Fermi National Accelerator Laboratory

FERMILAB-Conf-86/21-A
January 1986

COSMOLOGY AND PARTICLE PHYSICS

Michael S. Turner

NASA/Fermilab Astrophysics Center
Fermi National Accelerator Laboratory
Batavia, Illinois 60510

and

Departments of Physics and Astronomy and Astrophysics
Enrico Fermi Institute
The University of Chicago
Chicago, Illinois 60637

To be published in the Proceedings of the 1st and 2nd Jerusalem Winter Schools, eds. T. Piran and S. Weinberg (held in Jerusalem, December 1983, 1984).

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

COSMOLOGY AND PARTICLE PHYSICS

Michael S. Turner

NASA/Fermilab Astrophysics Center
Fermi National Accelerator Laboratory
Batavia, Illinois 60510

and

Departments of Physics and Astronomy and Astrophysics
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 x 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 6 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 Lecture 4 I will discuss how a very early ($t \lesssim 10^{-34}$ sec) phase transition associated with spontaneous symmetry breaking (SSB) has the potential to explain a handful of very fundamental cosmological facts, facts which can be accommodated by the standard cosmology, but which are not 'explained' by it. The 5th Lecture will be devoted to a discussion of structure formation in the Universe. For at least a decade cosmologists have had a general view of how structure developed, but have been unable to fill in details because of the lack of knowledge of the initial data for the problem (quantity and composition of the matter in the Universe and the nature of the initial density perturbations). The study of the very early Universe has provided us with important hints as to the initial data, and this has led to significant progress in our understanding of how structure formed in the Universe. There is perhaps no better example of the interplay between astrophysics, cosmology, and particle physics than the axion. In Lecture 6 I will review the very interesting case study of the axion -- the astrophysical and cosmological constraints on its properties, and its possible cosmological implications.

By selecting just 6 topics I have left out some other very important and interesting topics -- the deconfinement/chiral transition of QCD, supersymmetry/supergravity and cosmology, superstrings, and cosmology in extra dimensions -- to mention just a few. I refer the interested reader to references 1 and 2.

LECTURE 1 -- THE STANDARD COSMOLOGY

The hot big bang cosmology -- the so-called standard model of cosmology, neatly 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 in all likelihood ^{3}He and ^{7}Li as well). Light received from the most distant QSOs and galaxies observed left these objects when the Universe was only a few billion years old, and so observations of QSOs and distant galaxies allow us to directly probe the Universe to within a few billion years of 'the bang'. To date, all such observations are consistent with the standard cosmology. [Many QSOs with redshifts in excess of 3 have been observed, as well as many galaxies with redshifts in excess of 1; the current record holders are: for QSOs a redshift of 3.8 and for galaxies a redshift of 3.2; see ref. 3.] 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. Measurements made on wavelengths from 0.05 cm to 80 cm indicate its spectrum is consistent with that of a blackbody of temperature 2.75±0.05K (see Fig. 1.1 and ref. 4). Measurements of the isotropy indicate that the temperature is uniform to a part in 10^3 on angular scales ranging from 1' to 180° -- to a part in 10^4 after the dipole component is removed (see Fig. 1.2 and D. Wilkinson, ref. 4). The microwave background is a fossil record of

the Universe at that very early epoch. Both the measurement of the isotropy and the blackbody spectrum are consistent with the standard cosmology. According to 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 ^{4}He.] At present the standard cosmology passes this test with flying colors (as we shall see shortly).

I will now briefly review the standard cosmology (more complete discussions are given in ref. 5.). Throughout I will use 'high energy physics' units, where $\hbar = c = k_B = 1$. The following conversion factors may be useful

$$1\ \mathrm{GeV}^{-1} = 0.197 \times 10^{-13}\,\mathrm{cm}$$

$$1\ \mathrm{GeV}^{-1} = 0.658 \times 10^{-24}\,\mathrm{sec}$$

$$1\ \mathrm{GeV} = 1.16 \times 10^{13}\,\mathrm{K}$$

$$1\ \mathrm{GeV}^{4} = 2.32 \times 10^{17}\,\mathrm{g\ cm}^{-3}$$

$$1\ M_{\odot} = 1.99 \times 10^{33}\,\mathrm{g} \simeq 1.2 \times 10^{57}\ \text{baryons}$$

$$1\ \mathrm{pc} = 3.26\ \text{light-year} \simeq 3.09 \times 10^{18}\,\mathrm{cm}$$

$$1\ \mathrm{Mpc} = 3.09 \times 10^{24}\,\mathrm{cm}$$

$$G_N\ (\text{Newton's constant}) = 6.673 \times 10^{-8}\,\mathrm{cm}^{3}\mathrm{g}^{-1}\mathrm{sec}^{-2}$$

$$\equiv m_{pl}^{-2}$$

$$m_{pl} = 1.22 \times 10^{19}\,\mathrm{GeV}$$

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

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

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. [It is also possible to construct solutions which are locally, but not globally isotropic and homogeneous, and which are finite in extent and have spatial curvature $k \leq 0$.]

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

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

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

where ρ is the total energy density and p is the pressure. The expansion

rate H (also called the Hubble parameter or 'Hubble constant') sets the characteristic time for the growth of R(t); $H^{-1} \approx$ e-folding time for R. The present value of the 'Hubble constant' is parameterized as

$$H_o = 100h \text{ km sec}^{-1} \text{ Mpc}^{-1} ,$$
$$\approx h(10^{10}\text{yr})^{-1} .$$

The observational data strongly suggest that $1 \gtrsim h \gtrsim 0.4$ (ref. 6). 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:

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

$$\equiv \Omega - 1,$$

where $\Omega = \rho/\rho_{crit}$ and $\rho_{crit} = 3H^2/8\pi G \approx 1.88\ h^2 \times 10^{-29}\ \text{gcm}^{-3} \approx 1.05 \times 10^4 h^2 \text{eVcm}^{-3}$. A reliable and definitive determination of Ω has thus far eluded cosmologists -- the cosmic surveying required to directly determine ρ is far beyond our present 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: $\Omega \geq \Omega_{lum} \approx 0.01$.

Based upon dynamical techniques, all of which basically involve Kepler's third law in one guise or another, the observational data seems to indicate that the material which clusters with the visible galaxies

on scales $\lesssim$ 10-30 Mpc accounts for $\Omega_{GAL} \simeq 0.1$-0.3 (see Lecture 5 for further discussion).

Although Ω can, in principle, be determined by measurements of the deceleration parameter q_o,

$$q_o \equiv -(\ddot{R}/R)/H^2 ,$$
$$= \Omega(1+3p/\rho)/2 ,$$

the difficulty of reliably determining q_o probably only restricts Ω to be less than a few.[7]

The best upper limit to Ω is based upon the age of the Universe. The age of the Universe is related to the 'Hubble time', H^{-1}, by

$$t_U = f(\Omega)\ H^{-1} , \qquad (1.5)$$
$$\simeq 10^{10}\text{yr}\ h^{-1}\ f(\Omega) ,$$

where $f(\Omega)$ is a monotonically decreasing function of Ω; $f(0) = 1$, and $f(1) = 2/3$ (matter-dominated Universe) or $1/2$ (radiation-dominated Universe). [Throughout this discussion I will assume that the cosmological constant Λ is zero or negligibly small.] The dating of the oldest stars and of the elements strongly suggest that the Universe is at least 10 Byr old -- the best estimates being around 15 Byr old. From Eqn. (1.5) and $t_U \gtrsim t_{10}\ 10^{10}$yr it follows that $\Omega f^2/t_{10}^2 \gtrsim \Omega h^2$. The function Ωf^2 is monotonically increasing and bounded above by $\pi^2/4$. This means that <u>independent</u> of h, $\Omega h^2 \lesssim 2.5/t_{10}^2$. Assuming that $h \gtrsim 0.4$ and $t_{10} \gtrsim 1$, it follows that $\Omega h^2 \lesssim 1.1$ (see Fig. 1.3).

The energy density of the Universe quite naturally splits up into two parts -- that contributed by relativistic particles (today the microwave are photons and the cosmic neutrino backgrounds), and that contributed by non-relativistic particles (baryons and whatever else!). 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 ($T \gg m$) 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 $T \propto R(t)^{-1}$.

Today, the energy density contributed by relativistic particles (photons and 3 neutrino species) is negligible: $\Omega_{rel} \simeq 4 \times 10^{-5}\ h^{-2}\ (T/2.7\ K)^4$. 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) $< 4 \times 10^{-5}\ (\Omega h^2)^{-1}\ (T/2.7\ K)^4$, which corresponds to $t < 4 \times 10^{10}\ s\ (\Omega h^2)^{-2}\ (T/2.7\ K)^6$ and $T > 6\ eV\ (\Omega h^2)(2.7\ K/T)^3$, 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:

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

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

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

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

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

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

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

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 ($E \ll 10^{14}$ 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:

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

The constancy of S implies that

$$T \propto g_*(T)^{-1/3} \, R(t)^{-1}.$$

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

$$R(t) \simeq R(t_o)(t/t_o)^{1/2}, \tag{1.11}$$

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

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

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

$$d_H(t) = R(t) \int_0^t dt'/R(t') \tag{1.13}$$

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

If $R \propto t^n$ $(n < 1)$, then the horizon distance is finite and $\approx t \approx H^{-1}$. Note that even if $d_H(t)$ diverges (e.g., if $R \propto t^n$, $n \gtrsim 1$), 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:

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

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

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

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

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

where I have assumed that n_B/s has remained constant and has the value $\approx 10^{-10}$.

Although our verifiable knowledge of the early history of the Universe only takes us back to $t \approx 10^{-2}$ s and $T \approx 10$ 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 $\approx 10^{-43}$ s and temperatures as high as $\approx 10^{19}$ 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. 1.5.

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, $n_L/n_\gamma = (n_{e^-} - n_{e^+})/n_\gamma + (n_\nu - n_{\bar{\nu}})/n_\gamma \approx \eta + (n_\nu - n_{\bar{\nu}})/n_\gamma$, like the baryon

number ($\approx\eta$) is small. Having swallowed these assumptions, the rest follows like 1-2-3.

Frame 1: $t \approx 10^{-2}$ sec, $T \approx 10$ MeV. The energy density of the Universe is dominated by relativistic species: γ, e^+e^-, $\nu_i\bar{\nu}_i$ ($i = e, \mu, \tau, \ldots$); $g_* \approx 10.75$ (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 $\Delta m = m_n - m_p$. At high temperatures ($\gtrsim 0.3$ MeV) all the light isotopes are in nuclear statistical equilibrium with very small abundances, due to the very high photon to nucleon ratio, $\eta^{-1} \approx 10^{10}$:

$$n_D/n_B \approx \eta(T/m_N)^{3/2} \exp(2.2 \text{ MeV}/T) ,$$
$$\approx 10^{-13}$$
$$n_{He}/n_B \approx 0.2\, \eta^3\, (T/m_N)^3 \exp(28 \text{ MeV}/T) ,$$
$$\approx 4 \times 10^{-36} ,$$

where the abundances were evaluated for $T \approx 10$ MeV, and 2.2 MeV and 28 MeV are the binding energies of D and ^{4}He respectively. The fact that very little nucleosynthesis has taken place when $T \approx 1$ MeV is clearly traceable to the large value of η^{-1}. Of course, had the binding energies of the light nuclei been more like 30-100 MeV rather than 3-30 MeV nucleosynthesis would already be occurring when $T \approx 10$ MeV. Since D is the first stepping stone in the path of nucleosynthesis, the absence of substantial nucleosynthesis until much lower temperatures is usually

blamed on its small equilibrium abundance and binding energy, and this pheonomenon is often referred to as 'the deuterium bottleneck'.

Frame 2: $t \approx 1$ sec, $T \approx 1$ 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 $\approx 1/6$, very slowly decreasing, due to occasional free neutron decays. A little bit later ($T \approx m_e/3$), 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 $T_\nu \approx (4/11)^{1/3} T_\gamma$. The so-called 'deuterium bottleneck' continues to operate, preventing nucleosynthesis.

Frame 3: $t \approx 200$ sec, $T \approx 0.1$ MeV. If ^{4}He were to track its equilibrium abundance, then n_{He}/n_B would reach order unity at a temperature of about 0.3 MeV. However, the equilibrium abundances of ^{3}H, ^{3}He, and D are too small at this temperature to allow ^{4}He to be produced rapidly enough to achieve its equilibrium value. At a temperature of about 0.1 MeV, there is sufficient ^{3}H, ^{3}He, and D to produce ^{4}He at a rate comparable to the expansion rate, and nucleosynthesis begins in earnest. Essentially all the neutrons present are quickly incorporated into ^{4}He nuclei. As the D and ^{3}He are depleted, the rates at which they are burned into ^{4}He fall ($\Gamma \propto n(^3\text{He})$ or $n(D)$); eventually they drop below the expansion rate and so 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 (via electron capture) 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, $N_\nu \gtrsim 2$. 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 results in less D and ^{3}He remaining unprocessed. Because of large uncertainties in the rates of some reactions which create and destroy ^{7}Li, the predicted primordial abundance of ^{7}Li is only accurate to within about a factor of 2.

In 1946 Gamow[8] suggested the idea of primordial nucleosynthesis. In 1953, Alpher, Follin, and Herman[9] all but wrote a code to determine the primordial production of ^{4}He. Peebles (in 1966) and Wagoner, Fowler, and Hoyle (in 1967) wrote codes to calculate the primordial abundances[10]. Yahil and Beaudet[11] (in 1976) independently developed a nucleosynthesis code and also extensively explored the effect of large lepton number ($n_\nu - n_{\bar{\nu}} \simeq O(n_\gamma)$) 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 ^{7}Li 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. 1.6, 1.7 show the predicted abundances of D, ^{3}He, ^{4}He, and ^{7}Li, 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 ^{3}He. These determinations are consistent with a pre-solar value of (D/H) $\approx$ (2 $\pm$ 1/2) x 10^{-5}. An average ISM value for (D/H) $\approx$ 2 x 10^{-5} has been derived from UV absorption studies of the local ISM ($\lesssim$ few 100 pc), with individual measurements spanning the range (1 $\div$ 4) x 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 <u>lower</u> 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 <u>lower</u> 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 ^{3}He.

The abundance of ^3He 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 ^3He 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 ^3He during the sun's approach to the main sequence, these measurements represent the pre-solar sum of D and ^3He. These determinations of D + ^3He are all consistent with a pre-solar $[(D + {}^3He)/H] \simeq (4.0 \pm 0.3) \times 10^{-5}$. Earlier measurements of the $^3He^+$ hyperfine line in galactic HII regions and very recent measurements lead to derived present abundances of ^3He: $^3He/H \simeq (3\text{-}20) \times 10^{-5}$. The fact that these values are higher than the pre-solar abundance is consistent with the idea that the abundance of ^3He should increase with time due to the stellar production of ^3He by low mass stars.

^3He is much more difficult to destroy than D. It is very hard to efficiently dispose of ^3He without also producing heavy elements or large amounts of ^4He (environments hot enough to burn ^3He are usually hot enough to burn protons to ^4He). 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 ^3He without overproducing ^4He or heavy elements, ^3He can have been astrated (i.e. reduced by stellar burning) by a factor of no more than $f_a \simeq 2$. [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

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

$$\lesssim (1-f_a)\text{pre-solar}(D/H)+f_a\text{pre-solar}(D+{}^3He)/H;$$

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[47] 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 ^{7}Li in their outer layers. Indeed, they see this trend in their data, and deduce a primordial ^{7}Li 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 ^{7}Li 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 ^{7}Li), then concordance for ^{7}Li restricts η to the range $(1 - 7) \times 10^{-10}$. Note, of course, that their derived ^{7}Li 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 ^{3}He requires $\eta \approx (4 - 10) \times 10^{-10}$; moreover, concordance for D, ^{3}He, and ^{7}Li further restricts η: $\eta \approx (4 - 7) \times 10^{-10}$.

In the past few years the quality and quantity of ^{4}He observations has increased markedly. In Fig. 1.8 all the ^{4}He 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 ^{4}He is also synthesized in stars, some of the observed ^{4}He 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. 1.8. From Fig. 1.8 it is also clear that there is a large primordial component to ^{4}He: $Y_p \approx 0.22 - 0.26$. 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 ^{4}He (e.g., corrections for neutral ^{4}He, 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 several HII regions. Although observers have quoted statistical uncertainties of $\Delta Y \cong \pm 0.01$ (or lower), from the scatter in Fig. 1.8 it is clear that the systematic uncertainties must be larger. For example, different observers have derived ^{4}He 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 ^{4}He determinations for metal-poor objects. From Fig. 1.8 $Y_p \cong 0.23 - 0.25$ 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 Kunth 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 primordial abundance $Y_p \simeq 0.245 \pm 0.003$; allowing for a 3σ variation this suggests $0.236 \lesssim Y_p \lesssim 0.254$.

For the concordance range deduced from D, ^{3}He, and ^{7}Li ($\eta \gtrsim 4 \times 10^{-10}$) 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 $Y_p \cong 0.23 - 0.25$ (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: $2 \lesssim N_\nu \lesssim 4$; 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 <u>decrease</u> the primordial production of ^{4}He. Another possibility is neutrino degeneracy; a large lepton number ($n_\nu - n_{\bar{\nu}} \approx O(n_\gamma)$) 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: $4 \times 10^{-10} \leq \eta \leq 7 \times 10^{-10}$ and $N_\nu \lesssim 4$. This is the most precise determination we have of η and implies that

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

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

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

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 ^{4}He 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 ^{4}He synthesis is ≈ 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

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

$$=5.5 + 1.75N_\nu + \sum_{\text{Bose}} g_i(T_i/T)^4 + 7/8 \sum_{\text{Fermi}} g_i(T_i/T)^4. \qquad (1.19)$$

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 ^{4}He produced is $\gtrsim 0.25$ (which is greater than the observed abundance). More precisely, $N_\nu \lesssim 4$ implies

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

or

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

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 $\approx$ 190 MeV to the width of the Z^0. Preliminary results on the width of the Z^0 imply that $N_\nu \leq O(20)$[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.

Departure From Thermal Equilibrium, 'Freeze-out' and the Making of a Relic Species

Thus far I have tacitly assumed that the Universe is always in thermal equilibrium. Of course if that were always the case, the Universe would be a very boring place (Fe nuclei at 3K today!). In the strictest sense the Universe can never be in thermal equilibrium because of its expansion. However, if particle reaction rates are rapid compared to the expansion rate ($\Gamma \gg H$), then the Universe will pass through a succession of nearly equilibrium states characterized by $T \propto g_*^{-1/3} R^{-1}$. Interestingly, a massless (or very relativistic, $T \gg m$), non-interacting species which initially has an equilibrium phase space distribution will continue to do so with a temperature $T \propto R^{-1}$. The same is true for a very massive (i.e., $m \gg T$), non-interacting species, except that $T \propto R^{-2}$. [Both of these facts are straightforward to verify.]

Photons and ionized matter remain in thermal equilibrium until $T \approx 1/3$ eV, $R \approx 10^{-3} R_{today}$, and $t \approx 6.5 \times 10^{12} (\Omega h^2)^{-1/2}$ sec, when it becomes energetically favorable for the ions and electrons present to form neutral atoms (at which the scattering cross section for photons drops precipitously from $\sigma_T \approx 0.66 \times 10^{-24} cm^2$). This is the so-called decoupling or recombination epoch. After decoupling, the photons freely expand, the expansion preserves their thermal distribution with a temperature $T \propto R^{-1}$.

A given particle species can only remain in 'good thermal contact' with the photons 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 $H = \Gamma$ is called the freeze-out or decoupling temperature. [Note that if $\Gamma = aT^n$ and the Universe is radiation-dominated so that $H = (2t)^{-1} \approx 1.67\ g_*^{1/2}T^2/m_{pl}$, then the number of interactions which occur for $T \lesssim T_f$ is just: $\int_{T_f}^{0} \Gamma dt \approx (\Gamma/H)|_{T_f}/(n-2) \approx (n-2)^{-1}$]. If the species in question is very relativistic ($T_f \gg m_i$) (or very non-relativistic ($T_f \ll m$)) when it decouples, then its phase space distribution (in momentum space) remains thermal (i.e., Bose-Einstein or Fermi-Dirac) with a temperature $T_i \propto R^{-1}$ (or $T_i \propto R^{-2}$).

[It is interesting to note that based upon just the known interactions, one would not expect the Universe to be in thermal equilibrium during its earliest epochs. At high temperatures, the cross section for renormalizable interactions mediated by light gauge bosons scale as: $\sigma \approx \alpha^2/T^2$ (α = appropriate gauge coupling constant), and the number density of particles $n \approx g_*T^3$. Taking the expansion to be: $H \approx g_*^{1/2}T^2/m_{pl}$, it follows that $(\Gamma \approx n\sigma) \gtrsim H$, only for $T \lesssim g_*^{1/2}\ \alpha^2 m_{pl} \approx 10^{16}$GeV (for $g_*^{1/2}\ \alpha^2 \approx 10^{-3}$).]

Now consider the evolution of the temperature of a decoupled, relativistic species relative to that of the photons. For the decoupled species $T_i \propto R^{-1}$. However, due to the entropy release when various massive species annihilate (e.g., $e^{\pm}$ pairs when $T \approx 0.1$ MeV), the photon temperature does not always decrease as R^{-1}. Entropy conservation ($S \propto$

$g_* T^3$=constant) of course, can 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: $\langle\sigma v\rangle \approx 0.2 G_F^2 T^2$, and the number density of targets $n \approx T^3$, so that $\Gamma \approx 0.2\, G_F^2 T^5$. Equating this to H it follows that

$$T_f \approx (30\, m_{pl}^{-1} G_F^{-2})^{1/3} \tag{1.23}$$

$$\approx \text{few MeV},$$

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

$$T_\nu/T = [g_\gamma/(g_\gamma + 7/8\, g_{e^\pm})]^{1/3}$$

$$= (4/11)^{1/3}. \tag{1.24}$$

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:

$$T_i/T = [(g_\gamma + 7/8(g_{e^\pm} + N_\nu g_{\nu\bar{\nu}}))/g_{*d}]^{1/3}$$

$$\approx (10.75/g_{*d})^{1/3} \qquad (\text{for } N_\nu = 3). \tag{1.25}$$

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 $[g_* \simeq g_\gamma + g_{Gluon} + 7/8(g_{e^\pm} + g_{\mu^\pm} + g_{\nu\bar{\nu}} + g_{u\bar{u}} + g_{d\bar{d}} + g_{s\bar{s}} + ..) \gtrsim 62]$ hadronizes, and $T_i/T \simeq 0.56$; $(T_i/T)^4 \simeq 0.10$.

<u>'Hot' relics</u>- Consider a stable particle species X which decouples at a temperature $T_f \gg m_X$. 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

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

$$\simeq 0.278 g_{Xeff}/g_{*d}, \tag{1.26}$$

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

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

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

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

$$m_x/100 \text{ eV} \lesssim 0.13\, g_{*d}/g_{xeff}; \tag{1.29}$$

for a neutrino species: $T_d \cong$ few MeV, $g_{*d} \cong 10.75$, $g_{xeff} = 2 \times (3/4)$, so that $n_{\nu\bar{\nu}}/n_\gamma \cong 3/11$ 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. 1.9, 1.10.

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

$$n_x/n_\gamma \cong (g_x/2\zeta(3))(\pi/8)^{1/2}(m_x/T)^{3/2}\exp(-m_x/T), \tag{1.30}$$

$$n_x/s \cong 0.14(g_x/g_*)(m_x/T)^{3/2}\exp(-m_x/T), \tag{1.31}$$

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 $\Gamma_{ann} \approx n_x(\sigma v)_{ann} \gtrsim H$ the equilibrium abundance of Xs is maintained. When $\Gamma_{ann} \approx H$, 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 $T \lesssim T_f$:

$$n_x/s \approx (n_x/s)|_{T_f}. \tag{1.32}$$

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

$$m_x/T_f \approx \ln[0.04(\sigma v)_o m_x m_{pl} g_x g_*^{-1/2}]$$

$$+ (1/2 - n) \ln\{\ln[0.04(\sigma v)_o m_x m_{pl} g_x g_*^{-1/2}]\},$$

$$\approx 39+\ln[(\sigma v)_o m_x]+(1/2 - n)\ln[39+\ln[(\sigma v)_o m_x]], \tag{1.33}$$

$$n_x/s \approx 5\{\ln[0.04(\sigma v)_o m_x m_{pl} g_x g_*^{-1/2}]\}^{1+n}/[(\sigma v)_o m_x m_{pl} g_*^{1/2}],$$

$$\approx 4 \times 10^{-19}\{39+\ln[(\sigma v)_o m_x]\}^{1+n}/[(\sigma v)_o m_x g_*^{1/2}] \tag{1.34}$$

where $(\sigma v)_{ann}$ is taken to be $(\sigma v)_o(T/m_x)^n$ ($n = 0$ corresponds to s-wave annihilation, $n = 1$ to p-wave, <u>etc.</u>). In the second form of each equation $g_x = 2$, $g_* \approx 100$, and all dimensional quantities are to be measured in GeV units.

[The 'correct way' to solve for n_x/s is to integrate the Boltzmann equation which governs the X abundance, $d/dt\ (n_x/s) = -\langle\sigma v\rangle s[(n_x/s)^2-(n_{xeq}/s)^2]$. This has been done carefully in ref. 29 (also see ref. 30), and the 'freeze-out' approximation used in Eqns. 1.33, 1.34 is found to be reasonably good. As discussed in ref. 29 a more accurate analytic approximation gives

$$m_x/T_f \simeq \ln[0.04(n+1)\langle\sigma v\rangle_o m_x m_{pl} g_x g_*^{1/2}]$$

$$- (n + 1/2)\ln\{\ln[0.04(n+1)\langle\sigma v\rangle_o m_x m_{pl} g_x g_*^{1/2}]\},$$

$$n_x/s \simeq 3.79(n+1)(m_x/T_f)^{n+1}/[\langle\sigma v\rangle_o m_x m_{pl} g_*^{1/2}].$$

This analytic approximation for n_x/s agrees with the numerical results for n_x/s to better than 5%.]

