ldots)$.] In the standard cosmology $H^{-1} \propto t$ grows monotonically; a perturbation only crosses the horizon once (see Fig. 4.5). Thus it should be clear that microphysical processes cannot create adiabatic perturbations (on scales $\gtrsim H^{-1}$) since microphysics only operates on scales $\lesssim H^{-1}$. In the standard cosmology adiabatic (or curvature) perturbations were either there ab initio or they are not present. Microphysical processes can create isothermal (or pressure perturbations) on scales $\gtrsim H^{-1}$ (of course, they cannot grow until $\lambda \lesssim H^{-1}$). Fig. 4.4 shows the evolution of a galactic mass ($\simeq 10^{12} M_\odot$) adiabatic perturbation: for $t \lesssim 10^8$ s, $\lambda > H^{-1}$ and $\delta\rho/\rho \propto t$; for 10^{13} s $\gtrsim t \gtrsim 10^8$ s, $\lambda < H^{-1}$ and $\delta\rho/\rho$ oscillates as a sound wave since matter and radiation are still coupled ($v_s \simeq c$) and hence $\lambda_J \simeq H^{-1}$; for $t \gtrsim 10^{13}$ s, $\lambda < H^{-1}$ and $\delta\rho/\rho \propto t^{2/3}$ since matter and radiation are decoupled ($v_s << c$) and $\lambda_J < \lambda_{Galaxy}$. [Note: in an $\Omega = 1$ Universe the mass inside the horizon $\simeq (t/\mathrm{sec})^{3/2} M_\odot$.]

Finally, at this point it should be clear that a convenient epoch to specify the amplitude of a density perturbation is when it crosses the horizon. It is often supposed (in the absence of knowledge about the origin of perturbations) that the spectrum of fluctuations is a power law (i.e., no preferred scale):

$$(\delta\rho/\rho)_H = \varepsilon M^{-\alpha}.$$

If $\alpha > 0$, then on some small scale perturbations will enter the horizon with amplitude $\gtrsim O(1)$--this leads to black hole formation; if this scale is $\gtrsim 10^{15}$ g (mass of a black hole evaporating today) there will be too many black holes in the Universe today. On the other hand, if $\alpha < 0$ then the Universe becomes more irregular on larger scales (contrary to observation). In the absence of a high or low mass cutoff, the $\alpha = 0$ (so-called Zel'dovich spectrum[116]) of density perturbations seems to be the only 'safe' spectrum. It has the attractive feature that all scales cross the horizon with the same amplitude (i.e., it is scale-free). Such a spectrum is not required by the observations; however, such a spectrum with amplitude of $O(10^{-4})$ probably leads to an acceptable picture of galaxy formation (i.e., consistent with all present observations--microwave background fluctuations, galaxy correlation function, etc.; for a more detailed discussion see ref. 3.)

Origin of Density Inhomogeneities in the New Inflationary Universe

The basic result is that quantum fluctuations in the scalar field ϕ (due to the deSitter space event horizon which exists during the exponential expansion (inflation) phase) give rise to an almost scale-free (Zel'dovich) spectrum of density perturbations of amplitude

$$(\delta\rho/\rho)_H = (4 \text{ or } 2/5) H \, \Delta\phi/\dot{\phi}(t_1), \tag{4.6}$$

where 4 applies if the scale in question reenters the horizon when the Universe is radiation-dominated and $(\delta\rho/\rho)_H$ is then the amplitude of the sound wave; 2/5 applies if the scale in question reenters the horizon when the Universe is matter-dominated and $(\delta\rho/\rho)_H$ is then the amplitude of the growing mode perturbation at horizon crossing; H is the value of the Hubble parameter during inflation; $\dot{\phi}(t_1)$ is the value of $\dot{\phi}$ when the perturbation left the horizon during the deSitter phase; and $\Delta\phi \simeq H/2\pi$ is the fluctuation in ϕ. This result was derived independently by the authors of refs. 117-120. Rather than discussing the derivation in detail here, I will attempt to physically motivate the result. This result turns out to be the most stringent constraint on models of new inflation.

The crucial difference between the standard cosmology and the inflationary scenario for the evolution of density perturbations is that H^{-1} (the 'physics horizon') is not strictly monotonic; during the inflationary (deSitter) epoch it is constant. Thus, a perturbation can cross the horizon ($\lambda = H^{-1}$) twice (see Fig. 4.5)! The evolution of two scales (λ_G = galaxy and λ_H = presently observable Universe) is shown in Fig. 4.5. Earlier than t_1 (time when $\lambda_G = H^{-1}$) $\lambda_G < H^{-1}$ and microphysics (quantum fluctuations, etc.) can operate on this scale. When $t = t_1$ microphysics 'freezes out' on this scale; the density perturbation which