As an example, consider a heavy (Dirac) neutrino species (m_x >> MeV), for which $\langle\sigma v\rangle \simeq O(1)\ m_x^2 G_F^2$. In the absence of annihilations this species would decouple at T ≃ few MeV which is << m_x, and so the X will become a 'cold relic'. Using Eqns. 1.33, 1.34, we find that today:

$$n_x/s \simeq 5 \times 10^{-9}/(m_x/\mathrm{GeV})^3, \tag{1.35}$$

$$\Omega_x h^2 \simeq 2(m_x/\mathrm{GeV})^{-2}, \tag{1.36}$$

implying that a stable, heavy neutrino species must be <u>more</u> massive than a few GeV. [This calculation was first done by Lee and Weinberg,[34] and independently by Kolb.[34]] Note that $\rho_x \propto n_x m_x \propto \langle\sigma v\rangle_o^{-1}$ -- implying that

the more weakly-interacting a particle is, the more 'dangerous' it is cosmologically. If a particle species is to saturate the mass density bound and provide most of the mass density today ($\Omega_x h^2 \approx 1$) then its mass and annihilation cross section must satisfy the relation:

$$(\sigma v)_o g_*^{1/2} \approx 10^{-10}\{39+\ell n[m_x(\sigma v)_o]\}^{1+n} \tag{1.37}$$

where as usual all dimensional quantities are in GeV units.

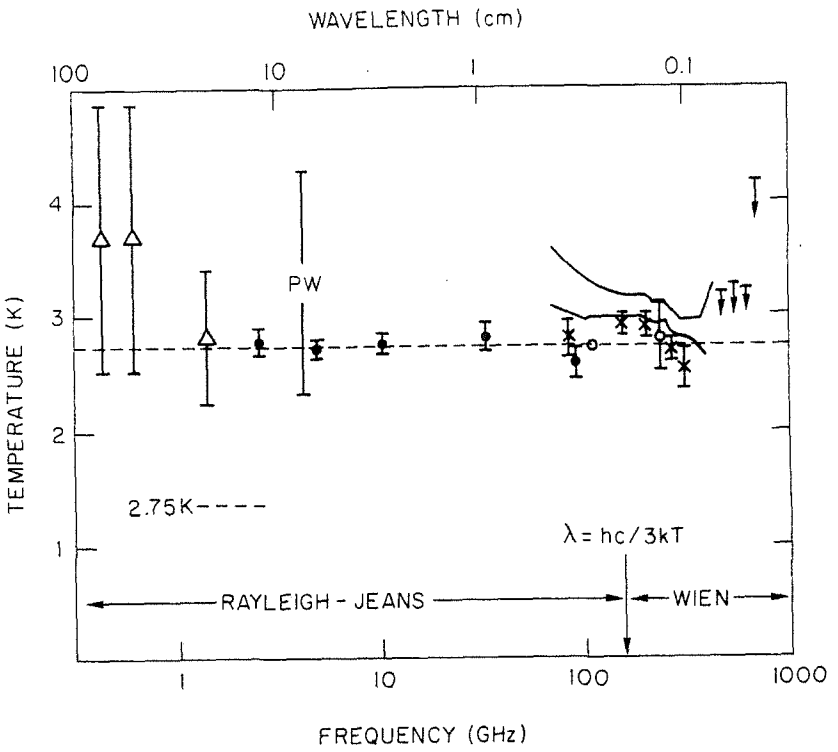


Fig. 1.1 ≈ Summary of microwave background temperature measurements from λ ≈ 0.05 to 80cm (see refs. 4). Measurements indicate that the background radiation is well-described as a 2.75 ± 0.05K blackbody. PW denotes the discovery measurement of Penzias and Wilson.

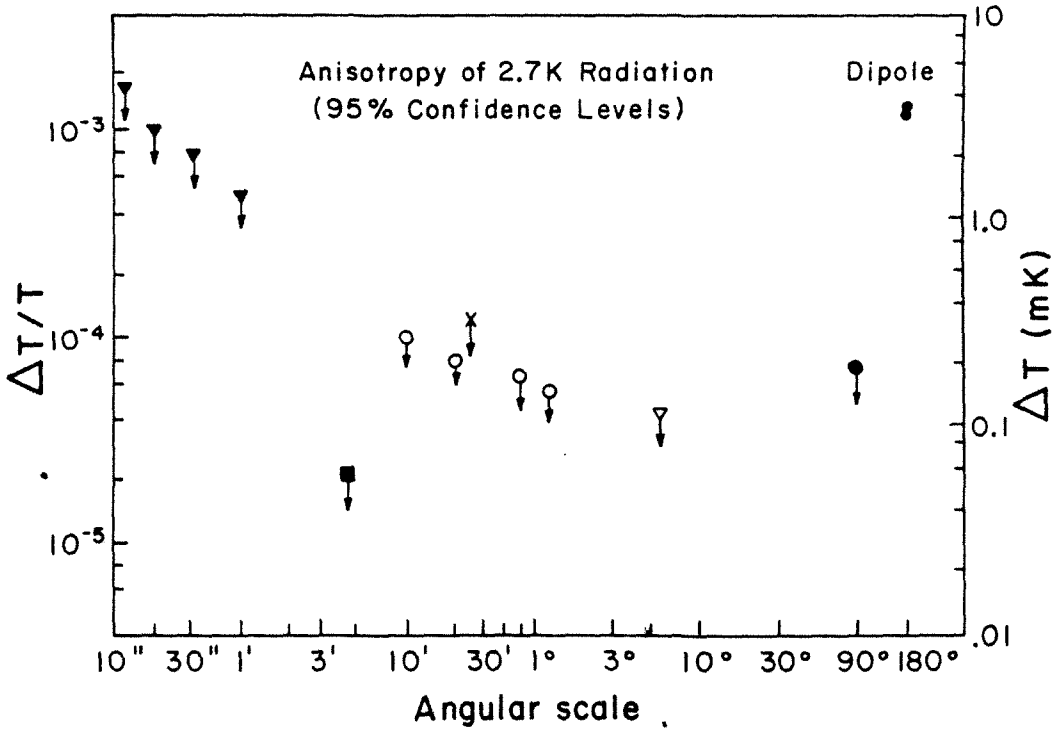


Fig. 1.2 ≈ Summary of microwave background anisotropy measurements on angular scales from 10" to 180° (see D. Wilkinson, ref. 4). With the exception of the dipole measurements, the rest are 95% confidence upper limits to the anisotrpy.

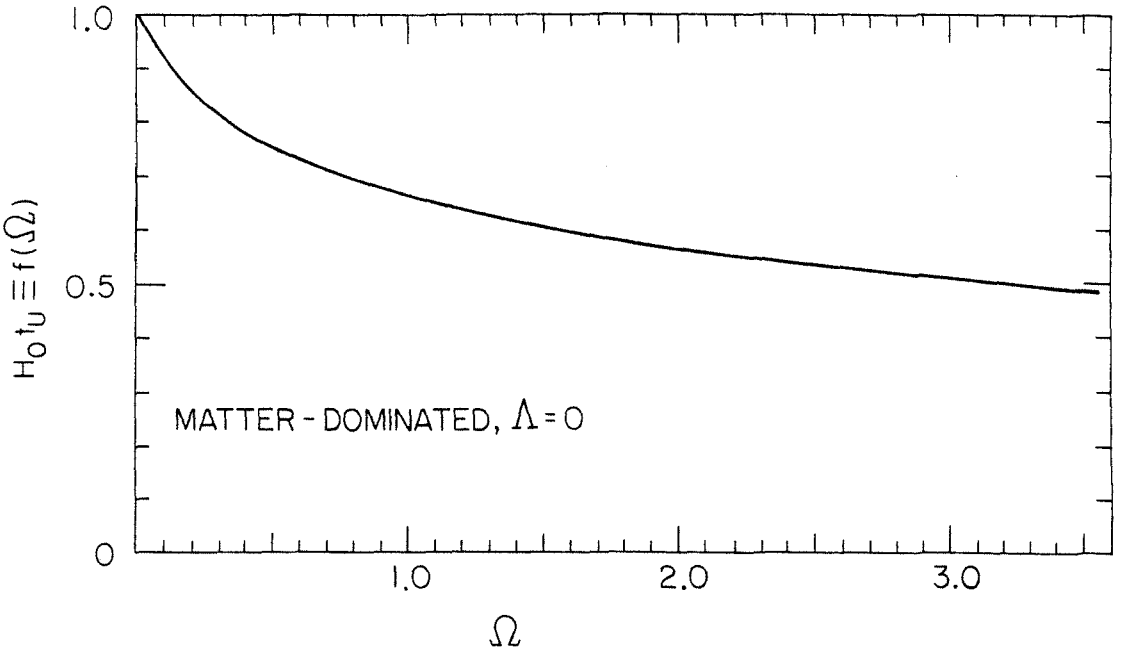


Fig. 1.3 The age of a matter-dominated, $\Lambda = 0$ Universe in Hubble units, $f(\Omega) \equiv h_0 t_U$, as a function of Ω.

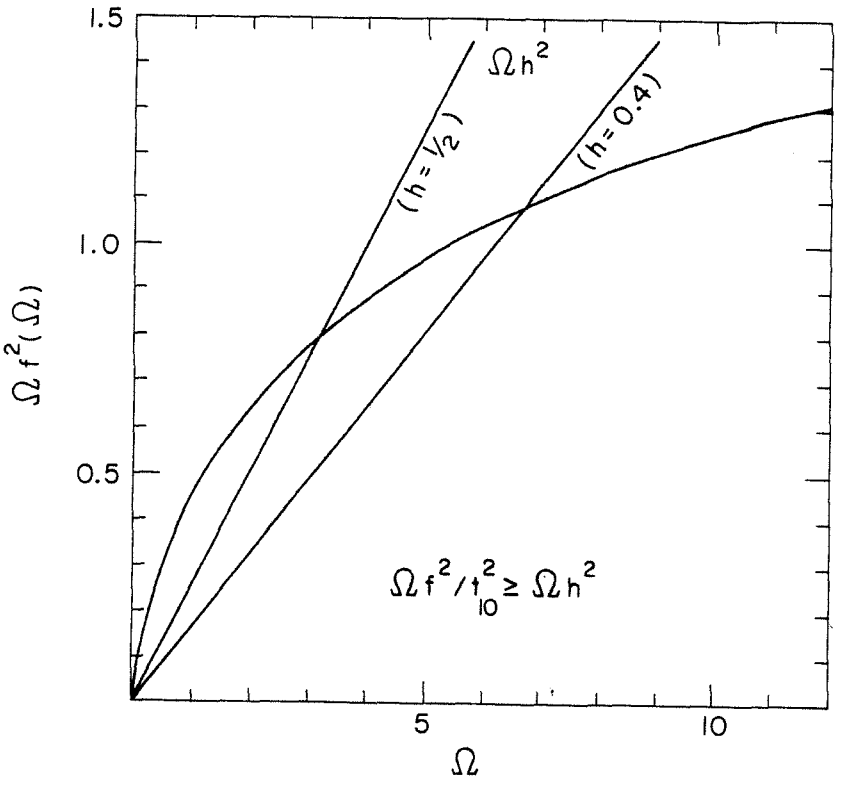
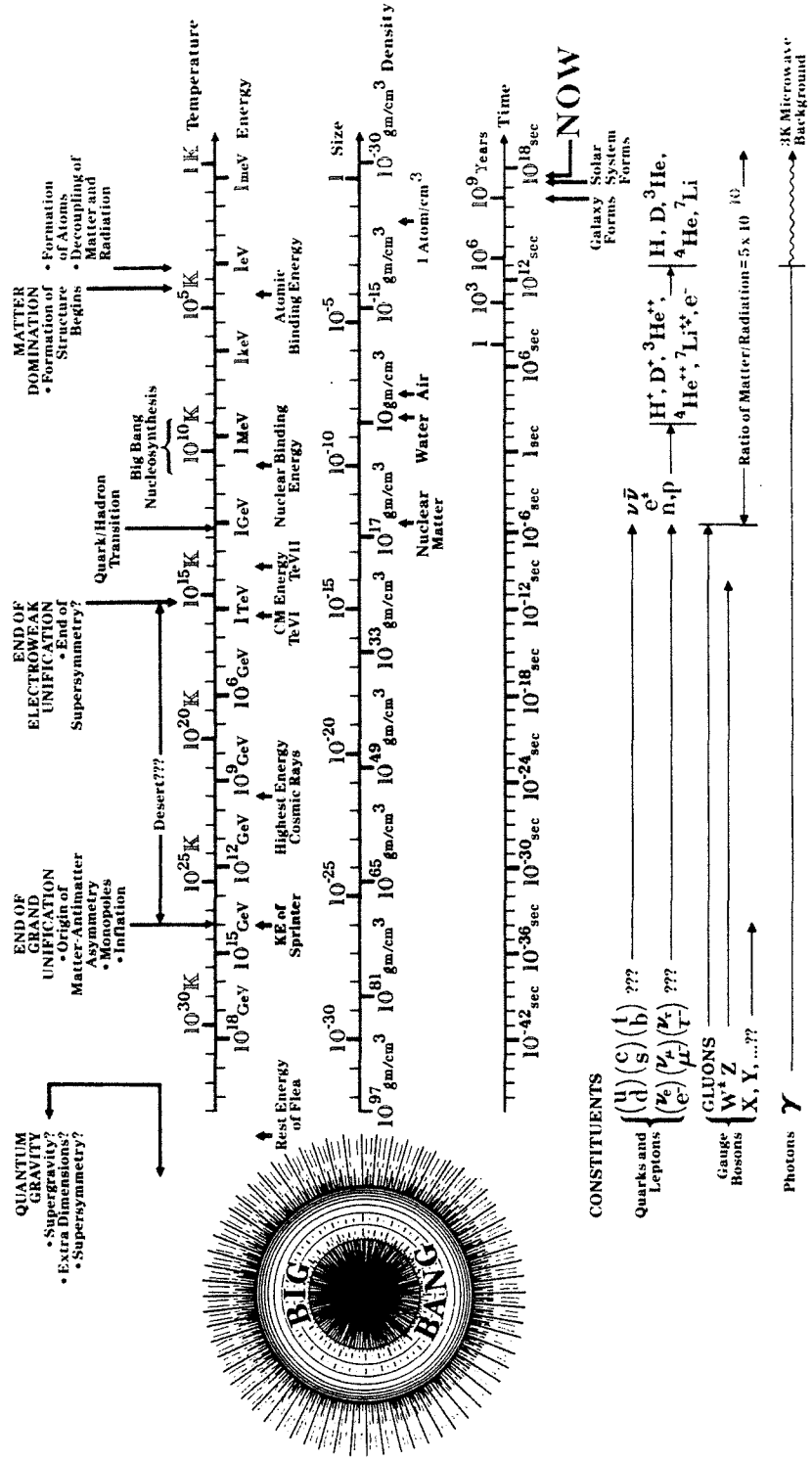


Fig. 1.4 The functions $\Omega f^2(\Omega)$ and Ωh^2 ($h = 0.4$ and 0.5). The function $\Omega f^2/t_{10}^2$ provides an upper bound to Ωh^2 (where $t_U \gtrsim t_{10}$ 10 Byr). For $t_{10} = 1$ and $h \gtrsim 0.4$ (0.5), this leads to the bounds $\Omega h^2 \lesssim 1.1$ (0.8).



• Fig. 1.5 'The Complete History of the Universe'. Highlights include: decoupling ($t \approx 10^{13}$ s, $T \approx 1/3$ 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 ($t \approx 10^{+2}$ s, $T \approx 10$ 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 ($t \approx 10^{-5}$ s, $T \approx$ 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) → U(1); GUT epoch (?? $t \approx 10^{-34}$ s, $T \approx 10^{14}$ 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 ($t \approx 10^{-43}$ s, $T \approx 10^{19}$ GeV).

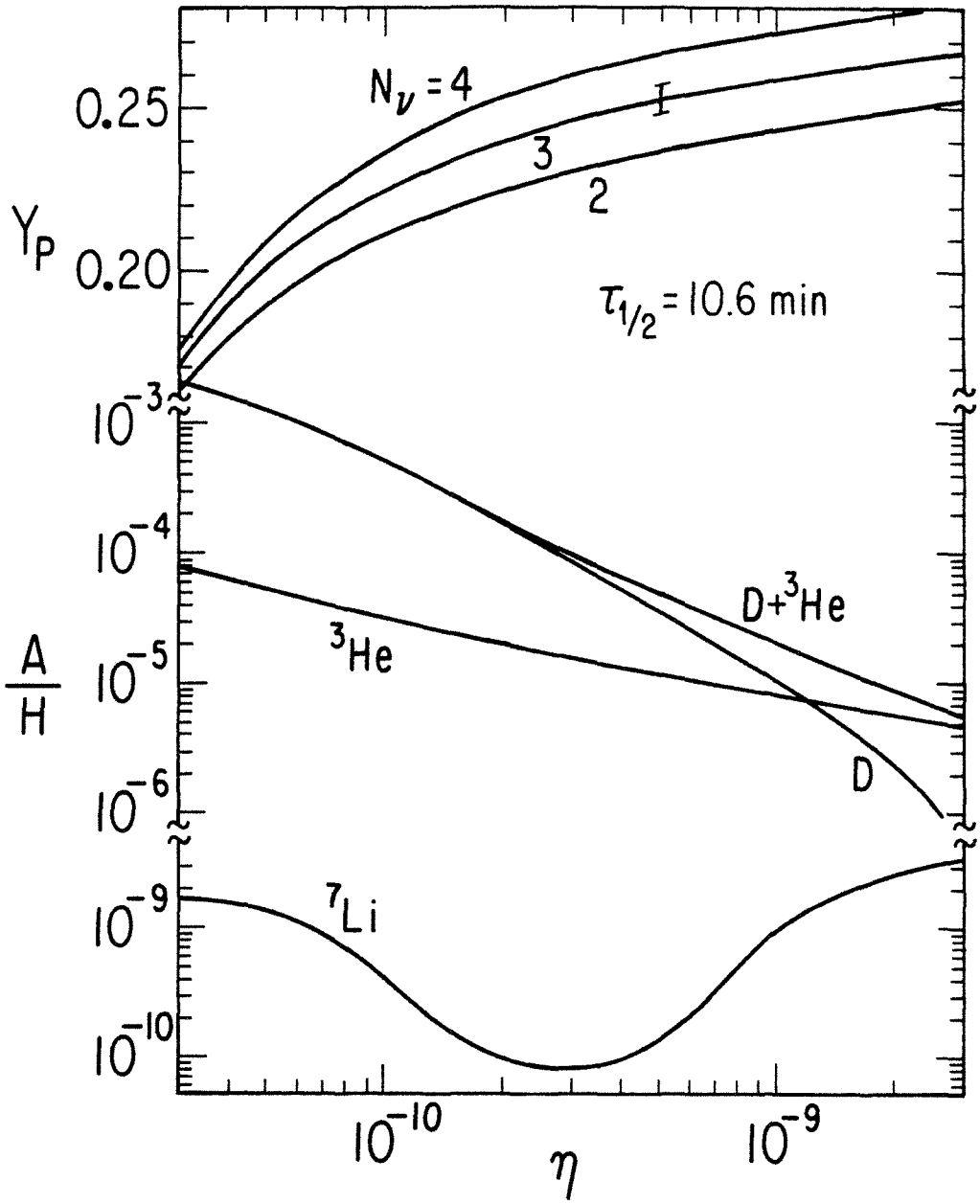


Fig. 1.6 The predicted primordial abundances of D, ^{3}He, ^{4}He, and ^{7}Li. [Note $\tau_{1/2}$ = 10.6 min was used; error bar shows $\Delta\tau_{1/2}$ = ± 0.2 min; Y_P = mass of ^{4}He; N_ν = equivalent number of light neutrino species.] Inferred primordial abundances: $Y \approx 0.23 - 0.25$; $(D/H) \gtrsim 1 \times 10^{-5}$; $(D + {}^3He)/H \lesssim 10^{-4}$; $^7Li/H \approx (1.1 \pm 0.4) \times 10^{-10}$. Concordance requires: $\eta \approx (4-7) \times 10^{-10}$ and $N_\nu \lesssim 4$.

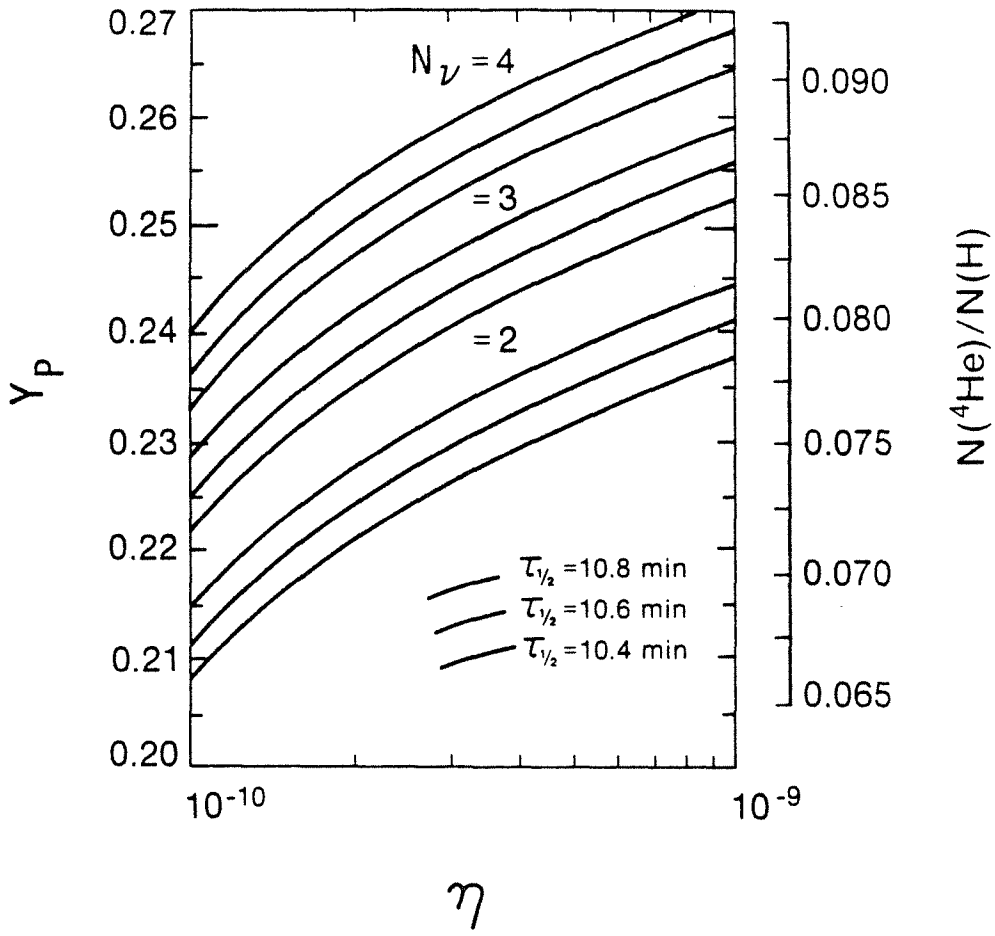


Fig. 1.7 The predicted primordial abundance of ^{4}He. Note that Y_p increases with increasing values of $\tau_{1/2}$, η, and N_ν. Hence lower bounds to η and $\tau_{1/2}$ and an upper bound to Y_p imply an upper bound to N_ν. Taking $\tau_{1/2} \approx 10.4$ min, $\eta \gtrsim 4 \times 10^{-10}$ (based on D + ^{3}He production), and $Y_p \lesssim 0.25$, it follows that N_ν must be $\lesssim 4$.

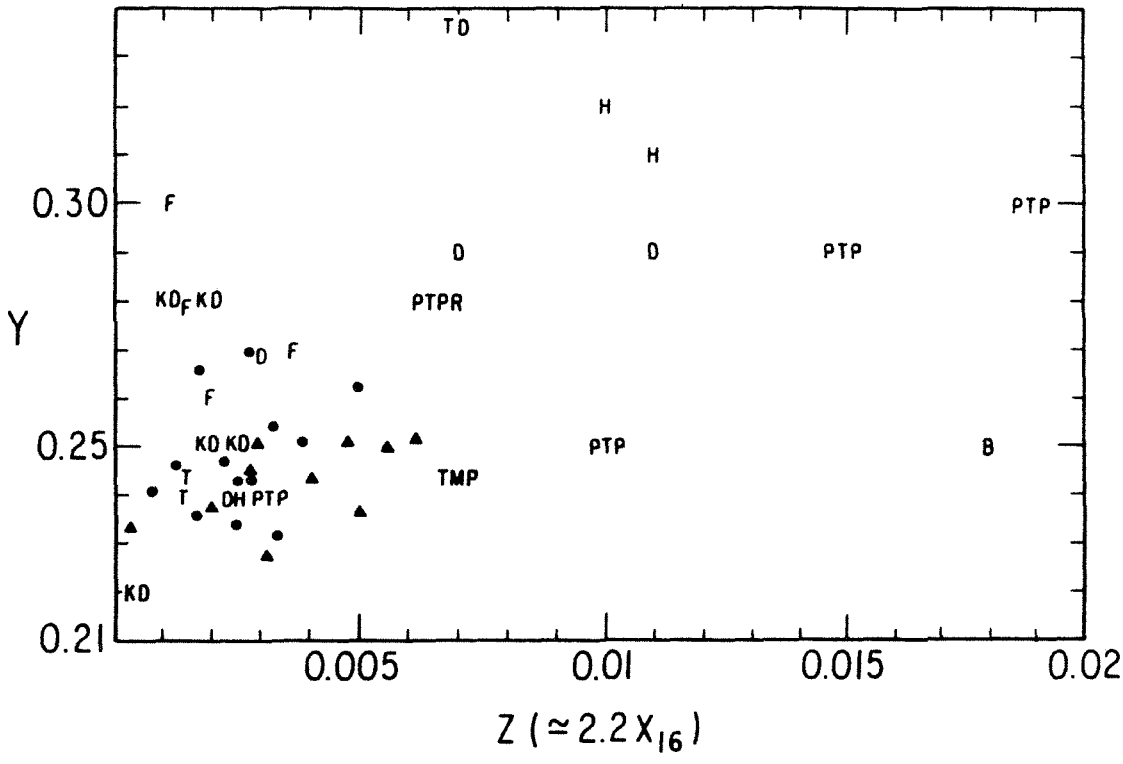


Fig. 1.8 Summary of ^{4}He abundance determinations (galactic and extragalactic) from recombination lines in HII regions vs. mass fraction of heavy (A $\geq$ 12) elements Z (≃ 2.2 mass fraction of ^{16}O). Note, observers do not usually quote errors for individual objects; scatter is probably indicative 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]).

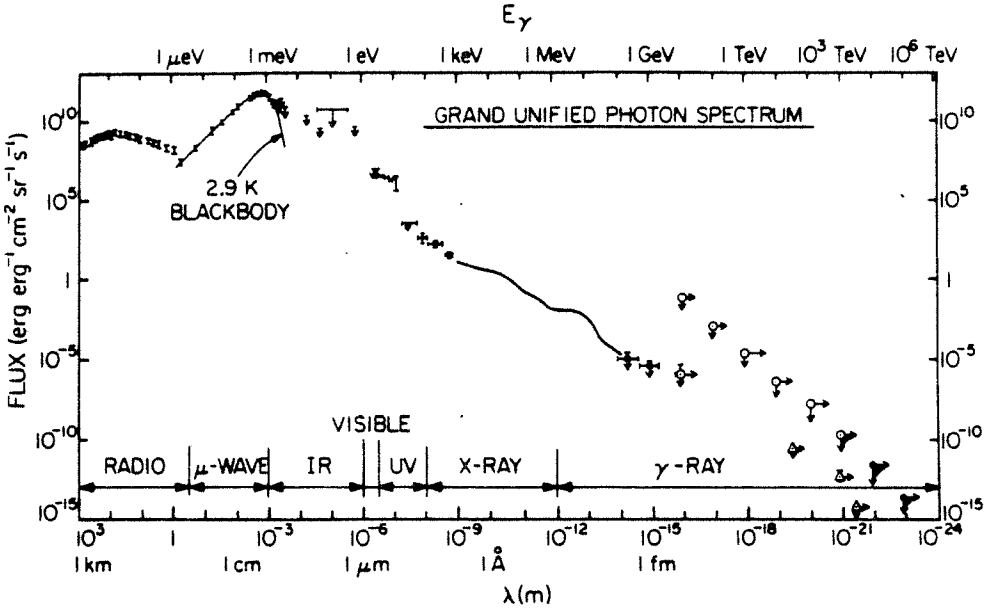


Fig. 1.9 The diffuse photon spectrum of the Universe from λ = 1 km to 10^{-24} m. Vertical arrows indicate upper limits; horizontal arrows indicate integrated flux, i.e., Flux (>E).

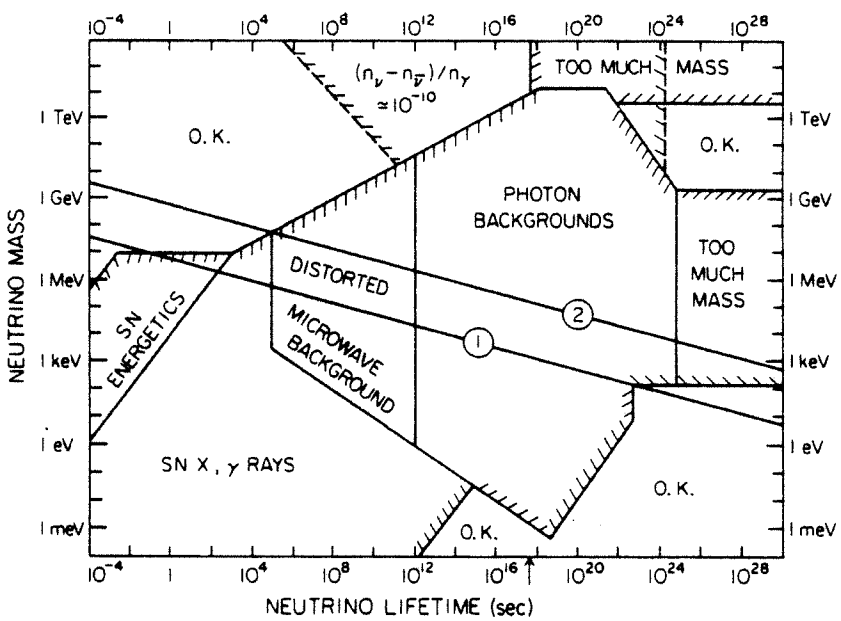


Fig. 1.10 Summary of astrophysical/cosmological constraints on neutrino masses/lifetimes. Lines 1 and 2 represent mass/lifetime relationships: $\tau = a \times 10^{-6}$ sec $(m_\mu/m_\nu)^5$, for a = 1, 10^{12}.

LECTURE 2 - BARYOGENESIS

I'll begin by briefly summarizing the evidence for the baryon asymmetry of the Universe and the seemingly insurmountable problems that render baryon symmetric cosmologies untenable. For a more detailed discussion of these I refer the reader to Steigman's review of the subject[33]. For a review of recent attempts to reconcile a symmetric Universe with both baryogenesis and the observational constraints, I refer the reader to Stecker[34].

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 abovementioned 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 $(1\text{-}100)\ M_{galaxy} \approx 10^{12}\text{-}10^{14} M_\odot$. 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 $< O(2)$. 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 $T \approx 0.5$ 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

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

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 $\sim g_* n_\gamma \approx g_* n_N \approx O(10^2) n_N$. The constancy of $n_B/s \approx O(10^{-10})$ requires that for $t < 10^{-6}$ s, $(n_N - n_{\bar{N}})/n_N (\approx 10^2 n_B/s) \approx O(10^{-8})$. 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: $T_f \approx O(20 \text{ MeV})$. Because of the incompleteness of the annihilations, residual nucleon and antinucleon to photon ratios (given by Eqn. 1.34) $n_{\bar{N}}/n_\gamma = n_N/n_\gamma \approx 10^{-18}$ 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($T \approx 20$ MeV).

Statistical fluctuations: One possible mechanism for doing this is statistical (Poisson) fluctuations. The co-moving volume that encompasses our galaxy today contains $\approx 10^{12}\ M_\odot \approx 10^{69}$ baryons and $\approx 10^{79}$ photons. Earlier than 10^{-6} s after the bang this same comoving volume contained $\approx 10^{79}$ photons and $\approx 10^{79}$ baryons and antibaryons. In order to avoid the annihilation catastrophe, this volume would need an excess of baryons over antibaryons of $\approx 10^{69}$, but from statistical fluctuations one would expect $N_b - N_{\bar{b}} \approx O(N_b^{1/2}) \approx 3 \times 10^{39}$ — 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 $\approx 2t$. When $T \approx 20$ MeV ($t \approx 10^{-3}$ 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 $\lesssim 10^{-5}\ M_\odot \ll M_{galaxy} \approx 10^{12}\ M_\odot$. [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, $n_b \gg n_{\bar{b}}$ 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, $n_B/s \sim (6-10) \times 10^{-11}$ — 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

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

At the Planck time ($\approx 10^{-43}$ 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 $T \approx g_*^{-1/4} m_{pl} \approx 3 \times 10^{18}$ GeV >> M. (Here I have taken $g_* \approx O(100)$; in minimal SU(5) $g_* \approx 160$.) So at the Planck time X, $\bar{X}$ bosons are very relativistic and up to statistical factors as abundant as photons: $n_X \approx n_{\bar{X}} \approx n_\gamma$. Nothing of importance occurs until $T \approx M$.

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

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

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 $H = |\dot{T}/T|$. The most important process in this regard is decay; other

processes (e.g. annihilation) are higher order in α. If $\Gamma_D \gg H$ for $T = M$, 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 $T = M$, 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 ($X = 1$) 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 $T \ll M$, $\Gamma_D \approx H$ (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 $\approx \exp(-M/T) \ll 1$, which greatly spresses inverse decay processes ($\Gamma_{ID} \approx \exp(-M/T)\Gamma_D \ll H$). 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 $B_1 = B_2$, $\varepsilon = 0$. 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 $\varepsilon = (r_\uparrow - \bar{r}_\uparrow + r_\downarrow - \bar{r}_\downarrow)/2$. 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 $K \equiv (\Gamma_D/H)|_{T=M} < 1$ 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 $\approx$ 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

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

$$\approx 3 \times 10^{17}\ \alpha\ \mathrm{GeV}/M.$$

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):

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

$$\Gamma_S \approx A\alpha(T/M)^5\ \Gamma_D, \tag{2.6}$$

where Γ_{ID} is the rate for inverse decays (ID), and Γ_S is the rate for $2 \leftrightarrow 2$ 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 $\Gamma_S \propto \alpha(T/M)^5\Gamma_D \propto \alpha^2T^5/M^4$. $\Gamma_S \approx n\langle\sigma v\rangle$; $n \approx T^3$ and for $T < M$, $\langle\sigma v\rangle \propto \alpha^2T^2/M^4$. 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 $\Gamma_S \approx A\alpha(T/M)^3\Gamma_D$, and $2 \leftrightarrow 2$ $\not{B}$ 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

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

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

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

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:

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

[In supersymmetric models which have dim=5 $\not{B}$ operators, $K_c(\ell nK_c)^{-1.2} \approx$

$18/A\alpha$ and the analog of Eqn. 2.9 for $K > K_c$ is: $n_B/s \approx (\varepsilon/g_*)\ A\alpha K \exp[-2(A\alpha K)^{1/2}]$.]

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

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, $\rho = \rho_x + \rho_{rad} \approx \rho_x = Mn_x$. After they decay $\rho_x \approx \rho_{rad} = (\pi^2/30)g_*\ T_{RH}^4 = (3/4)sT_{RH}$ (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

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

$$\approx (3/4)\varepsilon\ T_{RH}/M \qquad (2.10)$$

[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 $T_{RH} \leq M$. 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 <u>irrelevant</u>; it

could be by $X \to q$'s ℓ's, or equally well by $X \to \phi s \to q$'s ℓ's (ϕ = any other particle species).]

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 $\tau_X \approx \Gamma^{-1}$; when the Xs decay ($t \approx \tau_X$, $H \approx t^{-1} \approx \Gamma$): $\Gamma^2 = H^2 = 8\pi\rho_X/3m_{pl}^2$. Using the fact that $\rho_X \approx g_*(\pi^2/30)T_{RH}^4$ it follows that

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

<u>The C,CP Violation ε</u>

The crucial quantity for determining n_B/s is ≈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:

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

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

where α is the coupling of the particle exchanged in loop (i.e., α =

$g^2/4\pi$), $N \gtrsim 1$ 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 $n_B/s \approx 10^{-10}$. 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 = 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 $\varepsilon_H \approx 10^{-8}$—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 $\varepsilon_H \approx 10^{-8}$ easy to achieve.

In more complicated theories, e.g., E6, S(10), etc., $\varepsilon \approx 10^{-8}$ 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 $n_B/s \approx 10^{-10}$. 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.]

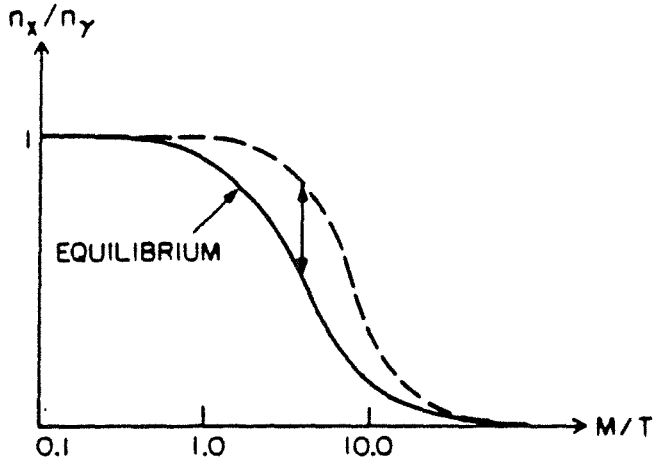


Fig. 2.1 The abundance of X bosons relative to photons. The broken curve shows the actual abundance, while the solid curve shows the equilibrium abundance.

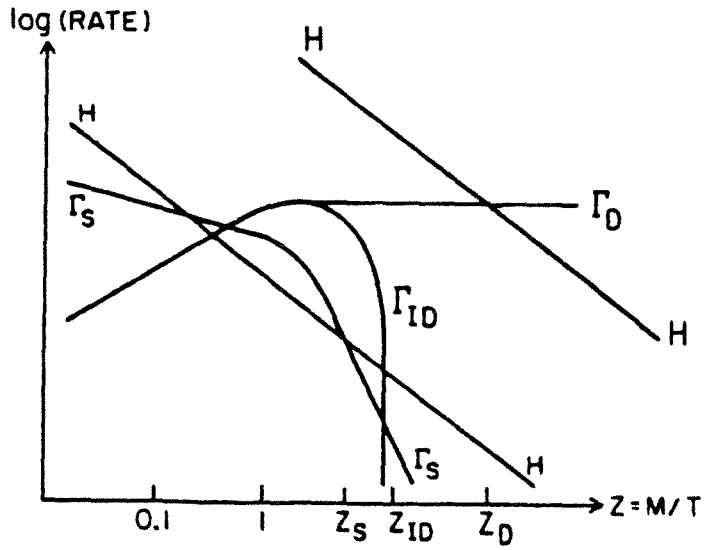


Fig. 2.2 Important rates as a function of z = M/T. H is the expansion rate, Γ_D the decay rate, Γ_{ID} the inverse decay rate, and Γ_s the $2 \leftrightarrow 2$ Ø scattering rate. Upper line marked H corresponds to case where K << 1; lower line the case where K > 1. For K << 1, Xs decay when $z = z_D$; for K > 1, freeze out of IDs and S occur at $z = z_{ID}$ and z_s.

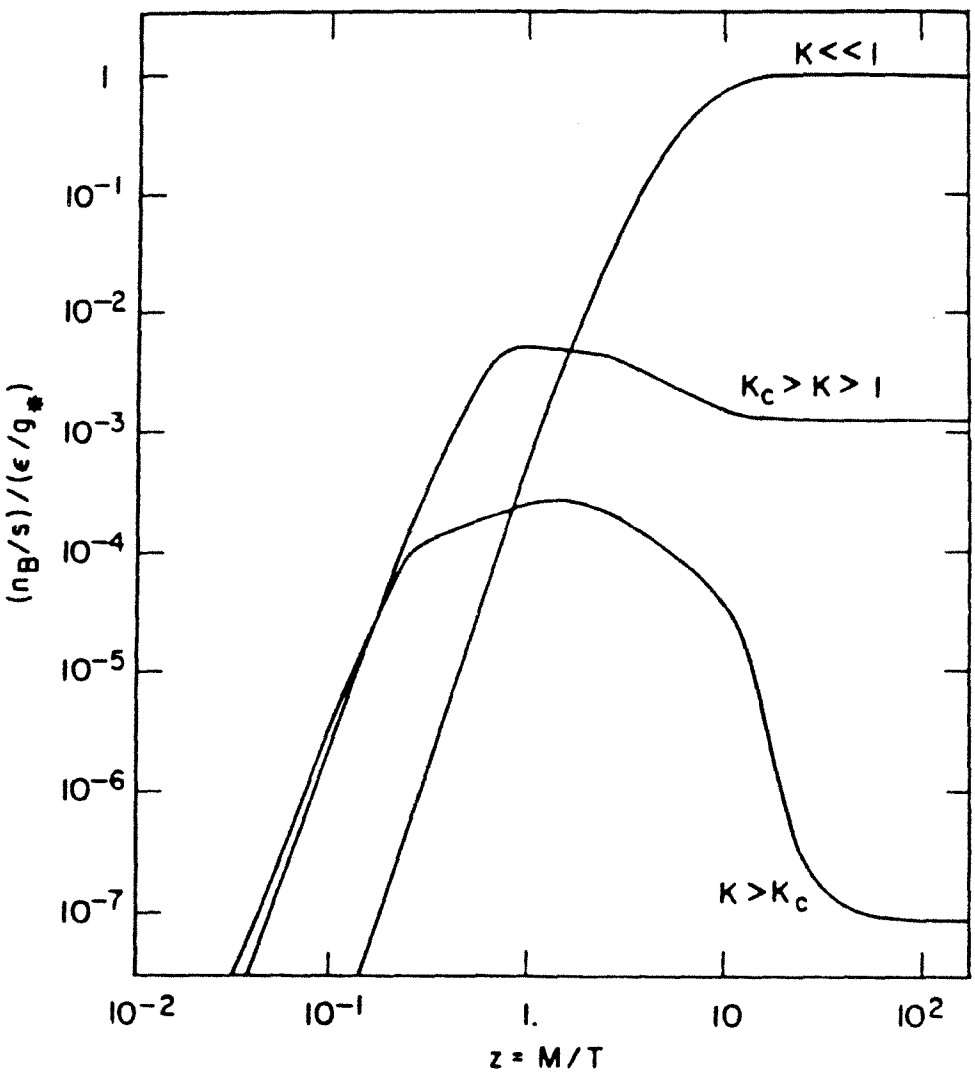


Fig. 2.3 Evolution of n_B/s as a function of $z = M/T$ ($\sim t^{1/2}$). For $K \ll 1$, n_B/s is produced when Xs decay out-of-equilibrium ($z \gg 1$). For $K_c > K > 1$, $n_B/s \propto z^{-1}$ (due to IDs) until the IDs freeze out ($z \simeq 10$). For $K > K_c$ $2 \leftrightarrow 2$ scatterings are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

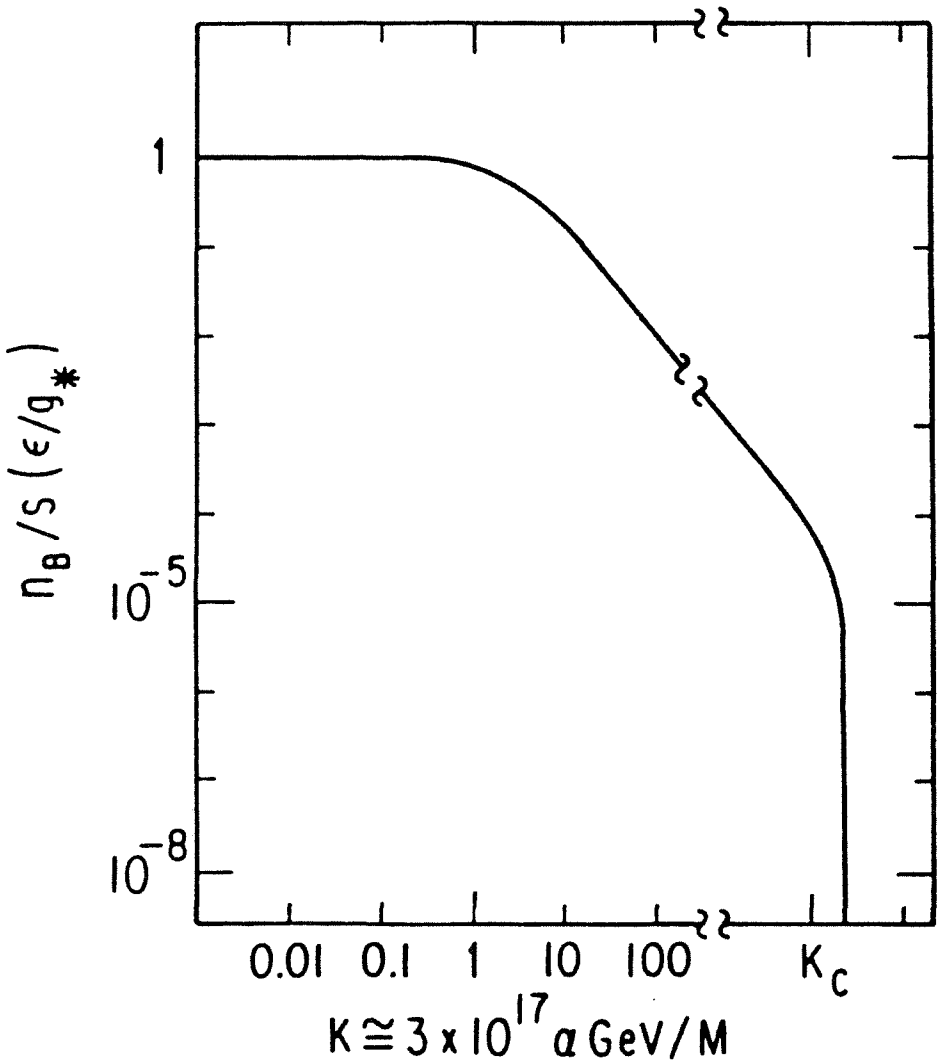


Fig. 2.4 The final baryon asymmetry (in units of ε/g_*) as a function of $K = 3 \times 10^{17}\ \alpha$ GeV/M. For $K \lesssim 1$, n_B/s is independent of K and $\simeq \varepsilon/g_*$. For $K_c > K > 1$, n_B/s decreases slowly, $\propto 1/(K(\ln K)^{0.6})$. For $K > K_c$ (when $2 \leftrightarrow 2$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

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

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

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

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 $m_M \cong O(M/\alpha)$ [$\cong 10^{16}$ 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 $\cong O(10^{14}$ 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

↑ ↑ ↑ ↑ ↑ ↑	↑ ← → ↓
no monopole	monopole

Clearly monopole configurations cannot exist until the SSB [G → G' x 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: $\xi \approx m_H^{-1}(T_G)$ (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: $\xi \lesssim d_H$($\approx$ 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 ($\approx \xi^3$) 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 ($\approx O(M)$), and as before the monopole mass by $m_M \approx O(M/\alpha)$. The age of the Universe when $T \approx T_c$ is given in the standard cosmology by: $t_c \approx 0.3\, g_*^{-1/2} m_{pl}/T_c^2$, cf. Eqn. 1.12. For SU(5): $T_c \approx 10^{14}$ GeV, $m_M \approx 10^{16}$ GeV and $t_c \approx 10^{-34}$ 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: $(n_M)_i \approx O(1)\, t_c^{-3}$. Comparing this to our fiducials S_{HOR} and $N_{B\approx HOR}$, 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:

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

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

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

$$\approx 10^{14}\langle F_M\rangle(m_M/10^{16}GeV)\ , \tag{3.3b}$$

where the monopole velocity has been assumed to be $\approx 10^{-3}c$ (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 $\approx 10^{-10}$ 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 $T \approx T_c$, this estimate for n_M/s translates into:

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

$$\Omega_M \approx 10^{11}\ (T_c/10^{14}\ GeV)^3(m_M/10^{16}\ GeV) \tag{3.4b}$$

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 $T_c \approx 10^{14}$ GeV). This is the so-called 'Monopole Problem'. The statement that $\Omega_M \approx 10^{11}$ for $T_c \approx 10^{14}$ 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 $T \approx 10^3$ GeV $(T_c/10^{14}\ \text{GeV})^3\ (m_M/10^{16}\ \text{GeV})$, and eventually reaching a temperature of 3 K at the young age of $t \approx 10^4\ \text{yrs}(T_c/10^{14}\ \text{GeV})^{-3/2}\ (m_M/10^{16}\ \text{GeV})^{-1/2}$. The requirement that $\Omega_M \lesssim O(\text{few})$ implies that

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

where I have taken m_M to be $O(100\ T_c)$. Note, given the generous estimate for ξ, even this is probably <u>not</u> safe; if one had a GUT in which $T_c \approx 10^{11}$ 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, $\langle F_M \rangle \lesssim 10^{-15}\ \text{cm}^{-2}\ \text{sr}^{-1}\ \text{s}^{-1}$, results in a slightly more stringent constraint:

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

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, $\langle F_M \rangle \lesssim 10^{-27}\ \text{cm}^{-2}\ \text{sr}^{-1}\ \text{s}^{-1}$, implies that

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

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 unlikely), 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 more 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 $\approx r_b^{-3}$. After the bubbles coalesce, and the Universe reheats, the entropy density is once again $s \approx g_* T_c^3$, so that the resulting monopole to entropy ratio is: $n_M/s \approx (g_* r_b^3 T_c^3)^{-1}$. Guth and Weinberg[57] have calculated r_b and find that $r_b \approx (m_{pl}/T_c^2)/\ln(m_{pl}^4/T_c^4)$, leading to a relatively accurate estimate for the monopole abundance:

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

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

The bottom line is that we have a serious problem here—the standard cosmology extrapolated back to $T \approx T_c$ 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):

$$\underset{T_c \approx 10^{14}\ \mathrm{GeV}}{SU(5) \rightarrow SU(3) \times SU(2) \times U(1)} \rightarrow \underset{\text{superconducting phase}}{\underset{T_1}{SU(3)} \rightarrow \underset{T_2}{SU(3) \times U(1)}}$$

(note T_1 could be equal to T_c). The key feature of their scenario is the existence of the epoch ($T \approx T_1 \rightarrow T_2$) 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 $T \approx T_c$ 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

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

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

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

resulting in

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

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

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 $T_c \approx 10^{14}$ GeV and $T_2 \approx 10^3$ 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 <u>generous</u>) 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 $T_c \ll 10^{14}$ GeV or by a perfect annihilation epoch), then the estimate for ξ <u>must</u> 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) $\rightarrow$ 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 $\ll O(10^{10}$ GeV).] The numbers produced are intrinsically small because monopole configurations do not exist in the theory until SSB occurs ($T_c \approx M$ = scale of SSB), and have a mass $O(M/\alpha) \approx 100\ M \approx 100\ T_c$. 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]

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

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

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

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

In general, $m_M/T_{max} \approx O(100)$ so that $\Omega_M \approx O(10^{-40})$ and $\langle F_M \rangle \approx O(10^{-32}\ \mathrm{cm}^{-2}\ \mathrm{sr}^{-1}\ \mathrm{s}^{-1})$—a negligible number of monopoles. However, the number produced is <u>exponentially</u> 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 $\propto$ 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!

<u>From Birth Through Adolescence ($t \approx 10^{-34}$sec to $t \approx 3\times10^{17}$sec)</u>

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 + $e^{\mp}$ $\leftrightarrow$ monopole + $e^{\mp}$) and thereby stay in <u>kinetic equilibrium</u> (KE $\approx$ 3T/2) until the epoch of $e^{\pm}$ annihilations (T $\approx$ 1/2 MeV, t $\approx$ 10 s). At the time of $e^{\pm}$ annihilations monopoles and antimonopoles should have internal velocity dispersions of:

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

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

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

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 ($\lambda_{Jeans} \approx 10^{-10}$ LY).