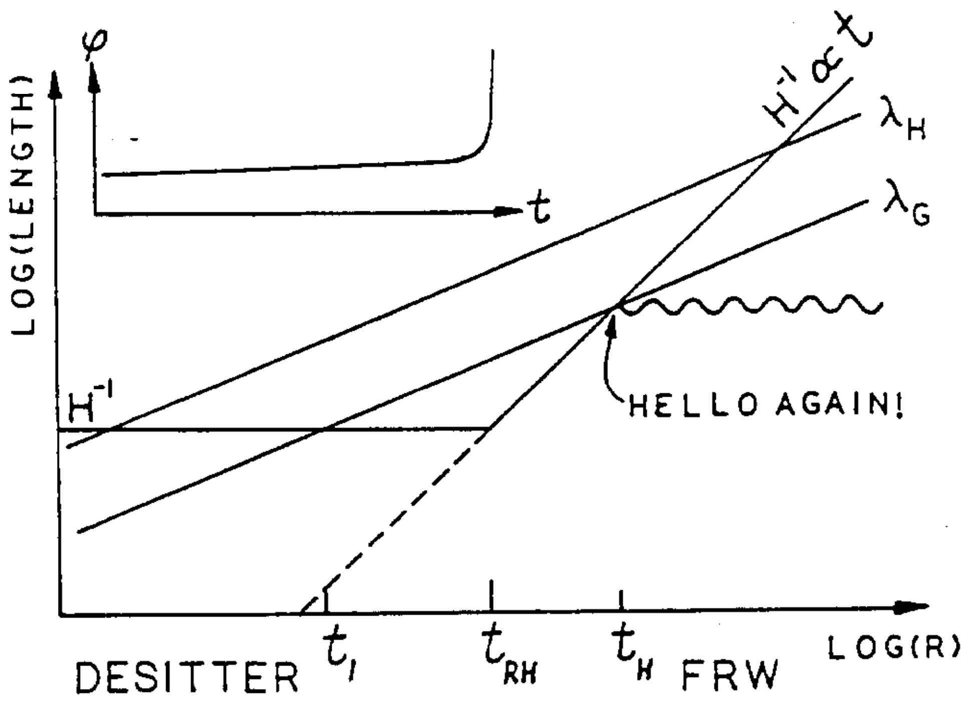

Fig. 4.5 The evolution of the 'physics horizon' ($\simeq H^{-1}$) and the physical sizes of perturbations on the scale of a galaxy (λ_G) and on the scale of the present observable Universe (λ_H). Reheating occurs at $t = t_{RH}$. For reference the evolution of ϕ is also shown. The broken line shows the evolution of H^{-1} in the standard cosmology. In the inflationary cosmology a perturbation crosses the horizon twice, which makes it possible for causal microphysics (in this case, quantum fluctuaions in ϕ) to produce large-scale density perturbations.

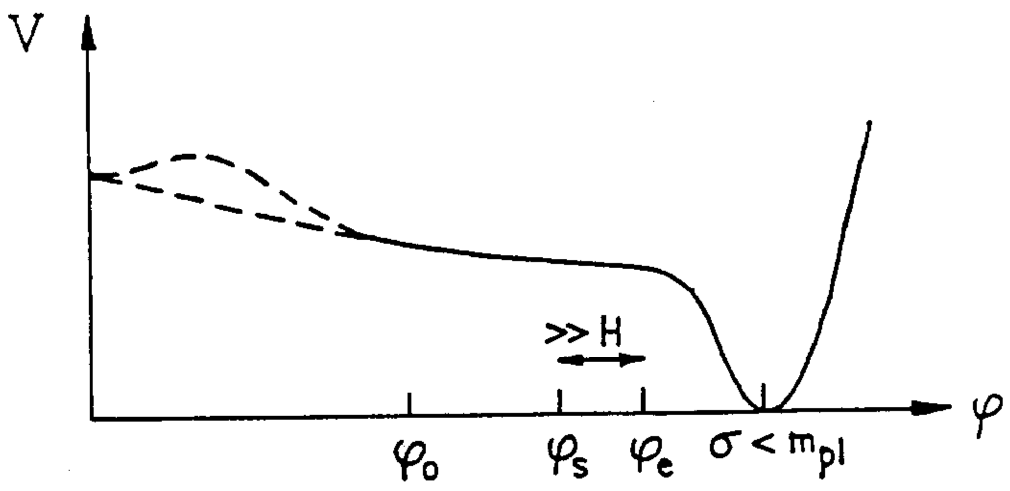

Fig. 4.6 The 'prescribed potential' for successful inflation.

exists on this scale, say $(\delta\rho/\rho)_1$, then evolves 'kinematically' until it reenters the horizon at $t = t_H$ (during the subsequent radiation-dominated FRW phase) with amplitude $(\delta\rho/\rho)_H$.

DeSitter space is exactly time-translationally-invariant; the inflationary epoch is approximately a deSitter phase - ϕ is almost, but not quite constant (see Fig. 4.3). [In deSitter space $p + \rho = 0$; during inflation $p + \rho = \dot{\phi}^2$.] This time-translation invariance is crucial; as each scale leaves the horizon (at $t = t_1$) $\delta\rho/\rho$ on that scale is fixed by microphysics to be some value, say, $(\delta\rho/\rho)_1$. Because of the (approximate) time-translation invariance of the inflationary phase this value $(\delta\rho/\rho)_1$ is (approxmately) the same for all scales. [Recall H, ϕ, $\dot{\phi}$ are all approximately constant during this epoch, and each scale has the same physical size ($= H^{-1}$) when it crosses outside of the horizon.] The precise value of $(\delta\rho/\rho)_1$ is fixed by the amplitude of the quantum fluctuations in ϕ on the scale H^{-1}; for a free scalar field $\Delta\phi = H/2\pi$ (the Hawking temperature). [Recall, during inflation V'' ($\simeq$ the effective mass-squared) is very small.]

While outside the horizon ($t_1 \lesssim t \lesssim t_H$) a perturbation evolves 'kinematically' (as a wrinkle in the geometry); viewed in some gauges the amplitude changes (e.g., the synchronous gauge), while in others (e.g., the uniform Hubble constant gauge) it remains constant. However, in all gauges the kinematic evolution is independent of scale (intuitively this makes sense since this is the kinematic regime). Given these 'two facts': $(\delta\rho/\rho)_1 \simeq$ scale-independent and the kinematic evolution $\simeq$ scale-independent, it follows that all scales reenter the horizon (at $t = t_H$) with (approximately) the same amplitude, given by Eqn. 4.6. Not only is this a reasonable spectrum (the Zel'dovich spectrum), but this is one of the very few instances that the spectrum of density perturbations has been calculable from first principles. [The fluctuations produced by strings are another such example, see, e.g. ref. 121; however, in a string scenario without inflation the homogeneity of the Universe must be assumed.]

Coleman-Weinberg SU(5) Model

The first model of new inflation[102,103] studied was the Coleman-Weinberg SU(5) model, with $T = 0$ effective potential

$$V(\phi) = 1/2\, B\sigma^4 + B\phi^4[\ln(\phi^2/\sigma^2) - 1/2], \qquad (4.7)$$

$$\simeq 1/2\, B\sigma^4 - \lambda(\phi)\phi^4 \qquad (\phi \ll \sigma)$$

where Φ is the 24 dimensional field responsible for GUT SSB, ϕ is the magnitude of Φ in the SU(3) x SU(2) x U(1) SSB direction, $B = 25g^4/256\pi^2$ (g = gauge coupling constant), $\sigma \simeq 1.2 \times 10^{15}$ GeV, and for $\phi \simeq 10^9$ GeV, $\lambda(\phi) \simeq 0.1$. [V may not look familiar; this is because ϕ is normalized so that its kinetic term is $1/2\ \dot{\phi}^2$ rather than the usual $(15/4)\dot{\phi}^2$.] Albrecht and Steinhardt[15] showed that when $T \simeq 10^8 - 10^9$ GeV the metastability limit is reached, and thermal fluctuations drive ϕ over the T-dependent barrier (height $\simeq T^4$) in the finite temperature effective potential. Naively, one expects that $\phi_0 \simeq T_{MS}$ since for $\phi \ll \sigma$ there is no other scale in the potential (this is a point to which I will return). The potential is sufficiently flat that the approximation $3H\dot{\phi} = -V'$ is valid for $\phi << \sigma$, and it follows that

$$(\phi/H)^2 = (3/2\lambda)[H(\tau_\phi - t)]^{-1}, \quad (4.8)$$

where $H\tau_\phi = (3/2\lambda)(H/\phi_0)^2$ (recall τ_ϕ = time it takes ϕ to traverse the flat portion of the potential). Physically, $H\tau_\phi$ is the number of e-folds of R which occur during inflation, which to solve the homogeneity-isotropy and flatness puzzles must be $\gtrsim 0(60)$. For this model $H \simeq 7 \times 10^9$ GeV; setting $\phi_0 \simeq 10^8 - 10^9$ GeV results in $H\tau_\phi \simeq 0(500-50000)$ - seemingly more than sufficient inflation.

There is however, a very basic problem here. Eqn. 4.8 is derived from the semi-classical equation of motion for ϕ [Eqn. 4.1], and thus only makes sense when the evolution of ϕ is 'classical', that is when $\phi >> \Delta\phi_{QM}$ (= quantum fluctuations in ϕ). In deSitter space the scale of quantum fluctuations is set by H: $\Delta\phi_{QM} \simeq H/2\pi$ (on the length scale H^{-1}). Roughly speaking then, Eqn. 4.8 is only valid for $\phi >> H$. However, sufficient inflation requires $\phi_0 \lesssim H$. Thus the Coleman-Weinberg model seems doomed for the simple reason that all the important physics must occur when $\phi \lesssim \Delta\phi_{QM}$. This is basically the conclusion reached by Linde[122] and Vilenkin and Ford[123] who have analyzed these effects carefully. Note that by artificially reducing λ by a factor of 10-100 sufficient inflation can be achieved $\phi_0 >> H$ (i.e., the potential becomes sufficiently flat that the classical part of the evolution, $\phi >> H$, takes a time $\gtrsim 60\ H^{-1}$). In the Coleman-Weinberg model $\Gamma >> H$ and the Universe reheats to $T \simeq M_G \simeq 10^{14}$ GeV.