After decoupling ($T \approx 1/3$ eV, $t \approx 10^{13}$ 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 nonexistent 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 x 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 <u>at least</u> 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

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

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 $(10^{16}\mathrm{GeV}/m_M)^{1/2}$
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

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 ($\varepsilon \approx 1$); 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 $\gtrsim 10^{42}$ cm.

What are Monopoles Doing Today?—Astrophysical Constraints

The three most conspicuous properties of a GUT monopole are: (i) macroscopic mass ($\approx M/\alpha \approx 10^{16}$ GeV $\approx 10^{-8}$ g for SU(5)); (ii) hefty magnetic charge $h \approx n\ 69e$ ($n = \pm 1, \pm 2, \ldots$); (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., $\Omega \approx 1$). 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

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

<u>cf.</u> Eqn. 3.3b. For comparison 10^{-14} cm^{-2} sr^{-1} s^{-1} $\approx$ 30 monopoles (soccerfield)$^{-1}$ yr.$^{-1}$

<u>If</u> 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} 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

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

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 kpc)(10^{-25} 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 ρ_{halo} x 30 kpc. Comparing this to the bound on the local column density of unseen material it follows that locally $\rho_{halo} \lesssim 10^{-25}$ g cm^{-3}. Using this as the limit to the density contributed by monopoles the flux bound 3.12 becomes

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

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: $\Delta KE = -\Delta[(B^2/8\pi) \times Vol]$. 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

$$KE = hB\ell = 10^{11} GeV(B/3x10^{-6}G)(\ell/300pc), \tag{3.14}$$

$$v_{mag} \approx (2hB\ell/m_M)^{1/2}$$

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

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}, $v_o << v_{mag} \approx (2hB\ell/m)+A^{1/2}$, 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:

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

For the galactic magnetic field $B \approx 3 \times 10^{-6}$ G, $\ell \approx 300$ pc, and $v_{mag} \approx 3 \times 10^{-3}c\ (10^{16}\ GeV/m)^{1/2}$. Since $v_o \approx 10^{-3}$ 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 $O(10^{20}\,\mathrm{GeV})$ 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 $O(10^{20}\ \mathrm{GeV})$ 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 $\approx O(10^{8}\ \mathrm{yr})$. 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:

$$F \lesssim 10^{-15}\,\mathrm{cm}^{-2}\,\mathrm{sr}^{-1}\,\mathrm{s}^{-1}\,(B/3 \times 10^{-6}\ \mathrm{G})\ (3 \times 10^{7}\ \mathrm{yr}/\tau) \times$$

$$(r/30\ \mathrm{kpc})^{1/2}\ (300\ \mathrm{pc}/\ell)^{1/2}, \qquad (3.17)$$

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

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

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 $\approx 3 \times 10^{19}$ 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 $v_{ph} \approx \omega_{pl}(\ell/2\pi)$ be greater than the gravitational velocity dispersion of the monopoles ($\approx 10^{-3}c$); $\ell \approx$ wavelength of the relevant mode $\approx$ 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

$$F_M \gtrsim 1/4\ m_M\ v^3_{grav}\ (h\ell)^{-2},$$

$$\gtrsim 10^{-14}(m_M/10^{16}\text{GeV})(1\text{kpc}/\ell)^2\,\text{cm}^{-2}\,\text{sr}^{-1}\,\text{s}^{-1}. \tag{3.20}$$

Incidently, this also implies an upper bound to the oscillation period: $\tau = 2\pi/\omega_{pl} \lesssim \ell/v_{grav} \approx 3 \times 10^6$ 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 $v_{ph} \approx (\omega_{pl}/2\pi)\ell$), 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: $(\sigma v) \approx 10^{-28}$ 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 $\approx M^{-2} \approx 10^{-56}$ cm^2 ($M^{-1} \approx$ 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: $(\sigma v) \approx$ (fermion energy)$^{-2}$, or for low energies $(\sigma v) \approx$ (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}

$(\sigma v)_{-28}(\rho/1 gcm^{-3})$ per monopole; only about 10^{30} monopoles in the sun ($\approx 10^{57}$ nucleons) are needed to produce the solar luminosity ($\approx 4 \times 10^{33}$ erg s^{-1}). Here and throughout I will parameterize (σv) by:

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

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}erg s^{-1} $(\sigma v)_{-28}$ $(\rho/3\times10^{14}$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 ($\approx 3\times10^{6}$ 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 $L \approx 3 \times 10^{30}$ 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 $\langle F_M \rangle$:

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

--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 $v_M \approx 10^{-3}c$ 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

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

--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 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 may 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 ($\gtrsim 10^{10}$ 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 ($\ll 10^{-18} cm^{-2} sr^{-1} sec^{-1}$). 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!).

Monopole Hunting

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 [$dE/dx \approx (10$ 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 $\propto L^{-1} \propto \text{Area}^{-1/2}$, the simplest loop detectors are limited in size to $\lesssim 1\text{m}^2$ ($1\text{m}^2 \times 2\pi$ sr $\times$ 1yr $\approx 10^{12}$ 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^{-1} 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 $3 \times 10^{-4} - 3 \times 10^{-3}$ 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 $(\sigma v) - v_M$ 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 no 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 believable 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 firm upper limit to the flux is provided by the Parker bound[73], $\langle F_M \rangle \leq 10^{-15}$ cm^{-2} sr^{-1} s^{-1}.

Note, this is not a predicted flux, it is only a firm upper 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 possible that the largest contribution to the local monopole flux is due to a cloud of monopoles orbiting the sun with velocities $\approx$ (1 - 2) x 10^{-4} c, I think that it is very unlikely.[99,100]]

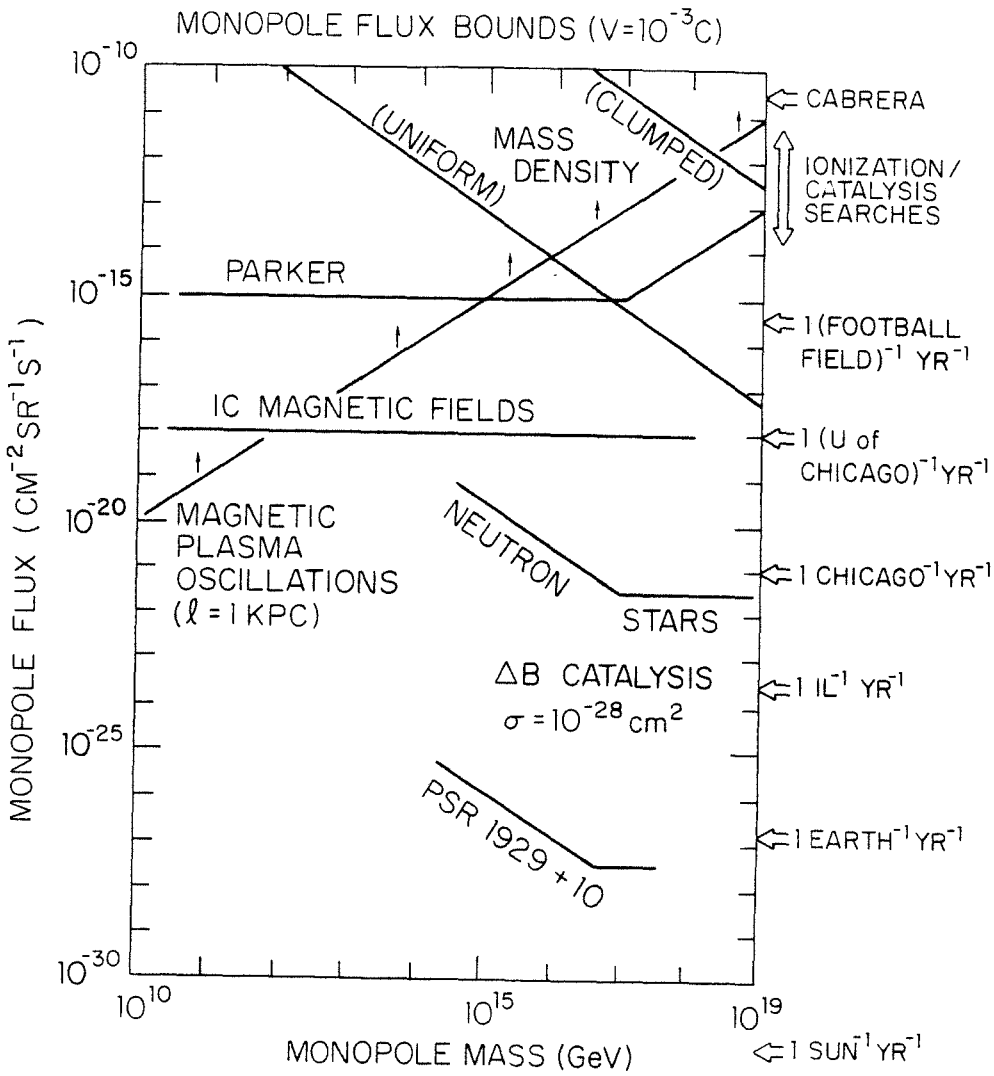


Fig. 3.1 Summary of the astrophysical/cosmological limits to the monopole flux as a function of monopole mass. Wherever necessary the monopole velocity is taken to be 10^{-3} c. The monopole catalysis bound based upon white dwarfs (ref. 93) is: $F_M \leq 2 \times 10^{-18} (\sigma v)_{-28}^{-1}$ cm^{-2} 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.

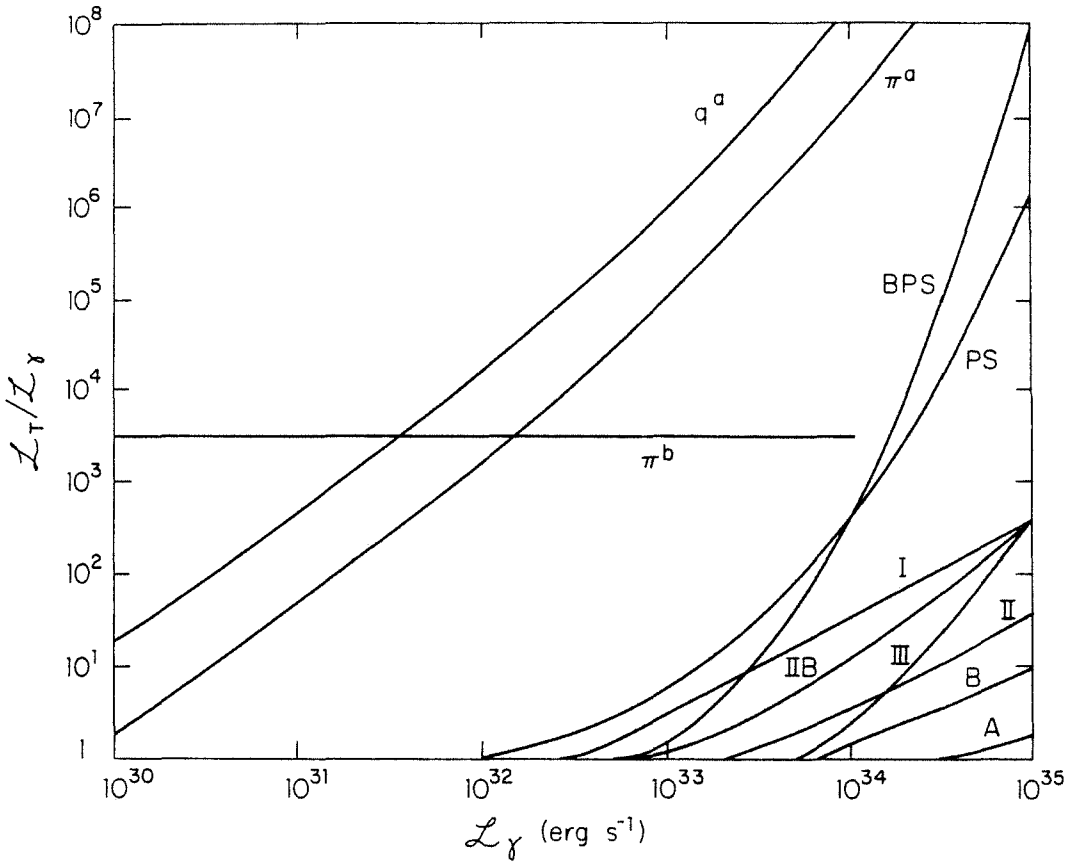


Fig. 3.2 The ratio of the total luminosity (= $L_\gamma + L_\nu$) of a hot neutron star to its photon luminosity as a function of L_γ. The different curves represent different neutron star equations of state: q (quark matter); π^a, π^b (pion condensate); the rest are more conventional equations of state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

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-Scale Homogeneity and Isotropy

The observable Universe ($d \approx H^{-1} \approx 10^{28}$ cm $\approx$ 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: $t \approx 10^{13}$ s, $T \approx 1/3$ eV) what was to become the presently observable Universe was comprised of $\approx 10^6$ 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 $\delta\rho/\rho \approx 10^5$ 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: $\theta \approx 1'\ h(M/10^{12}M_\odot)^{1/3}$.] 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, $R \approx 3 \times 10^{-5} R_{today}$ for $\Omega h^2 = 1$.] Density perturbations of amplitude $\delta\rho/\rho \approx 10^{-3}$ or so, on the scale of a galaxy ($\approx 10^{12} M_\odot$) 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 $x(t) \propto R(t)^n$ ($n = 1 \approx$ matter-dominated, or $2 \approx$ radiation-dominated). Since $\Omega \approx O(1)$ 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 $t \approx 300$ 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 $T \approx 0$ $SU(3) \times U(1)$ 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 $O(T_c^4)$ ($T_c \approx$ 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 $T \approx 0$ 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 $T \gtrsim T_c \approx M_G$ 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. 4.2). 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 ϕ($\propto R^{-3}$). 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 ($\Gamma^{-1} \equiv$ 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

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

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 $\phi = \phi_0$ and $\phi = \sigma$, then ϕ will evolve very slowly in that region, and the motion of ϕ will be 'friction-dominated' so that $3H\dot{\phi} \approx -V'$ (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 $\rho \approx V(\phi) \approx V(\phi = 0)$; both ρ_r and $1/2$ $\dot{\phi}^2$ are $\ll V(\phi)$. The expansion rate H is then just

$$H \approx (8\pi V(0)/3m_{pl}^2)^{1/2} \tag{4.5}$$

$$\approx M_G^2/m_{pl},$$

where V(0) is assumed to be of order M_G^4. While $H \approx$ constant R grows exponentially: $R \propto \exp(Ht)$; for $\tau_\phi = 100\ H^{-1}$ 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: $\omega^2 \approx V''(\sigma) \approx M_G^2$

$\gg H^2 \approx M_G^4/m_{pl}^2$. As ϕ oscillates about $\phi = 0$ its motion is damped by particle creation and the expansion of the Universe. If $\Gamma^{-1} \ll H^{-1}$, the coherent field energy density $(V + 1/2\,\dot{\phi}^2)$ is converted into radiation in less than an expansion time ($\Delta t_{RH} \approx \Gamma^{-1}$), and the patch is reheated to a temperature $T \approx O(M_G)$ ~ the vacuum energy is efficiently converted into radiation ('good reheating'). On the other hand, if $\Gamma^{-1} \gg H^{-1}$, 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 $t \approx \Gamma^{-1}$ the energy in radiation begins to dominate that in coherent field oscillations, and the patch is reheated to a temperature $T \approx (\Gamma/H)^{1/2} M_G \approx (\Gamma m_{pl})^{1/2} \ll M_G$ ('poor reheating'). The evolution of ϕ is summarized in Fig. 4.3.

For the following discussion let us assume 'good reheating' ($\Gamma \gg H$). After reheating the patch has a physical size $e^{100} H^{-1}$ ($\approx 10^{17}$cm for $M_G \approx 10^{14}$ 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 ($\approx 10^{28}$cm) had a physical size of about 10 cm (= 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. The flatness puzzle involves the smallness of the ratio of the curvature term to the energy density term. This ratio is exponentially smaller after inflation: x(after) $\approx e^{-200}$ 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 $1 \pm O(10^{-4BIG})$. 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 $\propto T_{RH}/(m_\phi \approx \omega)$ 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 $(M_G/T_{MS})^3 \exp(3H\tau_\phi) \approx$ 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, and in Lecture 5.]

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 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 $\lambda > \lambda_J \equiv v_s H^{-1}$ (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[15] 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 <u>kinematically</u> (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 <u>not</u> 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 <u>ab initio</u> 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 (≈ $10^{12} M_\odot$) 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 ($v_s \approx c$) and hence $\lambda_J \approx H^{-1}$; for $t \gtrsim 10^{13}$ s, $\lambda < H^{-1}$ and $\delta\rho/\rho \propto t^{2/3}$ since matter and radiation are decoupled ($v_s \ll c$) and $\lambda_J < \lambda_{Galaxy}$. [Note: in an $\Omega = 1$ Universe the mass inside the horizon ≈ $(t/\text{sec})^{3/2} M_\odot$.]

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 $\gtrsim 0(1)$--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 $0(10^{-4})$ 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

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

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 $\Delta\phi \approx H/2\pi$ 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 ($\lambda = H^{-1}$) <u>twice</u> (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 $\lambda_G \approx H^{-1}$) $\lambda_G < H^{-1}$ and microphysics (quantum fluctuations, etc.) can operate on this scale. When $t = t_1$ microphysics 'freezes out' on this scale; the density perturbation which 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 $\dot{\phi}$ 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 <u>same for all scales.</u> [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'' (= the effective mass-squared) is very small.]

While outside the horizon ($t_1 \leq t \leq t_H$) 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 <u>independent</u> of scale (intuitively this makes sense since this is the kinematic regime). Given these 'two facts': $(\delta\rho/\rho)_1$ = scale-independent and the kinematic evolution = 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.]

<u>Coleman-Weinberg SU(5) Model</u>

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

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

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

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), $\sigma \approx 1.2 \times 10^{15}$ GeV, and for $\phi \approx 10^9$ GeV, $\lambda(\phi) \approx 0.1$. [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 $T \approx 10^8 - 10^9$ GeV the metastability limit is reached, and thermal fluctuations drive ϕ over the T-dependent barrier (height $\approx T^4$) in the finite temperature effective potential. Naively, one expects that $\phi_0 \approx T_{MS}$ 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 $3H\dot{\phi} \approx -V'$ is valid for $\phi \ll \sigma$, and it follows that

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

where $H\tau_\phi \approx (3/2\lambda)(H/\phi_0)^2$ (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 $H \approx 7 \times 10^9$ GeV; setting $\phi_0 \approx 10^8 - 10^9$ GeV results in $H\tau_\phi \approx 0(500-50000) \approx$ 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 $\phi \gg \Delta\phi_{QM}$ (= quantum fluctuations in ϕ). In deSitter space the scale of quantum fluctuations is set by H: $\Delta\phi_{QM} \approx H/2\pi$ (on the length scale H^{-1}). Roughly speaking then, Eqn. 4.8 is only valid for $\phi \gg H$. 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 $\phi_0 \gg H$ (i.e., the potential becomes sufficiently flat that the classical part of the evolution, $\phi \gg H$, takes a time $\gtrsim 60\ H^{-1}$). In the Coleman-Weinberg model $\Gamma \gg H$ and the Universe reheats to $T \approx M_G \approx 10^{14}$ 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

$$(\delta\rho/\rho)_H \approx (4 \text{ or } 2/5)100\lambda^{1/2}[1 + \ln(M/10^{12}M_\odot)/171 \qquad (4.9)$$

$$+ \ln(g\sigma/10^{15}\ \text{GeV})/57]^{3/2},$$

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 $\lambda \approx 0.1$, results in $(\delta\rho/\rho)_H \approx O(10^2)$ which is clearly a

disaster. [On angular scales >> 1° the Zel'dovich spectrum results in temperature fluctuations of[104] $\Delta T/T \approx 1/2(\delta\rho/\rho)_H$ which must be $\lesssim 10^{-4}$ to be consistent with the observed isotropy.] To obtain perturbations of an acceptable amplitude one must artificially set $\lambda \approx 10^{-12}$ or so. [In an SU(5) GUT λ is determined by the value of $\alpha_{GUT} = g^2/4\pi \approx 1/45$, which implies $\lambda \approx 0.1$.] As mentioned earlier the density fluctuation constraint is a very severe one; recall that $\lambda \approx 10^{-2} \approx 10^{-3}$ 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., $\ddot{\phi}$ term negligible so that $3H\dot{\phi} = -V'$. 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

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

(4) In order to achieve an acceptable amplitude for density fluctuations, $(\delta\rho/\rho)_H \approx H^2/\dot{\phi}(t_1)$, $\dot{\phi}$ must be $\approx 10^4 \cdot H^2$ 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 ($t \approx 10^{-2} - 10^2$ sec; $T \approx 10$ MeV - 0.1 MeV), and so that a baryon-asymmetry of the correct magnitude can evolve. As discussed earlier, the reheat temperature is:

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

this must exceed min{10 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 always 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.

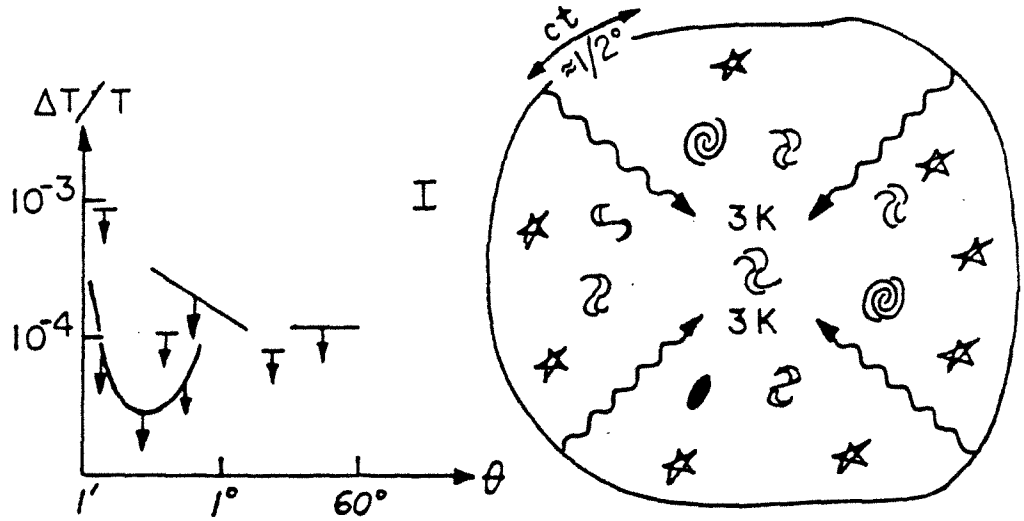


Fig. 4.1 Summary of measurements of the anisotropy of the 3K background on angular scales > 1' (from refs. 112, 113).

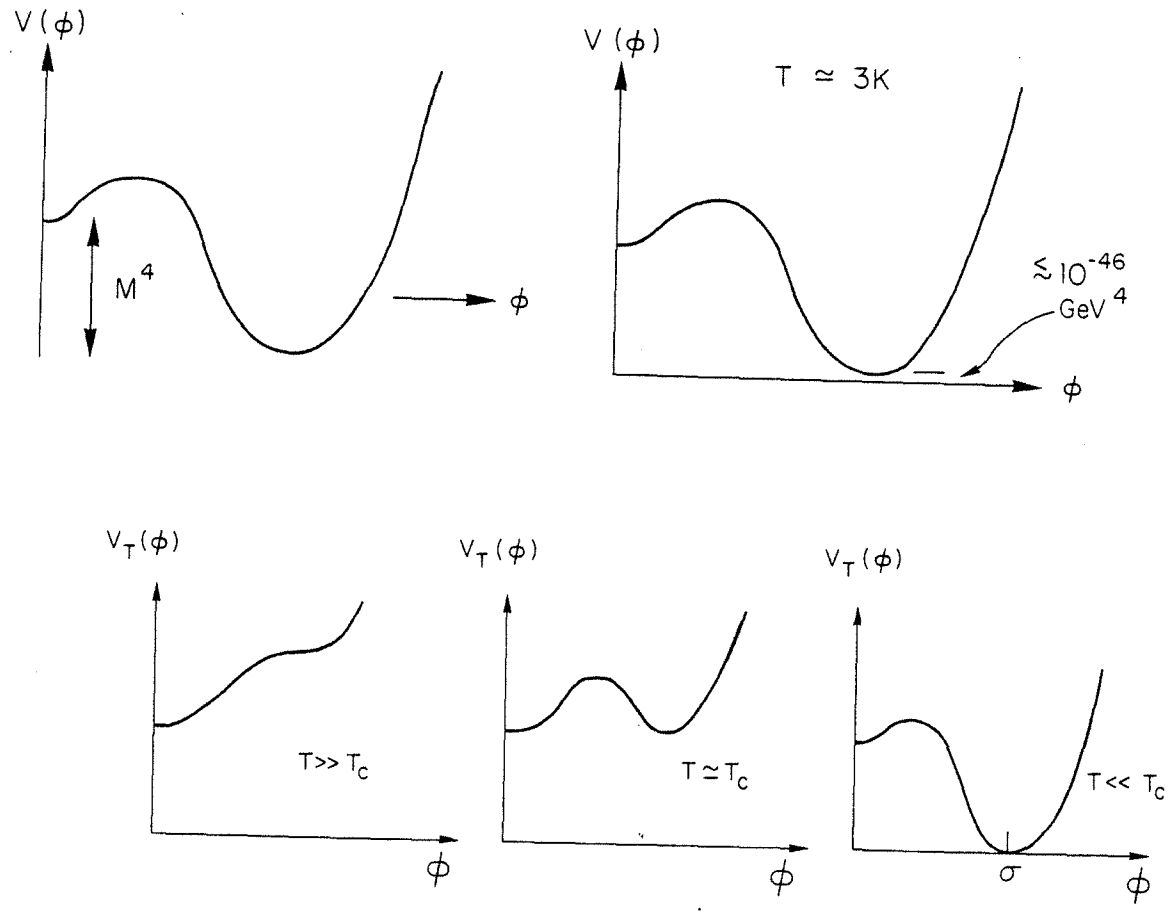


Fig. 4.2 The finite temperature effective potential V_T, for $T > T_c$; $T \approx T_c$; and $T \ll T_c$; here $\phi = \sigma$ is the SSB minimum of V.

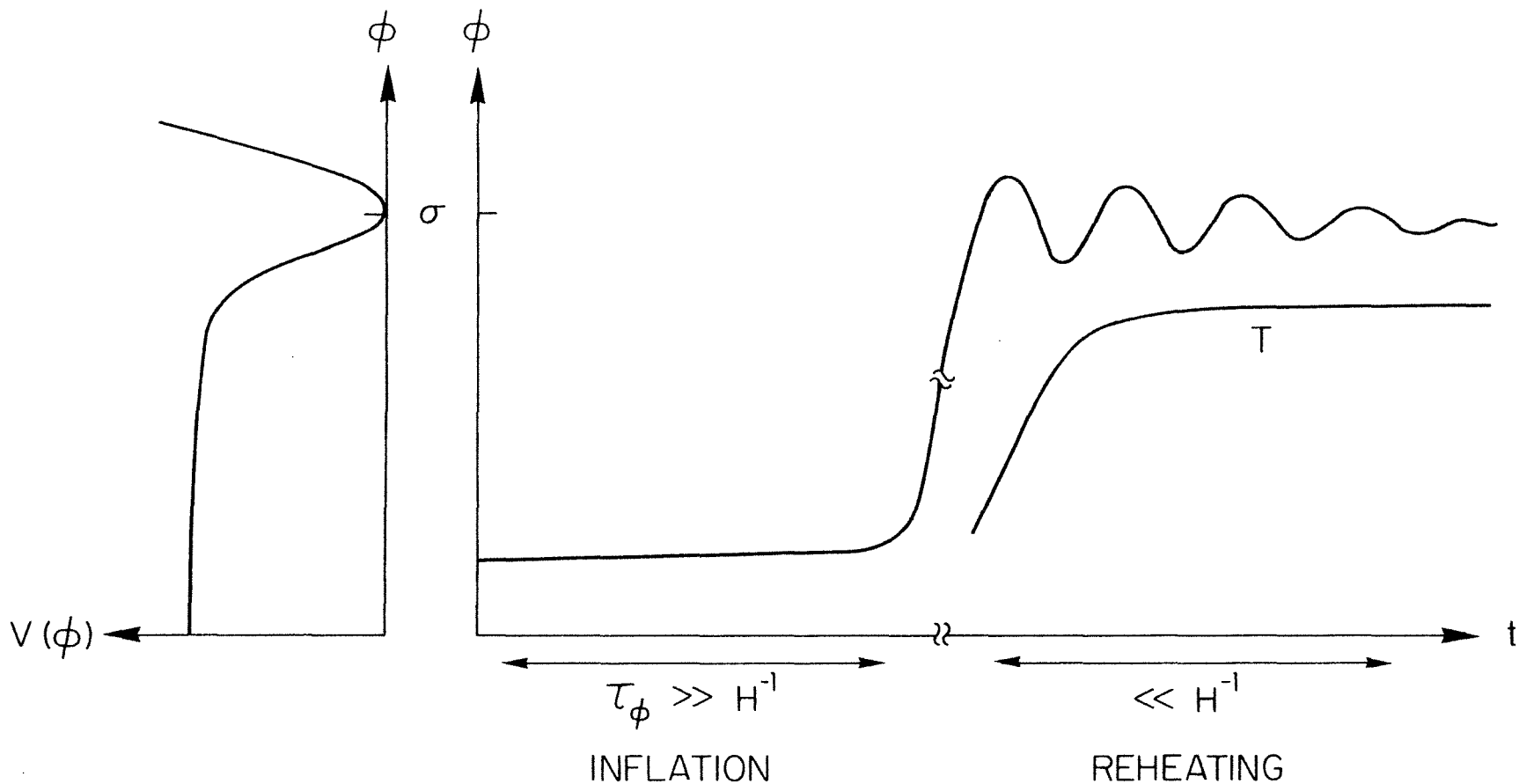


Fig. 4.3 The time evolution of ϕ. During the slow growth phase the time required for ϕ to change appreciably is $\gg H^{-1}$. As the potential steepens ϕ evolves more rapidly (timescale $\ll H^{-1}$), eventually oscillating about the SSB minimum. Particle creation damps the oscillations in a time $\approx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

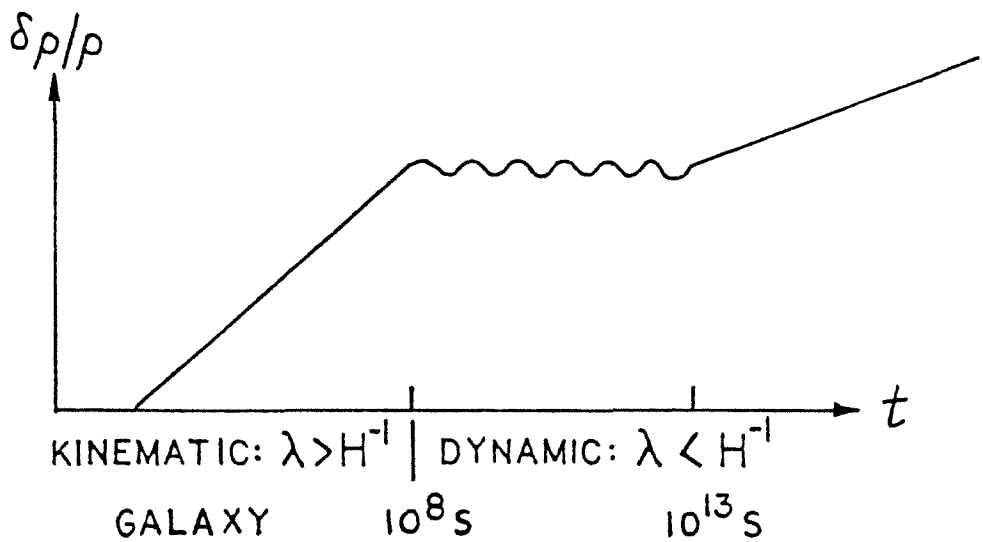


Fig. 4.4 Evolution of a galactic mass adiabatic density perturbation.

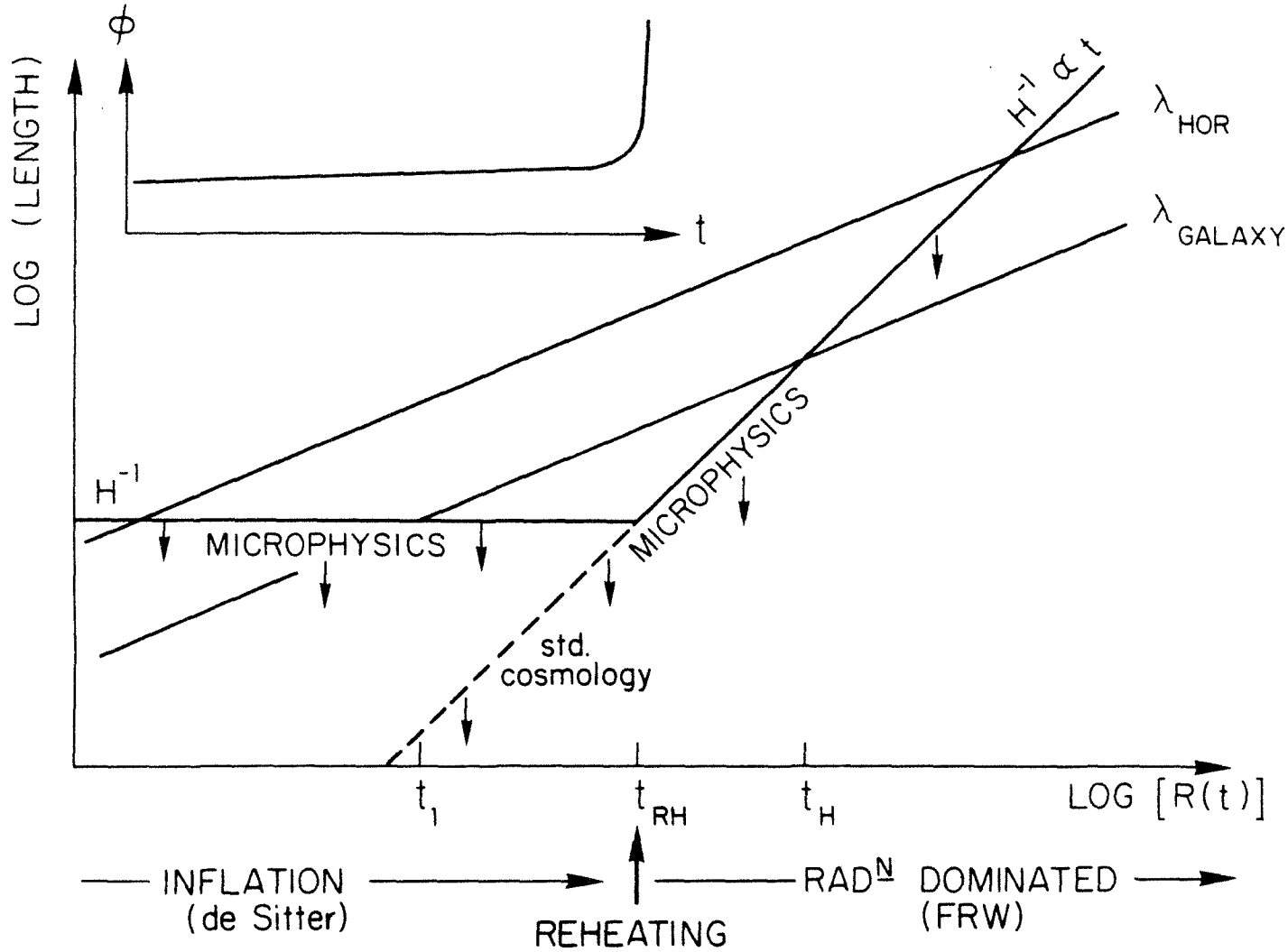


Fig. 4.5 The evolution of the 'physics horizon' (H^{-1}) and the physical sizes of perturbations on the scale of a galaxy (λ_G) and on the scale of the present observable Universe (λ_H). Reheating occurs at $t = t_{RH}$. **For reference the evolution of ϕ** is also shown. The broken line shows the evolution 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.

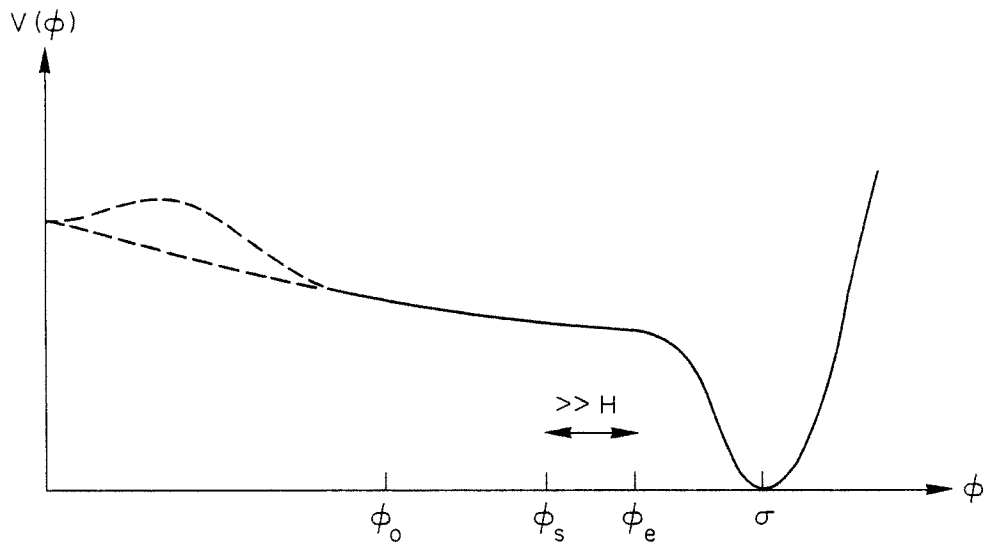


Fig. 4.6 The 'prescribed potential' for successful inflation.

LECTURE 5: FORMATION OF STRUCTURE IN THE UNIVERSE

Overview

On small scales the Universe today is very lumpy. For example, the average density in a galaxy ($\approx 10^{-24}$g cm^{-3}) is about 10^5 the average density of the Universe. The average density in a cluster of galaxies is about 100 times the average density in the Universe. Of course, on very large scales, say >> 100 Mpc, the Universe is smooth, as evidenced by the isotropy of the microwave background, number counts of radio sources, and the isotropy of the x-ray background.

The surface of last scattering for the 3K microwave background is the Universe at 200,000 years after 'the bang', when $T \approx 1/3$ eV and $R \approx 10^{-3}\ R_{today} = 10^{-3}$ (it is convenient to set $R_{today} \equiv 1$). Thus the μ-wave background is a fossil record of the Universe at that very early epoch. The isotropy of the μ-wave background, $\delta T/T \lesssim O(10^{-4})$ on angular scales ranging from 1' to 180° (see Fig. 5.1), implies that the Universe was smooth at that early epoch: $\delta\rho/\rho \ll 1$. There is a calculable relationship between $\delta T/T$ and $\delta\rho/\rho$ (which depends upon the nature of density perturbations present - type and spectrum), but typically

$$(\delta\rho/\rho)_{DEC} = (\delta T/T) \lesssim O(10^{-2}-10^{-3}), \qquad (5.1)$$

where # is O(10-100); for the detailed calculations I refer the interested reader to refs. 133-136.

So, the Universe was very smooth, and today it is very lumpy -- how did it get to here from there?? For the past decade, or so cosmologists have had a general picture of how this took place: small density inhomogeneities present initially grew via the Jeans, or gravitational instability, into the large inhomogeneities we observe today, i.e., galaxies, clusters of galaxies, etc. After decoupling, when the Universe is matter-dominated and baryons are free of the pressure support provided by photons, density inhomogeneities in the baryons and other components grow as

$$\delta\rho/\rho \propto \begin{cases} R & \delta\rho/\rho \lesssim 1 \\ \gtrsim R^3 & \delta\rho/\rho \gtrsim 1 \end{cases} \tag{5.2}$$

The isotropy of the μ-wave background allows for perturbations as large as 10^{-2} - 10^{-3} at decoupling, and the cosmic scale factor R(t) has grown by slightly more than a factor of 10^3 since decoupling, thus it is possible for the large perturbations we see today to have grown from small perturbations present at decoupling. This is the basic picture which is generally accepted as part of the 'standard cosmology'. [For a detailed discussion of structure formation in the Universe, see ref. 137.]

One would like to fill in the details, so that we can understand the formation of structure in the same detail that we do, say, primordial nucleosynthesis. The formation of structure (or galaxy formation as it is sometimes referred) began in earnest when the Universe became matter-dominated [$t_{eq} \approx 3\times10^{10}$sec $(\Omega h^2/\theta^3)^{-2}$; $T_{eq} \approx$

$6.8\,\mathrm{eV}\ \Omega h^2/\theta^3$]; that is the time when density perturbations in the matter component can begin to grow. In order to fill in the details of structure formation one needs the 'initial data' for that epoch; in this case, they include: the total amount of non-relativistic stuff in the Universe, quantified by Ω; the composition, i.e., fraction Ω_i of the various components [(i = baryons, relic WIMPs (weakly-interacting massive particles), cosmological constant, relic WIRPs (weakly-interacting relativistic particles)]; spectrum and type (i.e., 'adiabatic' or 'isothermal') of density perturbations initially present. Given these 'initial data' one can construct a detailed scenario (e.g., by numerical simulation), which can then be compared to the Universe we observe today.

I want to emphasize the importance of the 'initial data' for this problem; without such, it is clear that a detailed picture of structure formation cannot be put together. As I will discuss, it is in this regard that the Inner Space/Outer Space connection has been so very important in recent years. Events which we believe took place during the earliest moments of the history of the Universe and which we are just now beginning to understand, have given us important hints as to the 'initial data' for this problem. These hints include: $\Omega \approx 1.0$, adiabatic density perturbations with the Zel'dovich spectrum (from inflation); $\Omega_b \approx 0.014 \leftrightarrow 0.15$ (primordial nucleosynthesis); the possibility that the Universe is dominated by a massive, relic particle species (or WIMP) -- see Table 5.1 for a list of candidates; and other even more exotic possibilities -- topological strings, isothermal axion perturbations, a relic cosmological constant, and relativistic relic

particles, to mention a few. Because of these 'hints from the very early Universe', the problem has become much more focused, and at present two detailed scenarios exist the 'cold dark matter' picture and the 'hot dark matter' picture. Unfortunately, as I will discuss, neither appears to be completely satisfactory at present.

In the following subsections I will discuss: 'The Observed Universe' the final test of any scenario is how well it reproduces the observed Universe; the standard notation and definitions, associated with discussing the evolution of density inhomogeneities in an expanding Universe; the standard lore about the evolution of density perturbations; the hot and cold dark matter scenarios; and finally, I'll mention some of the pressing issues and current ideas.

The Observed Universe

The 'basic building blocks' of the Universe we see are galaxies; a typical galaxy has a mass of $O(10^{11}\ M_\odot)$. For reference, a solar mass ($M_\odot$) is 2×10^{33}g or about 10^{57} baryons. We observe galaxies with redshifts of greater than 1; the redshift z that a photon emitted at time t suffers by the present epoch is just:

$$(1 + z) = R_{today}/R(t) = R(t)^{-1}$$

The fact that we see galaxies with $z \gtrsim O(1)$ implies that galaxies were present and 'lit up' by the time the Universe was about 1/2 its present size. QSOs are the most distant objects we can see, and many QSOs with

redshifts in excess of 3 have been observed. [A typical QSO is 1-100 times as luminous as a galaxy, with a size of less than a light-month, or $\lesssim 10^{-5}$ that of a galaxy.] This fact implies that QSOs were present and 'lit up' when the Universe was 1/4 its present size.

To a first approximation galaxies are uniformly distributed in space; however, they do have a tendency to cluster and their clustering has been quantified by the galaxy-galaxy correlation function $\xi(r)$. The probability of finding a galaxy at a distance r from another galaxy is $(1+\xi(r))$ greater than if galaxies were just distributed randomly. The galaxy-galaxy correlation function is well-studied[137,138] and

$$\xi(r) \approx (r/5h^{-1}\mathrm{Mpc})^{-1.8} \quad .$$

Something like 10% of all galaxies are found in clusters of galaxies, a cluster being a bound and sometimes virialized system of O(100) galaxies. Particularly populous clusters are called rich clusters and many of these rich clusters are affectionately known by their Abell numbers, as the astronomer George Abell studied and classified many of these objects. There is some evidence that clusters cluster, and this has been quantified by the cluster-cluster correlation function[139]

$$\xi_{cc}(r) \approx (r/25h^{-1}\mathrm{Mpc})^{-1.8}$$

Interestingly enough, the cluster-cluster correlation function has the

same slope, and a larger amplitude. The larger amplitude is a fact which is not presently understood; a number of explanations have been suggested[140,141], including the possibility that the amplitude of the cluster-cluster correlation function has not yet been correctly determined. One very promising idea[140] is that this just reflects the fact that rich clusters are 'rare events' and that for Gaussian statistics rare events are more highly correlated than typical events (i.e. galaxies not in clusters).

There is some evidence for still larger-scale structure -- superclusters[142], objects which may contain many rich clusters and are not yet virialized; voids[143], large regions of space (perhaps as large as 100 Mpc) which contain far fewer than the expected number of galaxies; and filamentary structures[144], long chains of galaxies. As of yet, there are no unambiguous statistics to quantify these features of the Universe.

How much stuff is there in the Universe? This is usually quantified as the fraction of critical density Ω ($\equiv \rho/\rho_{crit}$), where

$$\rho_{crit} = 1.88h^2 \times 10^{-29} g\ cm^{-3}\ , \qquad (5.3)$$

$$= 1.05h^2 \times 10^4 eV\ cm^{-3}\ ,$$

$$= 0.810h^2 \times 10^{-46}\ GeV^4.$$

From primordial nucleosynthesis we know that the fraction contributed by baryons must be:

$$0.014 \lesssim \Omega_b \lesssim 0.15.$$

What can we say based on more direct observations of the amount of stuff in the Universe? The standard approach is not too different from that of the poor drunk faced with the task of locating his/her lost keys in a large, poorly-lit parking lot on a moonless night. He/she focuses his/her search in the vicinity of the only lamp post in the parking lot not because he/she thinks he/she lost his/her keys there, but rather because he/she realizes that this is the only place he/she could find them should they be there.

By determining the average mass per galaxy, one can convert the observed number density of galaxies into a mass density:

$$\langle\rho\rangle \approx \langle M_{GAL}\rangle\langle n_{GAL}\rangle \tag{5.4}$$

[To be more accurate, what astronomers actually do is to measure the mass to luminosity (or mass-to-light ratio, M/L, for short) for a typical galaxy and then multiply this by the observed luminosity density of the Universe to obtain the mass density; see ref. 145 for the dirty details of this procedure.]

The mass associated with a galaxy can be determined by a number of dynamical techniques, all of which basically involve Kepler's 3rd law in some way, or another. For a system with spherical symmetry:

$$GM = v^2 r \ , \tag{5.5}$$

where M is the mass interior to the orbit of an object with orbital velocity v and orbital radius r.

By studying the orbital motion of stars at the radius where the light has crapped out (the characteristic radius associated with the fall off in luminosity is called the Holmberg radius), one can measure the mass associated with the luminous material: M_{lum}. Converting this mass into an estimate for Ω one obtains:

$$\Omega_{LUM} \approx 0.01 \ ,$$

which is disappointingly distant from $\Omega = 1$, but consistent with baryons being the luminous material (thank God!).

Studies of the orbits of stars (in spiral galaxies) beyond the radius where the light has 'crapped out' have revealed an important and startling result[46] — their orbital velocities do not decrease (as they would if the luminous mass were the whole story, $v \propto r^{-1/2}$), but rather stay constant (this is the phenomenon referred to as 'flat rotation curves'). This, of course, indicates that mass continues to increase linearly with radius (or $\rho \propto r^{-2}$), whereas the light does not, indicating that the additional mass is dark. Flat rotation curves are the best evidence for the existence of dark matter. As of yet, there is no convincing evidence for a rotation curve that 'turns over' (which would indicate that the total mass in that galaxy has started to

converge). Thus the $\langle M_{GAL} \rangle$ obtained this way provides a lower limit to $\langle M_{GAL} \rangle$; this lower limit is <u>at least</u> 3–10 times the luminous mass, implying

$$\Omega_{HALO} \geq 0.03 - 0.10 \ .$$

The dark matter inferred from the rotation curves of spiral galaxies is often referred to as the 'halo material', as it is less condensed, seems to have a spherical distribution, and has a 'density run', $\rho \propto r^{-2}$, characteristic of a self-gravitating isothermal sphere of particles. Dynamical studies of the stars in our galaxy indicate that at our position (~ 8 kpc from the center) the disk component (i.e., the component distributed like the stars in the disk) dominates the halo component by about a factor of 30[147]. Incidently, these same studies[147] indicate that not all of the disk component is accounted for by stars, dust, gas, <u>etc.</u> Locally, about half the density, or ~ 6×10^{-24} g cm^{-3}, is unaccounted for. The 'dark matter in the disk' must be material which is capable of undergoing dissipation, since the formation of the disk clearly involved dissipation of gravitational energy.

[As of yet there is no undisputable evidence for dark matter in elliptical galaxies. They are more difficult to study since they lack 'test particles' far from the center of the galaxy. Attempts have been made to use x-ray measurements of the gravitational potential, and these seem to indicate the presence of dark matter.[148]]

By studying the dynamics of galaxies in bound, virialized systems (binary galaxy systems, small groups of galaxies, and galaxy clusters) one can try to infer $\langle M_{GAL} \rangle$. Again one is basically using Kepler's third law (in the guise of the virial theorem). The results indicate

$$\Omega_{BS} \approx 0.1 - 0.5 ,$$

have a higher degree of uncertainty, and are somewhat more ambiguous to interpret. For example, the highest values for Ω come from clusters, but only one galaxy in ten is found in a cluster. Most galaxies are found in binary systems or small groups. There is also the difficulty of identifying which galaxies are members of a given system, and determining whether or not a system has 'settled down' (i.e., is well-virialized) sufficiently so that the virial theorem is applicable.

There are many other techniques -- 'infall arguments': our motion toward the Virgo supercluster allows us to weigh the Virgo cluster, and our motion toward the Andromeda galaxy allows us to weigh our galaxy and Andromeda; the cosmic virial theorem[49], which relates the peculiar velocity field of the Universe[50] to Ω; all seem to point to a value of Ω in the range of $\approx 0.1 - 0.3$. Based upon the very non-trivial assumption that light is a good tracer of mass the observations seem to indicate that

$$\Omega_{OBS} \approx 0.2 \text{ '}\pm 0.1\text{'} ,$$

where '±0.1' is not meant as a formal error, but rather is meant to indicate the spread of the observations.

To summarize what we know about the amount of stuff in the Universe: (1) dark matter dominates $\approx$ by at least a factor of $3 \approx 10$, and is less $\approx$ condensed than the luminous component (which shows clear signs of having undergone dissipation); (2) the dark component <u>could</u> be baryonic $\approx$ until Ω_{DARK} is shown to be $\gtrsim 0.15$; (3) baryons cannot provide closure density as primordial nucleosynthesis restricts $\Omega_b \approx 0.014 \approx 0.15$; (4) no observation made yet indicates that Ω_{TOT} is close to 1.0! Our knowledge of Ω is summarized in Fig. 5.2.