Let's ignore for the moment the difficulties associated with the need to have $\phi_0 < H$, and examine the question of density fluctuations. Combining Eqns. 4.6 and 4.8 it follows that

$$(\delta\rho/\rho)_H \simeq (4 \text{ or } 2/5)100\lambda^{1/2}[1 + \ln(M/10^{12}M_\odot)/171 + \ln(g\sigma/10^{15}\text{ GeV})/57]^{3/2}, \quad (4.9)$$

where M is the comoving mass within the perturbation. Note that the spectrum is almost, but not quite scale-invariant (varying by less than a factor of 2 from $1M_\odot$ to $10^{22}M_\odot$ = present horizon mass). Blindly plugging in $\lambda \simeq 0.1$, results in $(\delta\rho/\rho)_H \simeq 0(10^2)$ which is clearly a disaster. [On angular scales >> 1° the Zel'dovich spectrum results in temperature fluctuations of[104] $\Delta T/T \simeq 1/2(\delta\rho/\rho)_H$ which must be $\lesssim 10^{-4}$ to be consistent with the observed isotropy.] To obtain perturbations of an acceptable amplitude one must artificially set $\lambda \simeq 10^{-12}$ or so. [In an SU(5) GUT λ is determined by the value of $\alpha_{GUT} = g^2/4\pi \simeq 1/45$, which implies $\lambda \simeq 0.1$.] As mentioned earlier the density fluctuation constraint is a very severe one; recall that $\lambda \simeq 10^{-2} - 10^{-3}$ would solve the difficulties associated with the quantum fluctuations in ϕ. To say the least, the Coleman-Weinberg SU(5) model seems untenable.

Lessons Learned--A Prescription for Successful New Inflation

Other models for new inflation have been studied, including supersymmetric models which employ the inverse hierarchy scheme,[124] supersymmetric/supergravity models[124-126] and just plain GUT models[127] No model has led to a completely satisfactory new inflationary scenario, some failing to reheat sufficiently to produce a baryon asymmetry, others plagued by large density perturbations, etc. Unlike the situation with 'old inflation' a few years ago, the situation does not appear hopeless. The early failures have led to a very precise prescription for a potential which will successfully implement new inflation.[128] Among the necessary conditions are:

(1) A flat region where the motion of ϕ is 'friction-dominated', i.e., term negligible so that $3\dot{H} = -V'$. This i.e., $\ddot{\phi}$ term negligible so that $3H\dot{\phi} = -V'$. This requires an interval where $V'' \lesssim 9H^2$.

(2) Denote the starting and ending values of ϕ in this interval by ϕ_s and ϕ_e respectively (note: ϕ_s must be $\gtrsim \phi_o$). The length of the interval should be much greater than H (which sets the scale of quantum fluctuations in ϕ): $\phi_e - \phi_s >> H$. This insures that quantum fluctuations will not drive ϕ across the flat region too quickly.

(3) The time required for ϕ to traverse the flat region should be $\gtrsim 60\ H^{-1}$ (to solve the homogeneity-isotropy and flatness problems). This implies that

$$\int H dt = -\int_{\phi_s}^{\phi_e} (3H^2\ d\phi/V') \gtrsim 60. \tag{4.10}$$

(4) In order to achieve an acceptable amplitude for density fluctuations, $(\delta\rho/\rho)_H \simeq H^2/\dot{\phi}(t_1)$, $\dot{\phi}$ must be $\simeq 10^4\ H^2$ when a galactic size perturbation crosses outside the horizon. This occurs about 50 Hubble times before the end of inflation.

(5) Sufficiently high reheat temperature so that the Universe is radiation-dominated at the time of primordial nucleosynthesis ($t \simeq 10^{-2} - 10^2$ sec; $T \simeq 10$ MeV - 0.1 MeV), and so that a baryon-asymmetry of the correct magnitude can evolve. As discussed earlier, the reheat temperature is:

$$T_{RH} \approx \min\{M_G,\ (\Gamma m_{pl})^{1/2}\}; \tag{4.11}$$

this must exceed $\min\{10\ \text{MeV},\ T_B\}$, where T_B is the smallest reheat temperature for which an acceptable baryon asymmetry will evolve.

(6) The potential be part of a 'sensible particle physics' model.

These conditions and a few others which are necessary for a successful implementation of new inflation are discussed in detail in ref.128. Potentials which satisfy all of the constraints tend to be very flat (for a long run in ϕ), and necessarily involve fields which are very weakly coupled (self couplings $\lesssim 10^{-10}$; see Fig. 4.6). To insure that radiative corrections do not spoil the flatness it is almost essential that the field ϕ be a gauge singlet field.