Finally, there is the 3K μ-wave background radiation, the fossil record of the Universe at 200,000 yrs after the bang when $T \approx 1/3$ eV and $R \approx 10^{\approx 3}$. The angular scale viewed on the sky and the corresponding length scale are related by

$$\lambda/\mathrm{Mpc} = 1.8(\phi/1')h^{\approx 1} , \tag{5.6a}$$

$$M/10^{12}M_{\odot} \approx 0.85(\phi/1')^3 \ \Omega h^{\approx 1} . \tag{5.6b}$$

Two additional scales of importance are: the thickness of the surface of last scattering ($\approx 6'$) and the size of the horizon at decoupling ($\approx 1°$). On small angular scales ($\lesssim 1°$) the temperature fluctuations in the μ-wave background are dominated by those intrinsic to the photons. For adiabatic perturbations:

$$(\delta T/T) \approx (1/3)(\delta n_{\gamma}/n_{\gamma}) \approx (1/3)(n_b/n_b); \tag{5.7}$$

in WIMP $\approx$ dominated model Universes perturbations in the WIMPs (which

began to grow as soon as the Universe becomes matter-dominated) have grown by a factor of $\approx$ (10-30) by decoupling, so that

$$(\delta T/T) \approx (\delta\rho_x/\rho_x)_{dec}/30 \ . \tag{5.8}$$

On very small angular-scales ($\lesssim 6'$) the fluctuations in the μ-wave background are 'washed-out' due to the thickness of the last scattering surface.

On large-angular scales ($\gtrsim 1°$), the temperature fluctuations are primarily due to the Sachs-Wolfe effect (i.e., induced by the fluctuations in the gravitational potential): $\delta T/T \approx \delta\phi/\phi \approx 1/2\ (\delta\rho/\rho)_H$, and

$$(\delta T/T) \approx 0.5(\delta\rho/\rho)_{HOR} \ , \tag{5.9}$$

where $(\delta\rho/\rho)_{HOR}$ is the amplitude of the density perturbation on the scale corresponding to θ when that density perturbation crossed inside the horizon. Note that since angular scales $\gtrsim 1°$ correspond to linear scales which were outside the horizon at decoupling, fluctuations on these scales provide us with information about the 'unprocessed, primordial spectrum of density perturbations.' [In the case that the Universe is reionized again after decoupling, the scale of 'virgin perturbations' may be somewhat larger, see ref. 151.] The main point here is that the 3K background is a fossil record of the early Universe, and density perturbations in particular. For this reason it provides a very stringent test of scenarios of galaxy formation. The

interpretation of microwave background fluctuations is discussed in greater detail in refs. 133-136.

Notation and Definitions

It is convenient to discuss density perturbations in terms of the density contrast,

$$\delta \equiv \delta\rho(x)/\bar{\rho} \ , \tag{5.10}$$

and to expand the density contrast δ in a Fourier expansion:

$$\delta(x) = \Sigma\ e^{-ikx}\delta_k \ , \tag{5.11}$$

$$\equiv V/(2\pi)^3 \int \delta_k e^{-ikx} d^3x \ ,$$

$$\delta_k \equiv V^{-1} \int \delta(x) e^{ikx} d^3k \ .$$

Here $\bar{\rho}$ is the average density of the Universe, periodic boundary conditions have been imposed, and V is the volume of the fundamental cube. Strictly speaking, this expansion is only valid in spatially-flat models (k=0); however, at early times the effect of spatial curvature is small and can usually be neglected. [An analogous expansion exists for spatially-curved models; see ref. 152.] One other very important warning: $\delta(x)$ is not a gauge invariant quantity; in fact it is always possible to make $\delta(x) = 0$ by a suitable gauge transform: $x^\mu \to x^\mu + \varepsilon^\mu$, $g \to g + O(\varepsilon)$, $\varepsilon^\mu \approx O(kt)$. J. Bardeen[52] has developed an elegant but

non-trivial gauge invariant treatment of this problem. Rather than develop his formalism here, I will instead discuss the standard lore in the synchronous gauge ($g_{oo} = -1$, $g_{oi} = 0$). [For more details concerning the standard lore in the synchronous gauge, see, e.g., ref. 153.] I should emphasize that the gauge non-invariance of $\delta(x)$ only rears its ugly head for perturbations larger than the horizon.

A particular Fourier component is characterized by an amplitude and a wavenumber k. Let x and k be co-ordinate (or comoving) quantities, so that the physical distance

$$dx_{phys} = R(t)\ dx$$

and the physical wavenumber

$$k_{phys} = k/R(t)\ . \tag{5.12}$$

The wavelength of a perturbation is

$$\lambda \equiv 2\pi/k\ , \tag{5.13}$$

$$\lambda_{phys} = R(t)\ \lambda\ . \tag{5.14}$$

In the linear regime ($\delta < 1$), the physical size of a given density perturbation grows with the expansion. The comoving label (k or λ) for that perturbation is quite useful because it does not change with time -- i.e., the same physical perturbation is characterized by the same

comoving label, while the physical labels (k(phys) and λ(phys)) do change with the expansion. It is also useful (and conventional) to characterize a density perturbation by the (invariant) rest mass M in a sphere of radius $\lambda/2$:

$$M \cong 1.5\times10^{11} M_{\odot}(\Omega h^2)\lambda^3_{Mpc} , \tag{5.15}$$

where as before Ω is the fraction of critical density in NR particles (today) and R(t) has been normalized so that $R_{today} = 1$. A given Fourier component then is labeled by λ (= its physical size today, assuming that $\delta(x)$ were still < 1), k, its comoving wavenumber, and M, the mass contained within a sphere of radius $\lambda/2$. For reference, a typical galaxy mass ($10^{12} M_{\odot}$ including the dark matter) corresponds to

$$\lambda_{GAL} \cong 1.9 \text{ Mpc } (\Omega h^2)^{-1/3} . \tag{5.16}$$

Of course the physical size of a galaxy is much less, more like 100 kpc, because galactic-sized perturbations have 'gone non-linear' ($\delta > 1$) and have ceased to continue to grow with the expansion; 1.9 Mpc, then, is the size a perturbation containing $10^{12} M_{\odot}$ would have today had it not 'gone nonlinear' and pulled away from the general expansion of the Universe.

With these definitions we can proceed to discuss quantities like: $\delta\rho/\rho$, $\delta M/M$, and $\xi(r)$, in terms of $|\delta_k|^2$. First consider $\delta\rho/\rho$:

$$\delta\rho/\rho = \langle\delta(x)\delta(x)\rangle^{1/2}, \tag{5.17}$$

where < > indicates the average over all space, i.e., $\delta\rho/\rho$ is the RMS value of $\delta(x)$. A bit of Fourier algebra yields:

$$(\delta\rho/\rho)^2 = V/(2\pi)^3 \int_0^\infty |\delta_k|^2 \, d^3k \, . \tag{5.18}$$

Evidently, the contribution to $(\delta\rho/\rho)^2$ from a given scale (specified by k) is:

$$(\delta\rho/\rho)^2|_k = V/(2\pi)^3 \, k^3 |\delta_k|^2 \, . \tag{5.19}$$

Now consider $(\delta M/M)$, the RMS mass fluctuation on a given mass scale. This is what most people mean when they discuss the density contrast on a given mass scale. Mechanically, one would measure $(\delta M)_{RMS}$ by taking a volume V_w, which on average contains mass M, and placing it at all points throughout the space and computing the RMS mass fluctuation. Although it is simplest to choose a spherical volume V_w with a sharp surface, to avoid surfaces effects we must take care to smooth the surface. This is done by using a 'window function' W(r) (see below), which smoothly defines a volume V_w and mass $M = \bar{\rho} \, V_w$, where

$$V_w = 4\pi \int_0^\infty r^2 W(r) dr \, . \tag{5.20}$$

A particularly simple window function is a Gaussian:

$$W(r) = \exp(-r^2/2r_0^2) \, , \tag{5.21a}$$

$$V_w = (2\pi)^{3/2} \, r_0^3 \, , \tag{5.21b}$$

$$W(k) = (2\pi)^{3/2} r_o^3 \exp(-k^2 r_o^2/2)/V \ . \tag{5.21c}$$

The RMS mass fluctuation on the mass scale $M \equiv \bar{\rho} V_W$ is then:

$$(\delta M/M)^2 = \langle(\delta M/M)^2\rangle \equiv \langle(\int \delta(\vec{x}+\vec{r}) W(\vec{r}) d^3r)^2/V_W^2\rangle; \tag{5.22}$$

after some simple Fourier algebra it follows that

$$(\delta M/M)^2 = V^3 (2\pi)^{-3} V_W^{-2} \int |\delta_k|^2 |W(k)|^2 d^3k \ . \tag{5.23}$$

For the Gaussian window function:

$$(\delta M/M)^2 = V(2\pi)^{-3} \int |\delta_k|^2 e^{-k^2 r_o^2} d^3k. \tag{5.24}$$

That is, $(\delta M/M)^2$ is equal to the integral of $|\delta_k|^2 d^3k$ over all scales larger than r_o (all wavenumbers smaller than r_o^{-1}). If $|\delta_k|^2$ is given by a power law, $|\delta_k|^2 \propto k^n$ with $n > -3$, then the dominant contribution to $(\delta M/M)$ comes from the Fourier component on scale $k \simeq r_o^{-1}$ (as one would have hoped!), and

$$(\delta M/M)^2 \simeq V(2\pi)^{-3} k^3 |\delta_k|^2/(n+3)|_{k \simeq r_o^{-1}} . \tag{5.25}$$

When one refers to '$\delta\rho/\rho$ on a given mass scale' what one means more precisely is $\delta M/M$, and as we have shown here that is just the power on that scale:

$$(\delta\rho/\rho)_M \propto k^{3/2} |\delta_k| \ , \tag{5.26}$$

$$\propto k^{3/2+n/2},$$

$$\propto M^{-1/2-n/6},$$

where we have used the fact that the comoving label $M \propto \lambda^3 \propto k^{-3}$, and assumed (as is usually done) that $|\delta_k|^2 \propto k^n$.

Finally, consider the galaxy-galaxy correlation function $\xi(r)$. Let $n(x)$ be the number density of galaxies at position x, and $\bar{n}$ be the average number density of galaxies: $\bar{n} = \langle n(x) \rangle$. The joint probability δP_{12} of finding galaxies in volumes δV_1 and δV_2 (centered at positions 1 and 2) is:

$$\delta P_{12} = \bar{n}^2 \delta V_1 \delta V_2 (1+\delta n(x_1)/\bar{n})(1+\delta n(x_2)/\bar{n}), \tag{5.27}$$

where $\delta n/\bar{n} = (n(x) - \bar{n})/\bar{n}$. If positions 1 and 2 are separated by a distance r, say, then the average probability of finding a galaxy a distance r from another galaxy is:

$$\langle \delta P_{12} \rangle = \bar{n}^2 \delta V_1 \delta V_2 (1+\langle \delta n(x+r)\delta n(x) \rangle/\bar{n}^2),$$

$$\equiv \bar{n}^2 \delta V_1 \delta V_2 (1 + \xi(r)) , \tag{5.28}$$

where $\xi(r)$ is the excess probability (over a random distribution), and is known as the galaxy-galaxy correlation function. Note that $\xi(r)$ is just $\langle \delta n(x+r)\delta n(x) \rangle/\bar{n}^2$. In terms of its Fourier transform $f(k) \equiv V^{-1} \int (\delta n(x)/\bar{n}) e^{ikx} dx$,

$$\xi(r) = V(2\pi)^{-3} \int |f(k)|^2 e^{-ikr} d^3k \ . \tag{5.29}$$

If, as is often done, we make the assumption that 'light faithfully traces mass', then $n(x) = a\rho(x)$ and $\delta(x) = \delta n(x)/\bar{n}$, so that

$$\xi(r) = \langle(\delta\rho(x+r)/\bar{\rho})(\delta\rho(x)/\bar{\rho})\rangle, \tag{5.30}$$

$$\xi(r) = V/(2\pi)^3 \int |\delta_k|^2 e^{-ikr} \, d^3k \ . \tag{5.31}$$

Note, the power spectrum $|\delta_k|^2$ is the Fourier transform of the galaxy-galaxy correlation function. Also note that

$$\xi(r=0) = \langle(\delta\rho/\rho)^2\rangle \ . \tag{5.32}$$

Perturbations are usually characterized as being <u>adiabatic</u> (more properly, 'curvature perturbations') or <u>isothermal</u> (more properly, 'isocurvature perturbations'). An adiabatic perturbation is an honest-to-God, gauge-invariant, wrinkle in the space-time manifold. That is '$\delta\rho \neq 0$'. By the equivalence principle, all components participate in an adiabatic perturbation:

$$\delta n_b/n_b = \delta n_\gamma/n_\gamma = \delta n_x/n_x \ , \tag{5.33}$$

where 'x' represents any other species, e.g., a relic WIMP. Note for an adiabatic perturbation $\delta(n_x/n_\gamma) = \delta(n_B/n_\gamma) = 0$.

On the other hand, an isothermal perturbation is not an honest-to-God wrinkle in the space-time manifold. That is '$\delta\rho = 0$'. Rather it is a spatial variation in the equation of state, a pressure perturbation, if you will. The usual example of an isothermal perturbation is a spatial fluctuation in the baryon-to-photon ratio, i.e., $\delta(n_B/n_\gamma) = f(x)$. Such perturbations are referred to as isothermal, because at very early times when $\rho_{TOT} \approx \rho_\gamma$, $\delta\rho = 0$ is equivalent to $\delta T = 0$. Generalizing, isothermal perturbations at early times are characterized by:

$$\delta n_\gamma/n_\gamma = 3(\delta T/T) = 0$$

$$(\delta n_i/n_i) \neq 0 \text{ (for some species i).}$$

This corresponds to some species (or quantum number) being 'laid down' non-uniformly. Note, that since an isothermal perturbation is really a 'pressure perturbation', pressure gradients can and do, by moving matter about, change a pressure perturbation into a density perturbation on scales $\lesssim v_s t$, (v_s = sound speed). Thus on physical scales $\lesssim v_s t$, a pressure perturbation results in a density perturbation of the same amplitude. [For further discussion of adiabatic and isothermal perturbations, see, e.g., refs. 152-153.]

The Standard Lore

In this subsection I will attempt to briefly summarize the standard results (in the synchronous gauge). For a more detailed discussion of

the evolution of density perturbations I refer the reader to refs. 137, 153; for a gauge-invariant discussion I refer the reader to Bardeen[152].

A very important scale when discussing the evolution of density perturbations is H^{-1}, the Hubble radius, or as I like to refer to it, 'the physics horizon'. H^{-1} is the expansion timescale; thus, in a time $\approx H^{-1}$ all physical distances roughly double. Therefore it is the timescale for 'coherent microphysics' — physical processes operating on longer timescales will have their effects distorted by the expansion. Therefore, H^{-1} is the distance over <u>causal</u>, <u>coherent</u> <u>microphysics</u> operates — hence, the phrase 'physics horizon'. If $R \propto t^n$ $(n < 1)$, then H^{-1} is also the particle horizon. The evolution of a given Fourier component of $\delta\rho/\rho$ is naturally divided into two regimes: (1) $\lambda_{phys} \geq H^{-1}$ — when the perturbations said to be 'outside the horizon'; (2) $\lambda_{phys} \leq H^{-1}$ — when the perturbation is said to be 'inside the horizon'. Since $\lambda_{phys} \propto R(t)$, so long as $R(t)$ increases more slowly than t, a perturbation begins outside the horizon and then at some point enters the horizon (see Fig. 5.3). For reference, during the radiation-dominated epoch, a perturbation on the mass scale M enters the horizon at time $t \approx (M/M_{\odot})^{2/3}$ sec.

(1) $\lambda_{phys} \geq H^{-1}$: While outside the physics horizon, a perturbation cannot be affected by microphysics; its evolution is purely <u>kinematical</u> — it evolves like a ripple in space-time. In the synchronous gauge, <u>adiabatic</u> perturbations grow:

$$\delta\rho/\rho \propto t^n \qquad n = \begin{cases} 1 & \text{radiation-dominated} \\ 2/3 & \text{matter-dominated} \\ 0 & \text{curvature-dominated} \end{cases}$$

By curvature-dominated, I mean $k/R^2 \gg 8\pi G\rho/3$ so that $H^2 \approx k/R^2$. <u>Isothermal</u> perturbations do not grow — after all, they are not honest-to-God wrinkles in space-time. [I should be careful here, as this is a gauge-dependent statement. What is true and gauge-invariant, is that an isothermal perturbation eventually becomes a density perturbation of the same amplitude; see ref. 153.]

(2) $\lambda_{phys} \lesssim H^{-1}$: Once a perturbation enters the horizon microphysics can be important. First consider pressure forces. Pressure forces can support a density perturbation against collapse on scales $\lesssim v_s t$, where $v_s^2 \equiv dp/d\rho$ is the sound speed and t is the age of the Universe. [More, precisely t is the dynamical timescale, i.e., timescale for gravitational collapse: $t_{dyn} \approx (G\rho)^{-1/2}$. If $\rho = \rho_{TOT}$, then t_{dyn} is also the age of the Universe; if $\rho < \rho_{TOT}$, then $t_{dyn} \gtrsim t$.] Perturbations on scales $\lesssim \lambda_J \approx v_s t$ oscillate as acoustic waves; λ_J is known as the Jeans length and $M_J = \pi\lambda_J^3\rho_{NR}/6$ is the Jeans mass (ρ_{NR} = mass density in NR particles). Perturbations on scales $\gtrsim \lambda_J$ are unstable against collapse, and grow:

$$\delta\rho/\rho \propto t^n \begin{cases} n = 2/3 & \text{matter-dominated} \\ n = 0 & \text{radiation-dominated, or curvature-dominated} \end{cases}$$

Note that if the Universe is radiation-dominated, perturbations in a NR component even on scales $\lambda \gtrsim \lambda_J$ cannot grow. Simply put, the energy density of the radiation drives the expansion so rapidly that the growth of the perturbation cannot keep up, and it just 'hovers'.

The equation governing the growth of perturbations in species i (i = baryons, exotic particle relic, photons, etc.) is:

$$\ddot{\delta}_i + 2H\dot{\delta}_i + k^2 v_{si}^2 \delta_i / R^2 = 4\pi G\rho\Sigma(\rho_i \delta_i/\rho), \tag{5.34}$$

where δ_i is the kth Fourier component in species i, v_{si}^2 is the sound speed squared in component i, ρ is the total energy density of the Universe, and ρ_i is the energy density in species i. For a one species fluid, it is straightforward to verify the results quoted above.

A few words about Eqn. (5.34). In a non-expanding Universe (i.e., H=0), the Jeans instability is exponential on scales $\lambda \gtrsim \lambda_J$, with characteristic time $\tau \approx t_{dyn} \approx (G\rho_i)^{-1/2}$. The expansion of the Universe slows the growth and results in a power law instability. When the Universe is radiation-dominated (i.e., $R \sim t^{1/2}$, $H \sim 1/2t$, and $\rho_i/\rho \ll 1$), the solution to Eqn. (5.34) is $\delta_i = a + b\ln t$, so a perturbation with an initial velocity ($\dot{\delta}_i \neq 0$) can actually grow, albeit logarithmically.

Finally, one last bit of important microphysics. Perturbations in a collisionless component (e.g., neutrinos, axions, etc.) are subject to Landau damping, or 'freestreaming'. Until they become Jeans unstable, they can 'stream out' of overdense regions and into underdense regions, in the process smoothing out density perturbations. [Note, this effect has not been taken into account in Eqn. (5.34); in order to correctly take this effect into account one must integrate the collisionless Boltzmann Equation for the collisionless component.] The comoving freestreaming scale is easily calculated:

$$\lambda_{FS}(t) = \int_0^t v(t')dt'/R(t') \ , \tag{5.35}$$

where $v(t)$ is the velocity of the species in question. Most of the contribution to the integral comes at or just after the time the species goes NR; after this epoch $v \propto R^{-1}$ (for a collisionless species) and

$$\lambda_{FS} \approx (t_{NR}/R_{NR}) \begin{cases} 3 & t_{eq} \leq t_{NR} \\ \ln(t_{eq}/t_{NR}) & t_{eq} \geq t_{NR} \end{cases} \tag{5.36}$$

where t_{NR} and R_{NR} are the time and the scale factor when the species became NR, and t_{eq} is the time when the Universe became matter-dominated. [Note, in the case that the Universe is still radiation-dominated when the species goes NR, λ_{FS} continues to grow after $t \approx t_{NR}$ -- albeit logarithmically.] For reference:

$$t_{eq} \approx 3\times10^{10} \ \mathrm{sec}(\Omega h^2/\theta^3)^{-2} \ , \tag{5.37a}$$

$$T_{eq} \approx 6.8\mathrm{eV} \ (\Omega h^2/\theta^3) \ , \tag{5.37b}$$

$$T_{NR} \approx (m_X/3)(T_\gamma/T_X) \ , \tag{5.37c}$$

$$R_{NR} \approx 7\times10^{-7}\theta(T_X/T_\gamma)(\mathrm{keV}/m_X) \ , \tag{5.37d}$$

$$t_{NR} \approx 2.2\times10^{7}\mathrm{sec}(m_X/\mathrm{keV})^{-2}(T_X/T_\gamma)^2 \ , \tag{5.37e}$$

where T_γ is the photon temperature, T_X/T_γ is the ratio the temperature of species X to the photons, e.g., for neutrinos $T_X/T_\gamma \approx (4/11)^{1/3}$, and as usual θ is the present photon temperature in units of 2.7K.

Assuming that $t_{NR} \lesssim t_{eq}$, which is almost always the case, it follows that:

$$\lambda_{FS} \approx 1 \text{ Mpc}(m_x/\text{keV})^{-1}(T_x/T_\gamma)[\ln(t_{eq}/t_{NR}) + 1]. \tag{5.38}$$

For a neutrino species, $t_{NR} \approx t_{eq}$ and $(T_\nu/T_\gamma) \approx 0.71$, so that the freestreaming scale is

$$\lambda_\nu \approx 30 \text{ Mpc } (m_\nu/30\text{eV})^{-1}. \tag{5.39}$$

Although photons and baryons are certainly not collisionless (at least until after decoupling), there is a similar effect due to the fact that the mean free path of a photon is finite (and so the two are not a perfect fluid). In this case the operative word is not 'free streaming', but rather 'diffusion', and the effect is known as 'Silk damping'.[154] The scale associated with photon diffusion is

$$\lambda_S \approx 60\text{Mpc}/(1 + 25\Omega_b^{3/4}\Omega^{1/4}h^2). \tag{5.40}$$

Because of photon diffusion, or freestreaming of a collisionless species all initial perturbations on scales less than λ_{FS} (and for baryons λ_s) are strongly damped.

The Spectrum of Density Perturbations

We know how to specify the spectrum of density perturbations, but when does one specify it? From our discussion of the evolution of density perturbations, it seems very sensible to specify the amplitude of a density perturbation when it crosses the horizon, before any microphysical processing can occur. The amplitude at horizon crossing also has a Newtonian interpretation -- it is the perturbation in the gravitational potential. Note that by doing such, the amplitude on different scales is specified at <u>different</u> times. It is traditional to suppose that the spectrum is a featureless power law. Until recently this was merely an assumption since one had no fundamental understanding of the origin of density inhomogeneities. As discussed in Lecture 4, the inflationary paradigm has changed that situation. Write the amplitude at horizon crossing $(\delta\rho/\rho)_{HOR}$ as a power law,

$$(\delta\rho/\rho)_{HOR} \propto M^{-\alpha} ,$$

and remember that $(\delta\rho/\rho)$ means $k^{3/2}|\delta_k|$. If $|\delta_k|^2 \propto k^n$, then $\alpha = -1/2 - n/6$. The inflationary Universe scenario predicts $\alpha=0$ or $n=-3$.

What can we say more generally about α. The formation of galaxies requires that $(\delta\rho/\rho)_{HOR} \approx 10^{-4}$ on the scale of a galaxy, i.e., $10^{12}M_\odot$. The measured isotropy of the μ-wave background implies that $(\delta\rho/\rho)_{HOR}$ on the scale of the present horizon ($\approx 10^{22}M_\odot$) is less than 10^{-3} (to be very conservative). This means that α must be ≥ -0.1 (or else the spectrum must have a cutoff).

Consider perturbations on scales $\ll 10^{12} M_{\odot}$. If a perturbation crosses the horizon with amplitude greater than order unity, black hole formation is inevitable — regions of space-time will pinch off before pressure forces can respond to prevent black hole formation.[155] Black holes less massive than order 10^{15}g will have evaporated before the present epoch (via the Hawking process[156]), however holes more massive than 10^{15}g will still be with us today. If $(\delta\rho/\rho)_{HOR}$ were $\gtrsim O(1)$ on a scale $\gtrsim 10^{15}$g, there would be far too many black holes with us today. This implies that α must be $\lesssim 0.2$ (or that the spectrum must be cut off on the low mass end). The $\alpha=0$, constant-curvature spectrum is clearly singled out. It is the so-called Harrison-Zel'dovich[157] spectrum.

It is sometimes convenient to specify the spectrum at a fixed time, e.g., at $t \gtrsim t_{eq}$ when structure formation is proceeding:

$$(\delta\rho/\rho)_t \propto M^{-\gamma} \; ; \tag{5.41}$$

for scales which are still outside the horizon at time t, $\gamma = \alpha + 2/3$. [This fact is straightforward to show.] Note that when specified at a fixed time, the Zel'dovich spectrum (on scales larger than the horizon) corresponds to $\gamma = 2/3$ and $n = 1$.

The Processed Spectrum

Given the initial spectrum, i.e., $(\delta\rho/\rho)_{HOR}$, we can use our knowledge of the microphysics to calculate 'the processed spectrum.' A convenient time to specify the processed spectrum is at $t \approx t_{eq}$ when

structure formation really begins. Let's assume that the initial spectrum is the Zel'dovich spectrum. If not, the slopes of the various regions of the processed spectrum are obtained by changing those shown by α. Scales which cross the horizon before t_{eq}, i.e., those with

$$\lambda \lesssim \lambda_{eq} \approx 2ct_{eq}/R_{eq} \approx 13\mathrm{Mpc}(\theta^2/\Omega h^2), \tag{5.42}$$

grow only logarithmically from horizon-crossing until $t \approx t_{eq}$ when the Universe becomes matter-dominated (and perturbations start to grow as $R(t)$), by a factor of $O(20)$ for scales $\ll \lambda_{eq}$. On scales $\gtrsim \lambda_{eq}$, $(\delta\rho/\rho) \propto M^{-2/3}$, as discussed above. Now add free-streaming and we have the fully-processed spectrum, from which structure formation will proceed (see Fig. 5.4).