Concluding Remarks

New inflation is an extremely attractive cosmological program. It has the potential to 'free' the present state of the Universe (on scales at least as large as 10^{28} cm) from any dependence on the initial state of the Universe, in that the current state of the observable Universe in these models depends only upon microphysical processes which occurred very early on ($t \lesssim 10^{-34}$s). [I should mention that this conjecture of 'Cosmic Baldness'[129] is still just that; it has not been demonstrated that starting with the most general cosmological solution to Einstein's equations, there exist regions which undergo sufficient inflation. The conjecture however has been addressed perturbatively; pre-inflationary perturbations remain constant in amplitude, but are expanded beyond the present horizon[130] and neither shear nor negative-curvature can prevent inflation from occurring[131].]

At present there exists no completely successful model of new inflation. However, one should not despair, as I have just described, there does exist a clear-cut and straightforward prescription for the desired potential (see Fig. 4.6). Whether one can find a potential which fits the prescription and also predicts sensible particle physics remains to be seen. If such a theory is found, it would truly be a monumental achievement for the Inner Space/Outer Space connection.

Now for some sobering thoughts. The inflationary scenario does not address the issue of the cosmological constant; in fact, the small value of the cosmological constant today is its foundation. If some relaxation mechanism is found to insure that the cosmological constant is <u>always</u> small, the inflationary scenario (in its present form at least) would vanish into the vacuum. It would be fair to point out that inflation is not the only approach to resolving the cosmological puzzles discussed above. The homogeneity, isotropy, and inhomogeneity puzzles all involve the apparent smallness of the horizon. Recall that computing the horizon distance

$$d_H = R(t) \int_0^t dt'/R(t') \tag{4.12}$$

requires knowledge of R(t) all the way back to t = 0. If during an early epoch ($t \lesssim 10^{-43}$s?) R increased as or more rapidly than t (e.g. $t^{1.1}$), then $d_H \to \infty$, eliminating the 'horizon constraint'. The monopole and flatness problems can be solved by producing large amounts of entropy since both problems involve a ratio to the entropy. Dissipating anisotropy and/or inhomogeneity is one possible mechanism for producing entropy. One alternative to inflation is Planck epoch physics. Quantum gravitational effects could both modify the behaviour of R(t) and through quantum particle creation produce large amounts of entropy [see e.g., the recent review in ref. 132].

Two of the key 'predictions' of the inflationary scenario, $\Omega = 1 \pm O(10^{-BIG})$ and scale-invariant density perturbations, are such natural and compelling features of a reasonable cosmological model, that their ultimate verification (my personal bias here!) as cosmological facts will shed little light on whether or not we live in an inflationary Universe. Although the inflationary Universe scenario is not the only game in town, right now it does seem to be the best game in town.

Due to the brevity of this course in particle physics/cosmology there are many important and interesting topics which I have not covered (some of which are discussed in refs. 1-3). I apologize for any omissions and/or errors I may be guilty of. I thank my collaborators who have allowed me to freely incorporate material from co-authored works; they include E. W. Kolb, P. J. Steinhardt, G. Steigman, D. N. Schramm, K. Olive and J. Yang. This work was supported in part by the DOE (at Chicago and Fermilab), NASA (at Fermilab), and an Alfred P. Sloan Fellowship.

REFERENCES

1. Eleventh Texas Sympsoium on Relativistic Astrophysics, ed. D. S. Evans, NY Acad. Sci. 422, especially parts I, III, IV (1984); K. Freese and D. N. Schramm, Nucl. Phys. B233, 167 (1984); J. R. Bond, A. S. Szalay, and M. S. Turner, Phys. Rev. Lett. 48, 1636 (1982); J. R. Bond, G. Efstathiou, and J. Silk, Phys. Rev. Lett. 45, 1980 (1980); S. White, C. Frenk, and M. Davis, Astrophys. J. 274, L1 (1983); A. Melott etal., Phys. Rev. Lett. 51, 935 (1983); G. R. Blumenthal, S. M. Faber, J. R. Primack and M. Rees, Nature, in press (1984); P. J. E. Peebles, Astrophys. J. 263, L1 (1982); S. D. M. White and M. J. Rees, Mon. Not. R. Astron. Soc. 183, 341 (1978); Formation and Evolution of Galaxies, eds. J. Audouze and J. Tran Thanh Van (Reidel: Dordrecht, 1983).
2. M. Srednicki, Nucl. Phys. B202, 327 (1982); D. V. Nanopoulos and K. Tamvakis, Phys. Lett. 110B, 449 (1982); J. Ellis, J. S. Hagelin, D. V.Nanopoulos, K. Olive, and M. Srednicki, Nucl. Phys. B 238, 453 (1984); A. Salam and J. Strathdee, Ann. Phys. (NY) 141, 316 (1982); P. G. O. Freund, Nucl. Phys. B209, 146 (1982); P. G. O. Freund and M. Rubin, Phys. Lett. 37B, 233 (1980); E. Kolb and R. Slansky, Phys. Lett. 135B, 378 (1984).
3. P. Sikivie, "Axions and Cosmology", in Internationale Univeersitätswochen für Kernphysik der Karl-Franzens- Universität XXIth (Schladming, Austria 1982); K. Sato and H. Sato, Prog. Theor. Phys. 54, 1564 (1975); D. Dicus, E. Kolb, V. Teplitz, and R. Wagoner, Phys. Rev. D27, 839 (1980); L. Abbott and P. Sikivie, Phys. Lett. 120B, 133 (1983); M. Dine and W. Fischler, Phys. Lett. 120B, 137 (1983); J. Preskill, M. Wise, and F. Wilczek, Phys. Lett. 120B, 127 (1983); M. S. Turner, F. Wilczek, and A. Zee, Phys. Lett. 125B, 35 and 519 (1983); P. Sikivie, Phys. Rev. Lett. 48, 1156 (1982).
4. R. Buta and G. deVaucouleurs, Astrophys. J. 266, 1 (1983); A. Sandage and G. A. Tammann, Astrophys. J. 256, 339 (1982).
5. S. Weinberg, Gravitation and Cosmology (Wiley: NY, 1972).
6. Physical Cosmology, eds. R. Balian, J. Audouze, and D. N. Schramm (North-Holland: Amsterdam, 1980).
7. G. Gamow, Phys. Rev. 70, 572 (1946).
8. R. A. Alpher, J. W. Follin, and R. C. Herman, Phys. Rev. 92, 1347 (1953).
9. P. J. E. Peebles, Astrophys. J. 146, 542 (1966).
10. R. V. Wagoner, W. A. Fowler, and F. Hoyle, Astrophys. J. 148, 3(1967).
11. A. Yahil and G. Beaudet, Astrophys. J. 206, 26 (1976).
12. R. V. Wagoner, Astrophys. J. 179, 343 (1973).
13. K. Olive, D. N. Schramm, G. Steigman, M. S. Turner, and J. Yang, Astrophys. J. 246, 557 (1981).

14. D. A. Dicus etal., Phys. Rev. D26, 2694 (1982).
15. J. Yang, M. S. Turner, G. Steigman, D. N. Schramm, and K. Olive, Astrophys. J., in press (1984).
16. R. Epstein, J. Lattimer, and D. N. Schramm, Nature 263, 198 (1976).
17. M. Spite and F. Spite, Astron. Astrophys. 115, 357 (1982).
18. D. Kunth and W. Sargent, Astrophys. J. 273, 81 (1983); J. Lequeux etal., Astron. Astrophys. 80, 155 (1979).
19. S. Hawking and R. Tayler, Nature 209, 1278 (1966); J. Barrow, Mon. Not. R. Astron. Soc. 175, 359 (1976).
20. R. Matzner and T. Rothman, Phys. Rev. Lett. 48, 1565 (1982).
21. Y. David and H. Reeves, in Physical Cosmology, see ref. 6.
22. S. Faber and J. Gallagher, Ann. Rev. Astron. Astrophys. 17, 135 (1979).
23. M.Davis and P. J. E. Peebles, Ann. Rev. Astron. Astrophys. 21, 109 (1983).
24. K. Freese and D. N. Schramm, see ref. 1.
25. V. F. Shvartsman, JETP Lett. 9, 184 (1969).
26. G. Steigman, D. N. Schramm, and J. Gunn, Phys. Rev. Lett. 43, 202 (1977); J. Yang, D. N. Schramm, G. Steigman, and R. T. Rood, Astrophys. J. 227, 697 (1979); refs. 13, 15.
27. G. Arnison etal., Phys. Lett. 126B, 398 (1983).
28. D. N. Schramm and G. Steigman, Phys. Lett. B, in press (1984).
29. S. Wolfram, Phys. Lett. 82B, 65 (1979); G. Steigman, Ann. Rev. Nucl. Part. Sci. 29, 313 (1979).
30. B. Lee and S. Weinberg, Phys. Rev. Lett. 39, 169 (1977).
31. E. W. Kolb, Ph.D. thesis (Univ. of Texas, 1978).
32. M. S. Turner, in Proceedings of the 1981 Int'l. Conf. on ν Phys. and Astrophys., eds. R. Cence, E. Ma, and A. Roberts, 1, 95 (1981).
33. G. Steigman, Ann. Rev. Astron. Astrophys. 14, 339 (1976).
34. F. Stecker, Ann. NY Acad. Sci. 375, 69 (1981).
35. T. Gaisser, in Birth of the Universe, eds. J. Audouze, J.Tran Thanh Van (Editions Frontiers: Gif-sur-Yvette, 1982).
36. A. Sakharov, JETP Lett. 5, 24 (1967).
37. M. Yoshimura, Phys. Rev. Lett. 41, 281; (E) 42, 746 (1978).
38. D. Toussaint, S. Treiman, F. Wilczek, and A. Zee, Phys. Rev. D19, 1036 (1979).
39. S. Dimopoulos and L. Susskind, Phys. Rev. D18, 4500 (1978).
40. A. Ignatiev, N. Krasnikov, V. Kuzmin, and A. Tavkhelidze, Phys. Lett. 76B, 486 (1978).
41. J. Ellis, M. Gaillard, and D. V. Nanopoulos, Phys. Lett. 80B, 360; (E) 82B, 464 (1979).
42. S. Weinberg, Phys. Rev. Lett. 42, 850 (1979).
43. E. Kolb and S. Wolfram, Nucl. Phys. 172B, 224; Phys. Lett. 91B, 217 (1980); J. Harvey, E. Kolb, D. Reiss, and S. Wolfram, Nucl. Phys. 201B, 16 (1982).
44. J. N. Fry, K. A. Olive, and M. S. Turner, Phys. Rev. D22, 2953; 2977; Phys. Rev. Lett. 45, 2074 (1980).
45. D. V.Nanopoulos and S. Weinberg, Phys. Rev. D20, 2484 (1979).
46. S. Barr, G. Segrè and H. Weldon, Phys. Rev. D20, 2494 (1979).
47. G. Segre and M. S. Turner, Phys. Lett. 99B, 339 (1981).
48. J.Ellis, M. Gaillard, D. Nanopoulos, S. Rudaz, Phys. Lett. 99B, 101 (1981).
49. E. W. Kolb and M. S. Turner, Ann. Rev. Nucl. Part. Sci. 33, 645 (1983).