From the epoch of matter domination until decoupling ($R \approx 10^{-3}$, $T \approx 1/3$ eV, $t_{dec} \approx 7\times10^{12}\mathrm{sec}(\Omega h^2)^{-1/2}$) perturbations in the WIMPs can grow ($\delta\rho/\rho \propto R \propto t^{2/3}$); however, perturbations in the baryons cannot yet grow since they are still tightly coupled to the photons. After decoupling, the baryons are free of the pressure support provided by the photons, and quickly fall into the potential wells formed by the WIMPs. In a few expansion times the baryon perturbations catch up with the WIMP perturbations (so long as $\Omega_{WIMP} \gg \Omega_b$), and then perturbations in both components grow together (see Fig. 5.5).

Starting at $t \approx t_{eq}$, all scales grow together, $\delta\rho/\rho \propto t^{2/3}$, so that the shape of the spectrum remains the same (while the overall amplitude increases). When $(\delta\rho/\rho)$ becomes unity on a scale, structures of that

mass begin to form bound systems whose self-gravitational attraction dominates that of the rest of the Universe. These structures cease to participate in the general expansion of the Universe -- 'they pull away from the expansion'. For the Zel'dovich spectrum (or any spectrum which decreases with increasing mass scale), the first scale on which structures form is set by λ_D (see Fig. 5.4).

There are two limiting cases: (1) $\lambda_D \approx \lambda_{eq} \approx 13h^{-2}$ Mpc (for relic WIMPs which go NR at $t \approx t_{eq}$ -- of the candidates in Table 5.1, this only applies to light neutrinos). In this case the damping scale is much greater than a galactic mass (closer to the mass of a supercluster). This case is known as 'hot dark matter'. (2) $\lambda_D \lesssim 1$ Mpc (i.e., $\lambda_D \ll \lambda_{eq}$), which occurs for relic WIMPs which go NR long before the epoch of matter-domination. In this case the first bound structures to form are galactic mass (or smaller). There is, of course, an intermediate possibility, $\lambda_D \approx 1$ Mpc - 10 Mpc, referred to as 'warm dark matter', although I will not discuss that case here. In this case the first objects to form are necessarily of galactic mass. I'll briefly review structure formation in the <u>hot</u> and <u>cold</u> dark matter scenarios.

Two Stories: Hot and Cold Dark Matter

<u>The Hints</u> -- Before going on to discuss the hot and cold dark matter scenarios, let me once again emphasize the hints which the early Universe has provided us with, and which has led to these two detailed scenarios. The structure-formation problem is basically an initial data problem, and the study of the very early Universe has helped to focus

our thinking in this regard. First, the density perturbations. It has long been realized that primordial perturbations were necessary, however until very recently their type, spectrum and origin were largely a mystery. Inflation has provided for the very first time a scenario for the origin of perturbations and makes a very definite prediction -- adiabatic perturbations with the Harrison-Zel'dovich spectrum. [Inflation in an axion-dominated Universe also results in isothermal, axion perturbations, see Seckel and Turner, ref. 3.] Next, the composition of the Universe, Ω_i. We know that the dominant component is dark, and since Ω_b must be $\lesssim 0.15$, the dominant component must be non-baryonic if Ω is to be $\gtrsim 0.15$. Inflation predicts $\Omega = 1.0$ (more precisely, $k = 0$). Structure formation also favors a large value of Ω, as in a $\Omega < 1$ Universe perturbations have less time to grow -- perturbations cease to grow when the Universe becomes curvature-dominated, at a scale factor $R \approx \Omega$. [Because of this fact larger initial perturbations are needed (implying larger $\delta T/T$); the small isotropy of the microwave background implies that $\Omega h^{4/3} > 0.2$, see ref. 135.] There is no lack of candidate particles which could have sufficient relic abundance to provide $\Omega = 1$ (see Table 5.1). The hint, then, is that the Universe is dominated by relic WIMPs, with $\Omega_b \approx 0.1$ or so and $\Omega_{WIMP} \approx 0.9$. This too aids the growth of density perturbations; while perturbations in the baryons cannot begin to grow until after decoupling ($R \approx 10^{-3}$), perturbations in the dominant WIMP component can begin to grow as soon as the Universe becomes matter-dominated ($R_{eq} \approx 3\times10^{-5}h^{-2}$) -- providing for an additional factor of $30h^2$ growth in $\delta\rho/\rho$ over an $\Omega = 1$ baryon-dominated Universe. [In that regard a $\Omega = \Omega_b \approx 0.1$ Universe is in sad shape; perturbations can grow only from $R \approx 10^{-3}$ to R

$\Omega_b \simeq 10^{-1}$, for a total growth factor of only 100! In fact, the small-scale microwave anisotropy measurement of Uson and Wilkinson rules out a baryon-dominated Universe with adiabatic perturbations; see, refs. 134-135.] These hints have led to two detailed scenarios -- hot and cold dark matter.

Hot dark matter - The first structures to form are of supercluster size ($10^{15} M_\odot$ or so), and they do so rather recently ($z \lesssim 3$). Zel'dovich has argued rather convincingly that the collapse of the objects should be very non-spherical, and very nearly 1-dimensional (like a pancake or 'blini'). [For a review of the pancake scenario, see, ref. 158.] Once the pancake forms and goes non-linear in one dimension, the baryons within it can collide with each other and dissipate their gravitational energy. Thereby, the baryons in the pancake can fragment and condense into smaller (say galaxy-sized) objects. The neutrinos, being so weakly interacting, do not collide with each other or the baryons, cannot dissipate their gravitational energy and therefore cannot collapse into more tightly-bound objects. Thus they should remain as a halo. Some slow-moving neutrinos may subsequently be captured by the baryon-dominated galaxies. Structure formation in the hot dark matter case is schematically illustrated in Fig. 5.6. In a phrase, the structure in a hot dark matter Universe forms from 'the top down'.

Several groups[159,160] have performed numerical simulations of the neutrino-dominated Universe. The simulations reproduce nicely the voids and filamentary structures which seem to exist in the Universe (see Fig. 5.8). However, in order to reproduce the observed galaxy-galaxy

correlation function, the epoch of pancaking must be made to occur yesterday -- that is at a very low redshift ($z \lesssim 1$). This fact is difficult to reconcile with the many galaxies with redshifts greater than 1 and QSO's with redshifts greater than 3 (the current record holder has $z \approx 3.8$). It should be mentioned that these simulations only simulate the behaviour of the neutrinos, whereas, the galaxy-galaxy correlation function clearly is determined by where the baryons are. The small-scale microwave anisotropy predicted in the neutrino scenario is marginally consistent with the upper limit[112] of Uson and Wilkinson on 4.5': $\delta T/T \lesssim 3\times10^{-5}$, so long as Ω is close to 1. To summarize, hot dark matter is down, but not quite out (yet). [For more details about how the baryons cool and fragment, and the hot dark matter scenario in general, see refs. 158-162.]

<u>Cold dark matter</u> -- Since the damping scale is less than a galactic mass in this case, the first structures to form should be galaxies. This should occur at a redshift of 10-20. Once galactic-sized objects have formed and have virialized, some of the baryons will collide and dissipate their gravitational energy, thereby settling into more compact objects at the center of the WIMP halo. These objects include stars, star clusters, and galactic disks.

Once galaxies form, they will tend to cluster together, forming larger aggregates such as small groups of galaxies, clusters of galaxies, and eventually superclusters. This process of 'hierarchical clustering' is helped along by the initial density perturbations which exist on these larger scales. The cold dark matter or 'bottom up' scenario is shown schematically in Fig. 5.7.

Numerical simulations of cold dark matter have been performed (see Fig. 5.8). Here there is no problem with getting galaxies to form early enough. The simulators[163,164] can get their model Universes to match the observed Universe in very many respects -- the galaxy-galaxy correlation function, masses and densities of galaxies, and many other details, but at a price: they find that they must require $\Omega h \approx 0.2$ to do so. The small-scale microwave anisotropy predicted in the cold dark matter scenario is a factor of 4 or so less than the observed upper limit at present. [For more details of the cold dark matter scenario see refs. 163-165.]

The Ω-Problem

This brings us to a pressing and very significant problem: the fact that $\Omega_{OBS} \approx 0.2 \pm$ '0.1', while theoretical prejudice would have us believe that $\Omega \approx 1.0$. As mentioned earlier, Ω_{OBS} is determined assuming that 'light' is a good tracer of the mass. A component of matter which is smoothly distributed on the scales which are being probed would not be detected. Ω_{OBS} only measures the material which is clumped on the scale being probed. In the cold dark matter scenario the Ω-problem is particularly acute as on the scales of galactic haloes and larger, baryons (light) should be a good tracer of the mass, as there is no mechanism to separate baryons and WIMPs. In the hot dark matter scenario the situation is less clear. Because of their large damping length, neutrinos are initially smooth on scales up to 30 Mpc or so, and Ω has yet to be reliably probed on scales this large.

The Ω-problem has received a great deal of attention recently and a number of solutions have been proposed. They all basically involve the same idea: the existence of a smooth component $\Omega_{SMOOTH} = 1-\Omega_{OBS} \approx 0.8\pm'0.1'$ which would provide the additional mass density required to bring Ω to 1. This component, of course, would go undetected because it is more smoothly distributed than the observed component. [It could, of course be detected by measuring the deceleration parameter q_o.] Suggestions for the smooth component include: (1) A relic cosmological term[166,167] (i.e., $\Lambda = 3\Omega_{SMOOTH}H_o^2$), which by definition is of course absolutely smooth. The possible origin of such a term, at present, has not even the slightest hint of an explanation. (2) Relativistic (or very fast-moving) particles,[166,168] produced by the recent (redshift 3-10) decay of an unstable WIMP (e.g., a 100 eV neutrino which decays to a light neutrino and a massless scalar). Very fast-moving particles cannot, by virtue of their high speeds, cluster. This scenario has received a great deal of attention lately. (3) A network of strings, or very fast-moving strings which too cannot cluster.[169] (4) A more smoothly-distributed component of galaxies which are either too faint to be seen or never lit up.[170]

The idea of 'failed galaxies' is an intriguing one. In this scenario, the very overdense regions ('3σ peaks') which collapse first would correspond to the galaxies we see. The more typical regions ('1 σ peaks' in the density distribution) would be the 'failed galaxies'. It is a well-known property of gaussian statistics that the rare events ('3 σ peaks' here) are more highly correlated. Put another way, the common events ('1 σ peaks') are less correlated, i.e., more smoothly

distributed. If this idea is correct, then today we are only seeing the tips of the icebergs so to speak. What is presently lacking in this scenario is a plausible mechanism for 'biasing' galaxy formation (i.e., inhibiting the 1 σ peaks from lighting up).

Epilogue

The hints provided by the very early Universe have helped to focus the efforts of those studying galaxy formation. We have at present two rather detailed stories -- hot and cold dark matter. Neither story however provides a totally satisfactory picture. We have at least one major problem -- the Ω problem. I have discussed some possible modifications of the two stories which might help to resolve the Ω problem.

Lest we become over confident, we should realize that the eventual sorting out of the details of structure formation may involve a bold departure from the two stories I discussed. For example, I did not discuss the role that cosmic strings may play. It may be that structure formation was initiated not by adiabatic density perturbations, but rather by strings and loops and the isothermal perturbations induced by them. I refer the interested reader to ref. 171, for further discussion of strings and galaxy formation.

The discussion of structure formation usually focuses on the role of gravitational forces, astrophysical fireworks (energy produced by the galaxies themselves by nuclear and other processes) are usually ignored.

During the very early stages of structure formation (before any structures form) this is probably a very good approximation, but once a few galaxies light up, the energy released by astrophysical processes within them may play an important role. This point has been particularly emphasized by Ostriker and his collaborators.[172] All of the processes suggested for biasing galaxy formation involve astrophysical processes.

While I have focused on the role of 'primordial perturbations' (i.e., those produced in the very early Universe), it could be that the relevant perturbations arise due to physical processes which occur rather late, e.g., processes which result in black hole formation (say masses 10^{-6}-$10^{6}M_{\odot}$), with these 'small' black holes then playing the role of seeds for structure formation. Carr[174] has particularly emphasized this possibility.

Although much progress has been made toward understanding structure formation, there may yet be some very interesting surprises in store for us! More than likely observational/experimental data will play an important role. For example, the list of candidate WIMPs may be whittled down by searches for SUSY partners at the CERN $Sp\bar{p}S$ and TeV 1 (or one of the candidate WIMPs may be detected!), the situation with regard to the mass of the electron neutrino may become clearer, the 'not so invisible' halo of axions may be detected by the technique advocated by Sikivie[173], further refinement of the measurement of the small-scale anisotropy of the microwave background may provide a signal(!) or even tighter constraints, deeper galaxy surveys may lead to the discovery (or

lack of discovery) of galaxies (or QSOs) at very high redshifts ($z > 4$), which would have important implications for whether things were 'top down' or 'bottom up' (or neither), and other results not even dreamed of by simple-minded theorists!

Table 5.1 WIMP Candidates for the Dark Matter

Particle	Mass	Place of origin	Abundance*
Invisible Axion	10^{-5}eV	10^{-30}sec, 10^{12}GeV	10^{9}cm^{-3}
Neutrino	30eV	1 sec, 1 MeV	100cm^{-3}
Photino/Gravitino/ Mirror Neutrino	keV	10^{-4}sec, 100MeV	10cm^{-3}
Photino/Sneutrino/Axino Gravitino/Shadow Matter/ Heavy Neutrino	GeV	10^{-3} sec, 10 MeV	10^{-5}cm^{-3}
Superheavy Magnetic Monopoles	10^{16}GeV	10^{-34}sec, 10^{14}GeV	10^{-21}cm^{-3}
Pyrgons/Maximons/Newtorites	$\gtrsim 10^{19}$GeV	10^{-43}sec, 10^{19}GeV	$\lesssim 10^{-24}$cm^{-3}
Quark Nuggets	$\approx 10^{15}$gram	10^{-5}sec, 300MeV	10^{-44}cm^{-3}
Primordial Black Holes	$\gtrsim 10^{15}$gram	$\gtrsim 10^{-12}$sec, $\lesssim 10^{3}$GeV	$\lesssim 10^{-44}$cm^{-3}

*Abundance required for closure density: $n_{WIMP} \approx 1.05h^2 \times 10^{-5}$ cm^{-3}/m_{WIMP} (GeV)

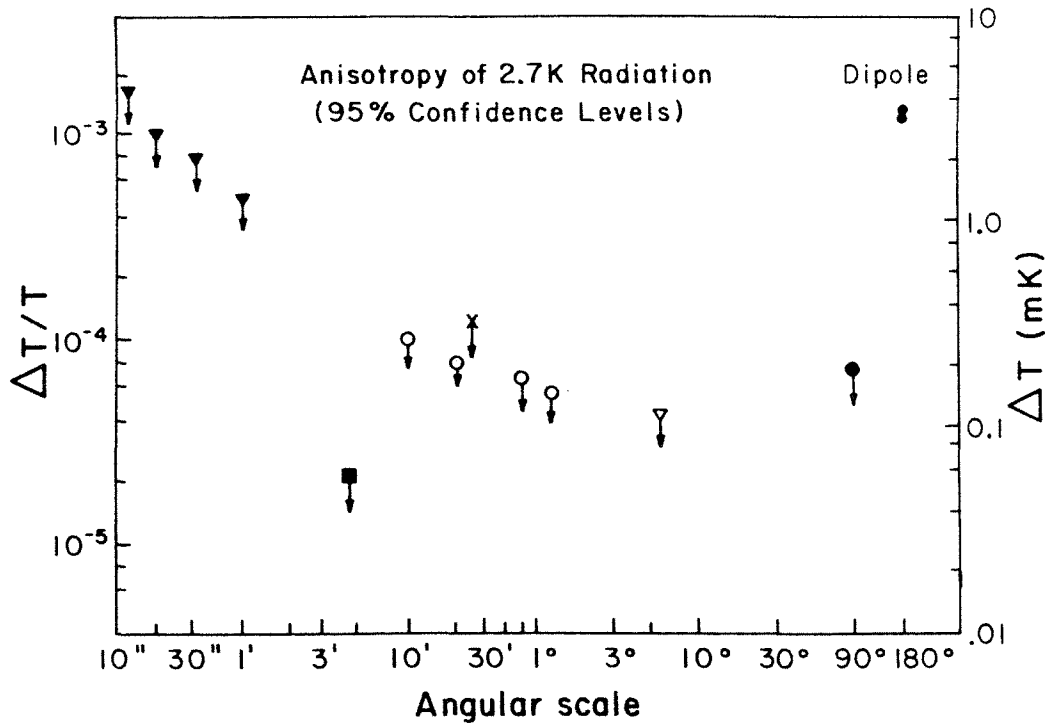


Fig. 5.1 Summary of the measurements of the isotropy of the microwave background (from ref. 113).

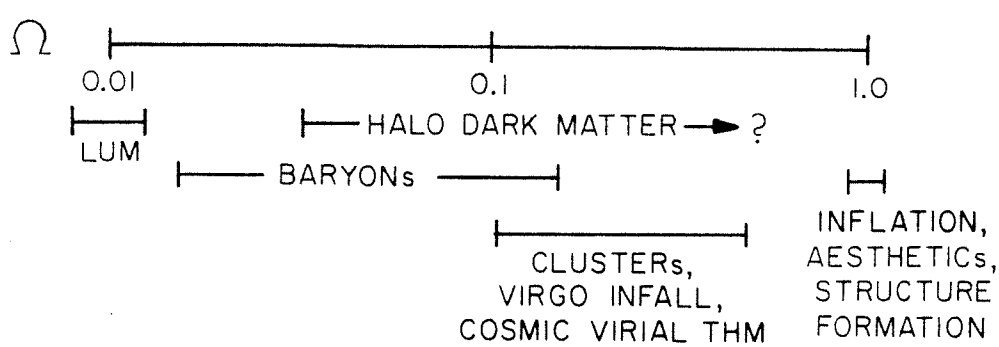


Fig. 5.2 Summary of our knowledge of Ω.

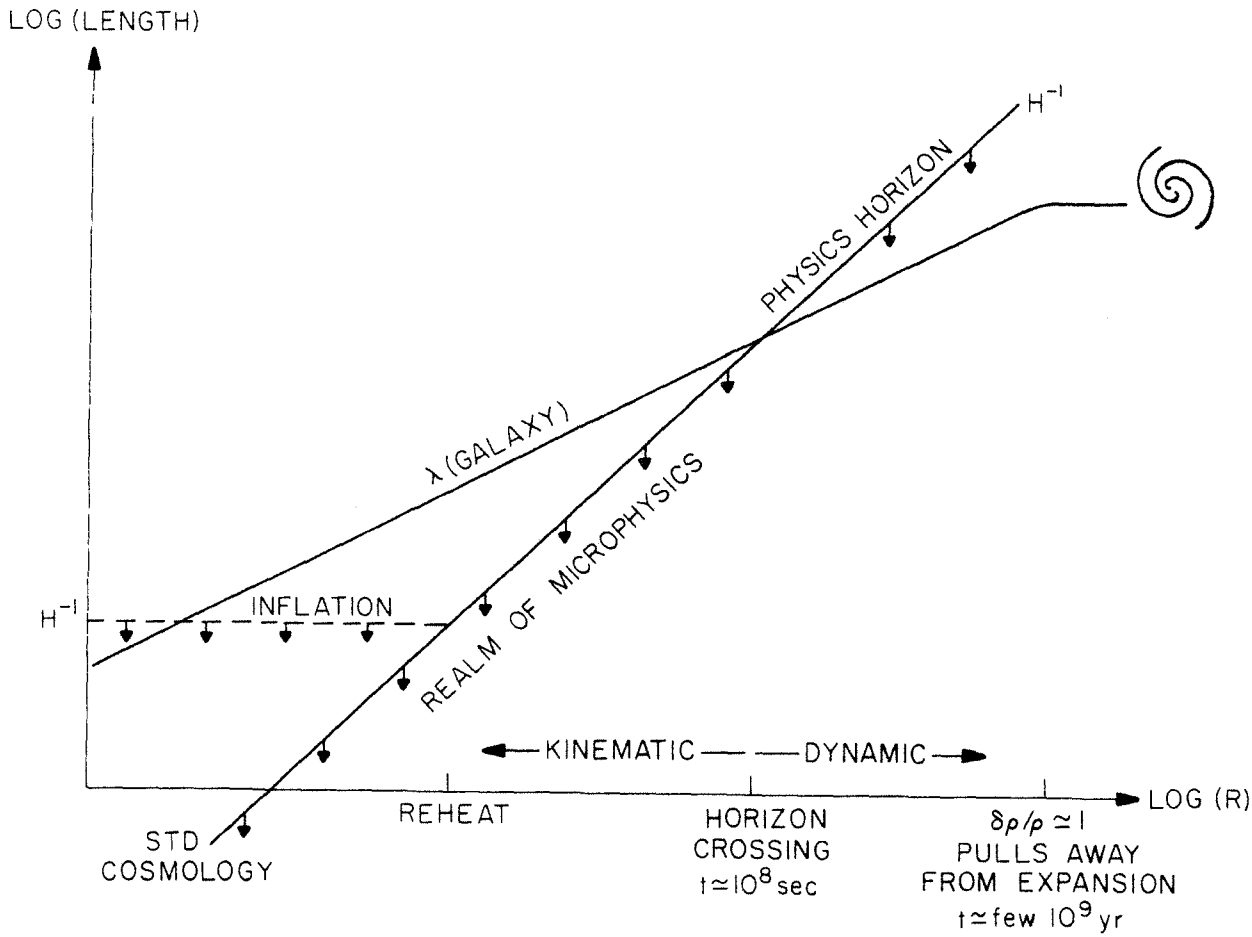


Fig. 5.3a Evolution of the physical size of a galactic-scale perturbation and of the 'physics horizon' (= H^{-1}) as a function of the scale factor R(t). Causal, coherent microphysics only operates on physical scales $\lesssim H^{-1}$. The kinematic ($\lambda \gtrsim H^{-1}$) and dynamical ($\lambda \lesssim H^{-1}$) regimes are indicated. Once $\delta\rho/\rho$ becomes of O(1) the galactic perturbation pulls away from the expansion, and its size remains constant.

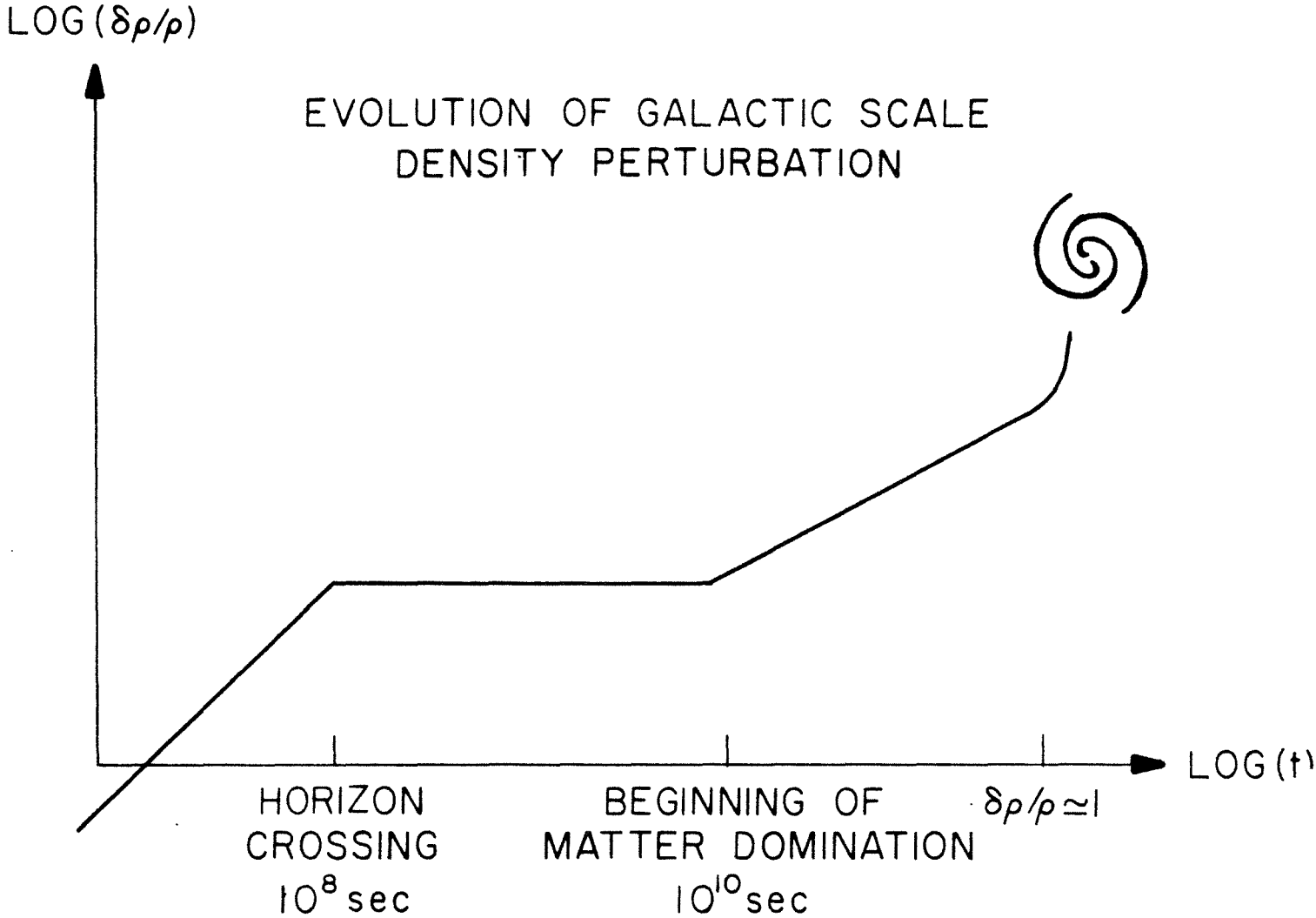


Fig. 5.3b Evolution of the amplitude of a galactic-scale perturbation ~~in~~ amplitude $\delta\rho/\rho \propto k^{3/2}|\delta_k|$.