50. P. A. M. Dirac, Proc. Roy. Soc. A133, 60 (1931).
51. G. 't Hooft, Nucl. Phys. B79, 276 (1974).
52. A. Polyakov, JETP Lett. 20, 194 (1974).
53. T. Kibble, J. Phys. A9, 1387 (1976).
54. J. Preskill, Phys. Rev. Lett. 43, 1365 (1979).
55. Ya. B. Zel'dovich and M. Yu. Khlopov, Phys. Lett. 79B, 239 (1978).
56. M. Einhorn, D. Stein, and D. Toussaint, Phys. Rev. D21, 3295 (1980).
57. A. Guth and E. Weinberg, Nucl. Phys. B212, 321 (1983).
58. J. Harvey and E. Kolb, Phys. Rev. D24, 2090 (1981).
59. T. Goldman, E. Kolb, and D. Toussaint, Phys. Rev. D23, 867 (1981).
60. D. Dicus, D. Page, and V. Teplitz, Phys. Rev. D26, 1306 (1982).
61. J. Fry, Astrophys. J. 246, L93 (1981).
62. J. Fry and G. Fuller, U. of Chicago preprint (1983).
63. F. Bais and S. Rudaz, Nucl. Phys. B170, 149 (1980).
64. A. Linde, Phys. Lett. 96B, 293 (1980).
65. G. Lazarides and Q. Shafi, Phys. Lett. 94B, 149 (1980).
66. P. Langacker and S.-Y. Pi, Phys. Rev. Lett. 45, 1 (1980).
67. E. Weinberg, Phys. Lett. 126B, 441 (1983).
68. M. Turner, Phys. Lett. 115B, 95 (1982).
69. G. Lazarides, Q. Shafi, and W. Trower, Phys. Rev. Lett. 49, 1756 (1982).
70. J. Preskill, in The Very Early Universe, eds. G. W. Gibbons, S. Hawking, and S. Siklos (Cambridge Univ. Press: Cambridge, 1983).
71. A. Goldhaber and A. Guth, in preparation (1983-84).
72. W. Collins and M. Turner, Phys. Rev. D29, 2158 (1984).
73. M. Turner, E. Parker, T.Bogdan, Phys. Rev. D26, 1296 (1982).
74. J. P. Vallee, Astrophys. Lett. 23, 85 (1983).
75. K. Freese, J. Frieman, and M. S. Turner, U. of Chicago preprint (1984).
76. M. S. Turner, Nature 302, 804 (1983).
77. M. S. Turner, Nature 306, 161 (1983).
78. J. Bahcall, IAS preprint (1984).
79. E. N. Parker, Astrophys. J. 160, 383 (1970).
80. G. Lazarides, Q. Shafi, and T. Walsh, Phys. Lett. 100B, 21 (1981).
81. Y. Rephaeli and M. Turner, Phys. Lett. 121B, 115 (1983).
82. J. Arons and R. Blandford, Phys. Rev. Lett. 50, 544 (1983).
83. E. Salpeter, S. Shapiro, and I. Wasserman, Phys. Rev. Lett. 49, 1114 (1982).
84. V. A. Rubakov, JETP Lett. 33, 644 (1981); Nucl. Phys. B203, 311 (1982).
85. C. Callan, Phys. Rev. D25, 2141 (1982); D26, 2058 (1982).
86. E. Kolb and M. S. Turner, Astrophys. J., in press (1984).
87. S. Dawson and A. Schellekens, Phys. Rev. D27, 2119 (1983).
88. E. Weinberg, D. London, and J. Rosner, Nucl. Phys. B236, 90 (1984).
89. J. Harvey, M. Ruderman, and J. Shaham, in preparation (1983-84).
90. K. Freese, M. S. Turner, and D. N. Schramm, Phys. Rev. Lett. 51 1625 (1983).
91. E. Kolb, S. Colgate, and J. Harvey, Phys. Rev. Lett. 49, 1373 (1982).
92. S. Dimopoulos, J. Preskill, and F. Wilczek, Phys. Lett. 119B, 320 (1982).
93. K. Freese, Astrophys. J., in press (1984).
94. B. Cabrera, Phys. Rev. Lett. 48, 1378 (1982).

95. B. Cabrera, M. Taber, R. Gardner, and J. Bourg, Phys. Rev. Lett. 51, 1933 (1983).
96. P. B. Price, S. Guo, S. Ahlen, and R. Fleischer, Phys. Rev. Lett. 52, 1265 (1984).
97. Magnetic Monopoles, eds. R. Carrigan and W. Trower (Plenum: NY, 1983).
98. Proceedings of Monopole '83, eds. J. Stone (Plenum: NY, 1984).
99. S. Dimopoulos, S. Glashow, E. Purcell, and F. Wilczek, Nature 298, 824 (1982).
100. K. Freese and M. Turner, Phys. Lett. 123B, 293 (1983).
101. A. Guth, Phys. Rev. D23, 347 (1981).
102. A. Linde, Phys. Lett. 108B, 389 (1982).
103. A. Albrecht and P. Steinhardt, Phys. Rev. Lett. 48, 1220 (1982).
104. R. Sachs and A. Wolfe, Astrophys. J. 147, 73 (1967).
105. P. J. E. Peebles, The Large-Scale Structure of the Universe (Princeton Univ. Press: Princeton, 1980).
106. R. H. Dicke and P. J. E. Peebles, in General Relativity: An Einstein Centenary Survey eds. S. Hawking and W. Israel (Cambridge Univ. Press: Cambridge, 1979).
107. S. Hawking, I. Moss, and J. Stewart, Phys. Rev. D26, 2681 (1982).
108. A. Albrecht, P. Steinhardt, M. Turner, and F. Wilczek, Phys. Rev. Lett. 48, 1437 (1982).
109. L. Abbott, E. Farhi, and M. Wise, Phys. Lett. 117B, 29 (1982).
110. A. Dolgov and A. Linde, Phys. Lett. 116B, 329 (1982).
111. M. S. Turner, Phys. Rev. D28, 1243 (1983).
112. J. Uson and D. Wilkinson, Astrophys. J. 277, L1 (1984).
113. D. Wilkinson, in Inner Space/Outer Space, eds. E. Kolb, M. Turner, D. Lindley, K. Olive, D. Seckel (U. of Chicago Press: Chicago, 1984).
114. G. Lazarides and Q. Shafi, in The Very Early Universe (see ref. 70).
115. J. Bardeen, Phys. Rev. D22, 1882 (1980).
116. Ya. B. Zel'dovich, Mon. Not. R. Astron. Soc. 160, 1p (1972); E. R. Harrison, Phys. Rev. D1, 2726 (1970).
117. S. Hawking, Phys. Lett. 115B, 295 (1982).
118. A. Starobinskii, Phys. Lett. 117B, 175 (1982).
119. A. Guth and S.-Y. Pi, Phys. Rev. Lett. 49, 1110 (1982).
120. J. Bardeen, P. Steinhardt, and M. Turner, Phys. Rev. D28, 679 (1983).
121. Ya. B. Zel'dovich, Mon. Not. R. Astron. Soc. 192, 663 (1980); A. Vilenkin, Phys. Rev. Lett. 46, 1169, 1496 (E) (1981).
122. A. Linde, Phys. Lett. 116B, 335 (1982).
123. A. Vilenkin and L. Ford, Phys. Rev. D26, 1231 (1982).
124. A. Albrecht etal., Nucl. Phys. B229, 528 (1983).
125. J. Ellis, D. Nanopoulos, K. Olive, and K. Tamvakis, Nucl. Phys. B221, 524 (1983); Phys. Lett. 120B, 331 (1983); D. Nanopoulos etal., Phys. Lett. 127B, 30; 123B, 41 (1983); G. Gelmini, etal, Phys. Lett. 131B, 161 (1983).
126. B. Ovrut and P. Steinhardt, Phys. Lett. 133B, 161 (1983); R. Holman, P. Ramond, and G. Ross, Phys. Lett. 137B, 343 (1984).
127. Q. Shafi and A. Vilenkin, Phys. Rev. Lett. 52, 691 (1984); S.-Y. Pi, Phys. Rev. Lett. 52, 1725 (1984).
128. P. Steinhardt and M. Turner, Phys. Rev. D29, 2162 (1984).
129. W. Boucher and G. Gibbons in The Very Early Universe, ref. 70; S. Hawking and I. Moss, Phys. Lett. 110B, 35 (1982).

130. J. Frieman and M. Turner, Phys. Rev. D30, in press (1984); R. Brandenberger and R. Kahn, Phys. Rev. D29, 2172 (1984).
131. G. Steigman and M. Turner, Phys. Lett. 128B, 295 (1983).
132. J. Hartle, in The Very Early Universe, ref. 70.