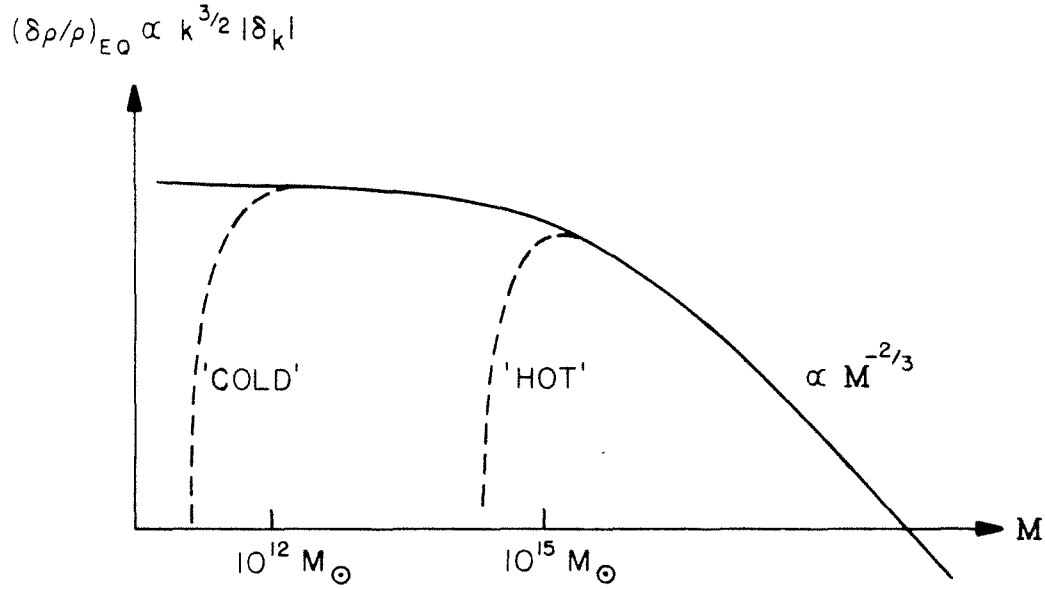


Fig. 5.4 The 'Processed Spectra' at $t \approx t_{eq}$.

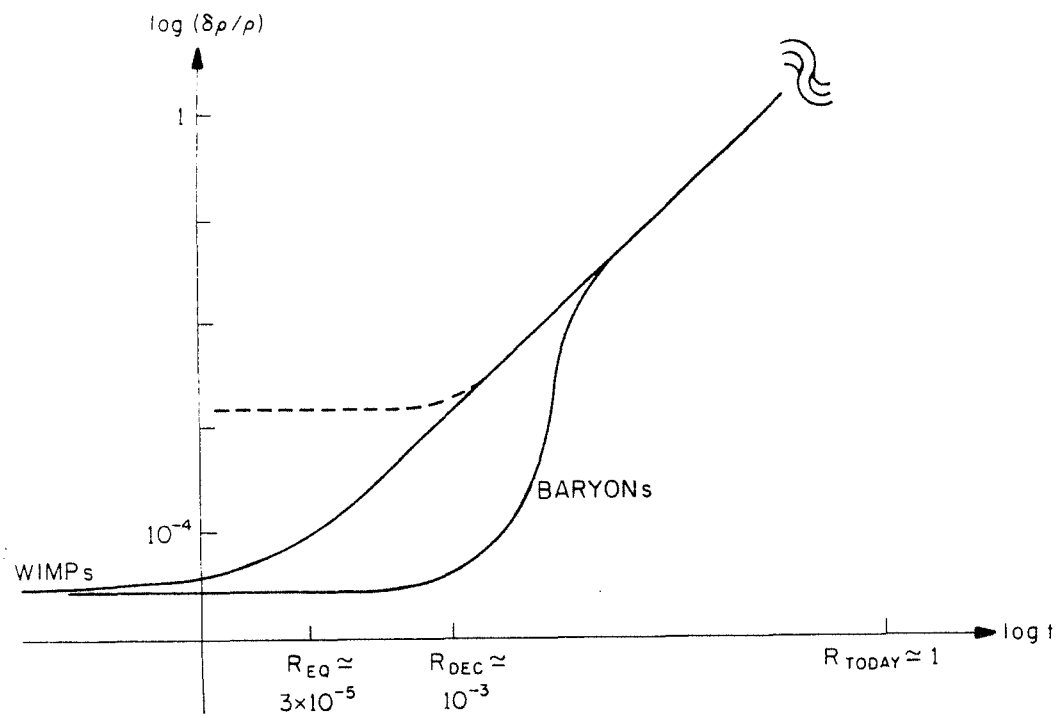


Fig. 5.5 The evolution of $(\delta\rho/\rho)$ in the Baryon and WIMP components. Note that in a baryon-dominated Universe $(\delta\rho/\rho)$ cannot begin to grow until decoupling and must necessarily begin with a larger amplitude.

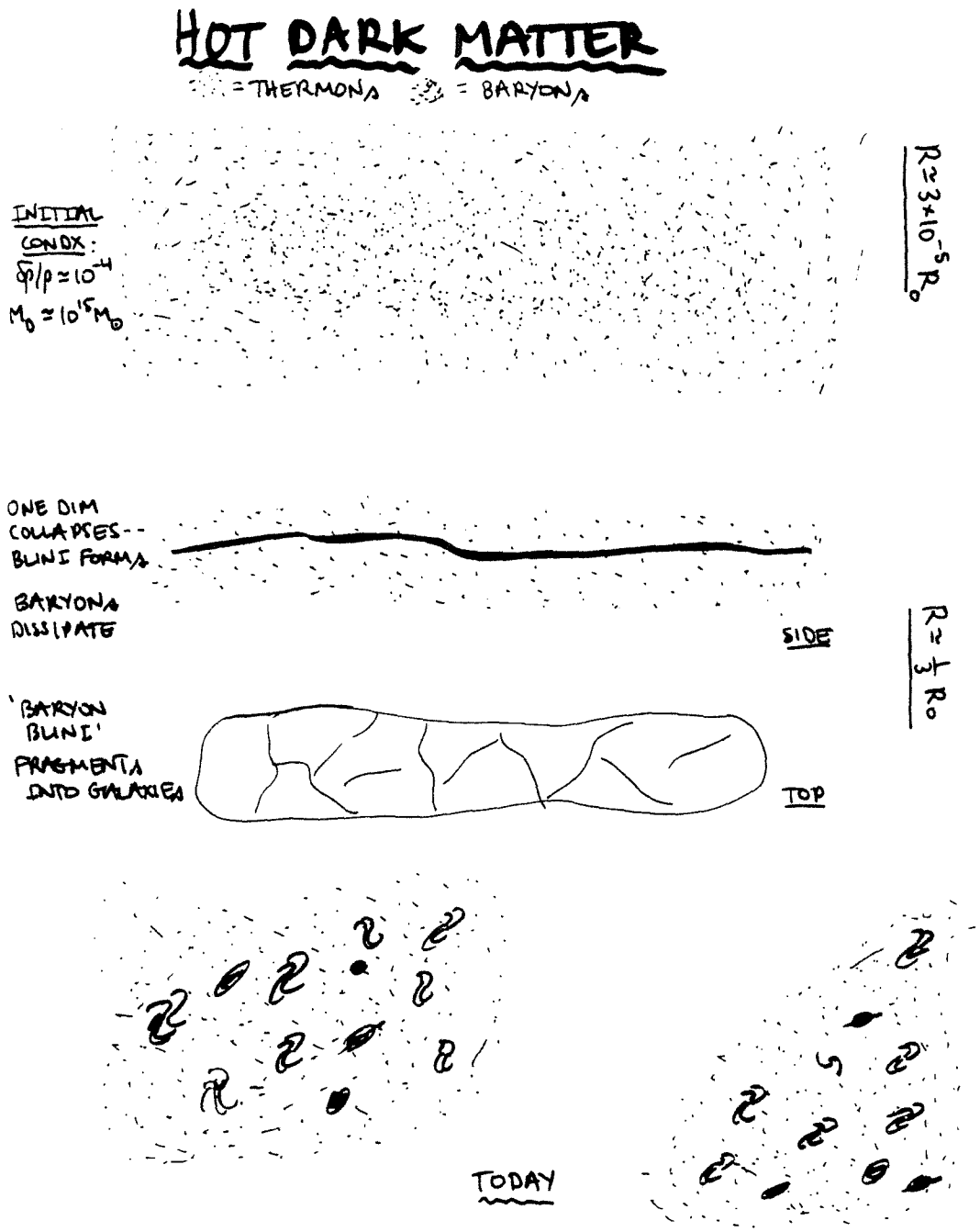


Fig. 5.6 Schematic illustration of structure formation in the 'hot dark matter' scenario.

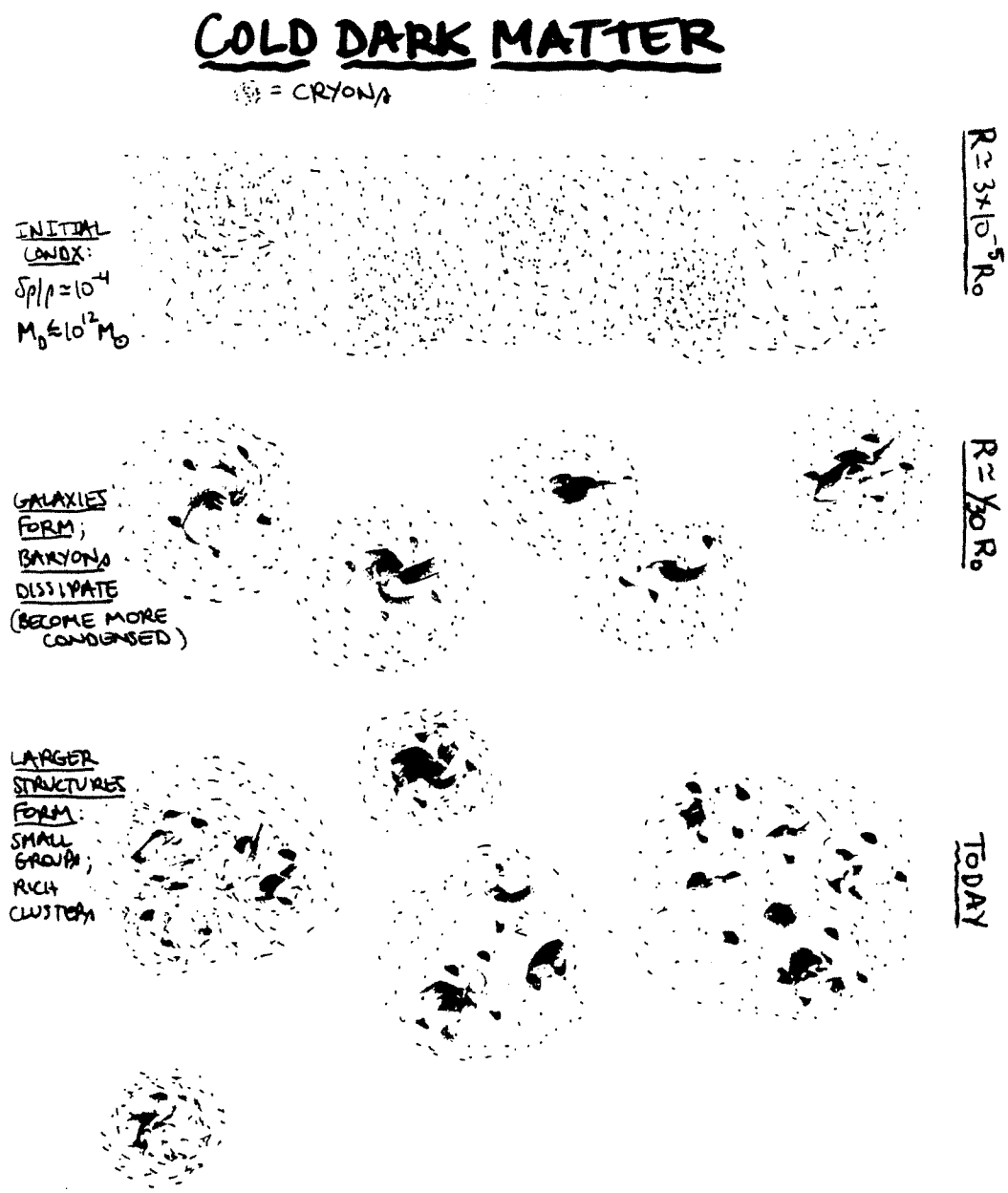


Fig. 5.7 Schematic illustration of structure formation in the 'cold dark matter' scenario.

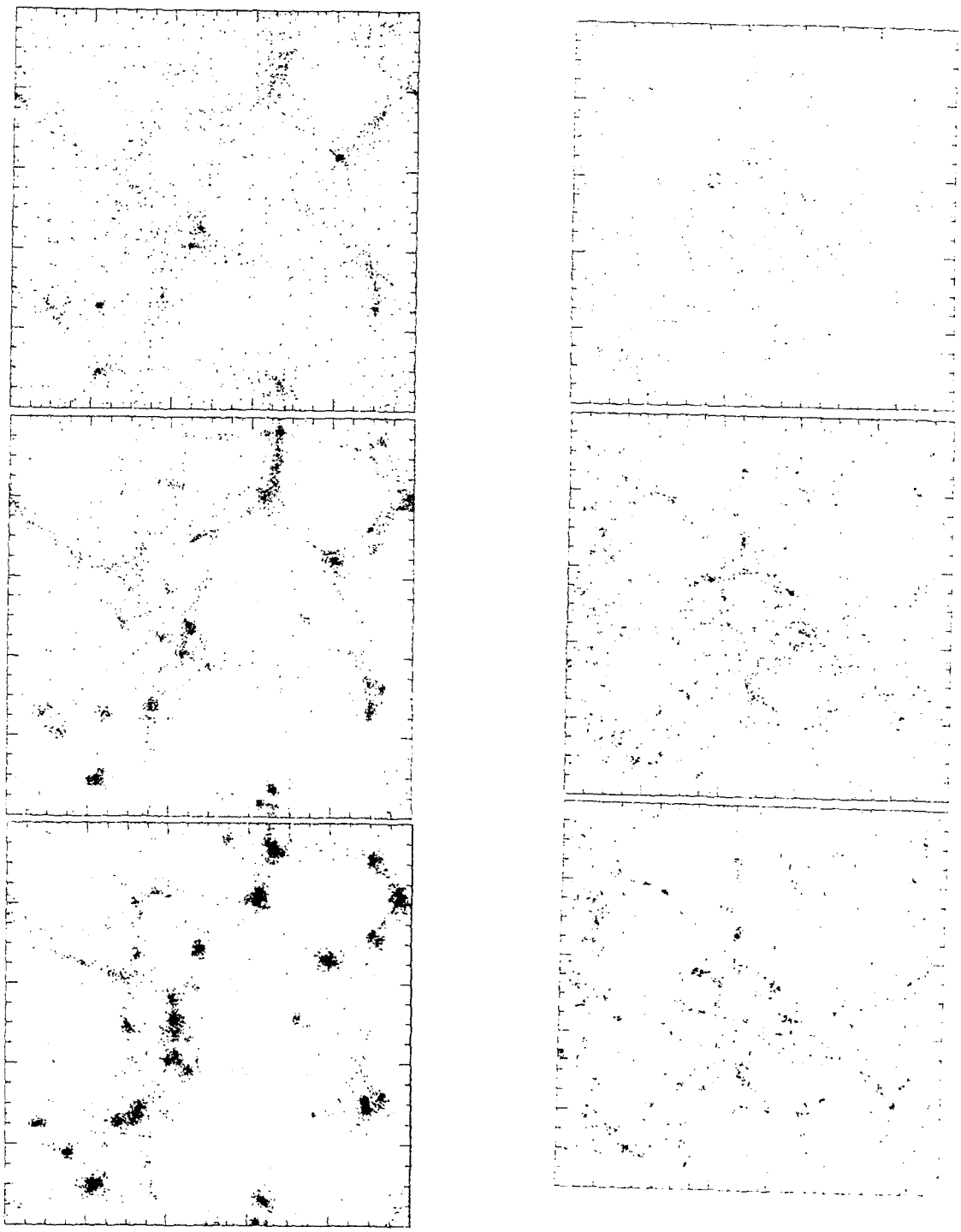


Fig. 5.8 Projected distribution of mass points (in comoving coordinates) for the numerical simulations of ref. 159; time evolves downward. For 'hot dark matter' (left), the panels represent the present Universe (z=0), in models where the onset of galaxy formation occurred at z_{GF} = .4, 1.1, 2.5 . For 'cold dark matter' (right), the last panel corresponds to the present Universe in a model with $\Omega h \approx 0.2$ ~~~ the model which gives the best fit to the observed Universe.

LECTURE 6 - AXIONS, ASTROPHYSICS, AND COSMOLOGY

The early Universe and contemporary astrophysical sites have proven to be invaluable 'non-traditional laboratories' for studying physics in regimes (i.e., energies and coupling strengths) that are not accessible in terrestrial experiments. In Lecture 1 I discussed the very stringent constraints on neutrino properties which follow from astrophysical and cosmological considerations. In this Lecture I will focus on an even more interesting case study -- the axion.

Introduction

QCD is almost universally accepted as 'the theory of the strong interactions.' It is a wonderful and remarkable theory, except for one very nagging flaw -- the so-called strong CP problem. The perturbative part of the theory automatically conserves C, P, T, and B. However, non-perturbative effects (due to color instantons[175]) lead to an additional term in the Lagrangian

$$\mathcal{L}_{QCD} = \mathcal{L}_{pert} + (\theta g^2/32\pi^2) G^a_{\mu\nu} \tilde{G}^{a\mu\nu}, \tag{6.1}$$

which does not respect CP or T invariance; here $G^a_{\mu\nu}$ is the color field strength tensor, $\tilde{G}^a_{\mu\nu}$ its dual, and θ is an arbitrary, unspecified parameter in the theory. This additional term is a perfect 4-divergence and thus only contributes a surface term to the action $S_{QCD} = \int \mathcal{L}_{QCD} d^4x$. Although surface terms can usually be neglected, they cannot be neglected in non-Abelian gauge theories because of the existence of finite energy, color instanton solutions.

Because of this term, QCD physics is θ-dependent. In particular, this term leads to an electric dipole moment for the neutron of order[176]

$$d_n \approx 5\times10^{-16}\theta \text{ e-cm} . \tag{6.2}$$

The present experimental upper bound of $|d_n| < 10^{-24}$e-cm (ref. 177) then implies that θ must be less than about 2×10^{-9}. Recall that θ is an <u>arbitrary</u> parameter whose value is left unspecified by QCD; why then is it so small? This is the strong CP problem.

[Under a chiral rotation of the quark fields, $q_j \to \exp(i\alpha\gamma_5)q_j$, $\theta \to \theta - 2N\alpha$, and the quark mass matrix picks up a phase $\exp(iN\alpha)$. A complex quark mass matrix also results in strong CP violation. In order to avoid strong CP violation θ-arg det (quark mass matrix) must vanish. Only if one of the quarks is massless is it possible to perform a chiral rotation which brings θ to zero, without introducing a complex phase into the quark mass matrix. In this case there is no strong CP problem. Although both the u and d quarks are very light, $m_u \approx 5$ MeV $m_d \approx 9$ MeV, $m_u = 0$ or $m_d = 0$ is not consistent with current algebra calculations of their masses.]

The most elegant solution to the strong CP problem yet is that due to Peccei and Quinn[178]. By introducing an additional $U(1)_{PQ}$ global symmetry in the theory they effectively turn θ into a dynamical variable whose potential $V(\theta)$ is minimized for $\theta = 0$ (the CP-conserving value). The PQ symmetry is a spontaneously-broken symmetry (broken at a scale f_{PQ}), and therefore gives rise to an (almost) massless Nambu-Goldstone boson[179], the axion. The axion is not exactly massless, because instanton effects also break the PQ symmetry (it is actually then a

$U(1)_{PQ}$ quasi symmetry). The axion's mass is $\approx m_\pi^2/f_{PQ}$. In their original model[178] Peccei and Quinn suggested $f_{PQ} \approx O(300\ \text{GeV})$ -- the weak scale, implying $m_a \approx$ few 100 keV. However, as we shall see astrophysical arguments (as well as accelerator searches) preclude $f_{PQ} \lesssim O(10^8\ \text{GeV})$. Subsequently, Kim[180] and Dine, Fischler, and Srednicki[181] pointed out that $f_{PQ} >> 300$ GeV solves the strong CP problem just as well and makes the axion 'invisible' (i.e., very weakly-interacting, as its coupling strength $\propto 1/f_{PQ}$) and hence astrophysically safe -- although, as we shall see not necessarily cosmologically safe.

The Axion

Let $\vec{\phi} = \phi e^{i\theta}$ be the complex scalar field whose vacuum expectation value, $\langle|\vec{\phi}|\rangle = \langle\phi\rangle = f_{PQ}$, spontaneously breaks the $U(1)_{PQ}$ symmetry. The axion, a, then is the Nambu-Goldstone boson associated with the breakdown of the global $U(1)_{PQ}$, and corresponds to the θ degree of freedom, $a = \theta f_{PQ}$. At high temperatures ($T >> \Lambda_{QCD} \approx$ few 100 MeV) the axion is very nearly massless which corresponds to $V(\vec{\phi})$ being flat in the θ - direction. At low temperatures ($T \lesssim \Lambda_{QCD}$), SU(3) color instanton effects lead to $V(\vec{\phi})$ developing minima at: $\theta = (2\pi/N)n$, $n = 0, 1, 2, \ldots, N-1$, and the axion a mass (see Figs. 6.1 a,b). Here N is a positive integer whose value depends upon the PQ charges of the quarks; for the simplest models N=6. [Both PQ symmetry breaking and the instanton effects leave a discrete Z_N symmetry unbroken.[182]] At low temperatures, $V(\vec{\phi})$ 'anchors' θ in a CP-conserving minima; for more details of exactly how this works, see refs. 178 and 183.

The axion terms in the Lagrangian relevant for our purposes can be written as

$$\mathcal{L}_{\text{axion}} = +\cdots + - 1/2\ \partial\mu a \partial^{\mu} a - 1/2\ m_a^2 a^2$$

$$+ i\sum_f \chi_f (a/f_{PQ}) m_f \bar{f}\gamma_5 f + 0.09(N\alpha/\pi)(a/f_{PQ}) F_{\mu\nu}\tilde{F}^{\mu\nu} , \qquad (6.3)$$

where $V(\theta)$ has been expanded about $\theta = 0$, $m_a^2 \equiv V''(0)$ is of course the temperature-dependent axion mass (squared), f is a fermion field of mass m_f, and the last term on the rhs of Eqn. (6.3) represents the a-γ-γ coupling due to the triangle anomaly ($F_{\mu\nu}$ is the electromagnetic field strength tensor). The dimensionless constants χ_f depend upon the PQ charge assignments in the model; in general χ_f is of order unity for all fermion fields. However, there are models in which none of the leptons, or even none of the ordinary quarks, couple to the axion (i.e., $\chi_\ell = 0$ or $\chi_\ell = \chi_q = 0$ at tree level). Note that the coupling strength of the axion to matter varies as f_{PQ}^{-1}, <u>cf.</u> Eqn. (6.3), so that the axion becomes 'invisible' so to speak in the limit of $f_{PQ} \to \infty$. For more details about the coupling of the axion to quarks and leptons, see, e.g., refs. 183, 184.

The a-γ-γ coupling allows the axion to decay, $a \to 2\gamma$, but with a very long lifetime

$$\tau_a \approx \alpha^{-2} N^{-2} m_a^{-3} f_{PQ}^2 , \qquad (6.4)$$

$$\approx 10^{31}\text{yrs}\ (f_{PQ}/10^{10}\ \text{GeV})^5 .$$

The mass of the axion is given by

$$m_a = [\sqrt{z}/(1+z)] N f_\pi m_\pi / f_{PQ} , \qquad (6.5)$$

$$\approx 0.6N \times 10^{-3}\ \text{eV}\ (f_{PQ}/10^{10}\ \text{GeV})^{-1},$$

where $z = m_u/m_d \approx 0.55$, $f_\pi \approx 95$ MeV, and $m_\pi \approx 140$ MeV (see, e.g., ref. 184 and references therein).

Axion Emission from Stars

The central temperatures of stars undergoing nuclear fusion range from a few $\times\ 10^6$K ($\approx$ 0.1 keV) for very low mass stars (0.1 $M_\odot$ or so) up to 10^8 K or more ($\approx$ 10 keV) for stars in their red giant phases. Neutron stars and white dwarfs are born very hot, temperatures up to 10^{11} or 10^{12} K (up to 100 MeV). In a very simplistic view, a star spends its lifetime trying to get rid of its nuclear free energy. The timescale for its evolution is set by how rapidly it can cool itself. Cooling by photon emission is very inefficient because of the very short mean free path of a photon in matter

$$\Lambda \lesssim (n_e \sigma_T)^{-1},$$

$$\approx (\rho/\text{g cm}^{-3})^{-1}\ 3\text{cm},$$

where n_e is the number density of free electrons and $\sigma_T \approx 0.66 \times 10^{-24}\text{cm}^2$ is the Thomson cross section. [I included the '$\lesssim$ sign' because often other processes dominate the scattering/absorption of photons.] For this reason, when the primary cooling mechanism is photon emission, the timescale for evolution is very long measured in millions or billions of years. [Note, the time required for a photon to travel the 10^{11}cm from the center of the sun to the surface is: $t \approx (10^{11}\text{cm}/\lambda)^2\ \lambda/c \approx 10^7$ yrs.]

On the other hand, cooling by the emission of weakly-interacting particles, such as neutrinos, is very efficient because most stars are transparent to them. However, because of the very feeble strength of neutrino interactions and the severe temperature dependence of their production cross section ($\mathcal{L}_\nu \propto T^8$ in main sequence stars), neutrino cooling does not become important until the interior temperature of a star is $\gtrsim$ few 10^8K. Such temperatures are reached in stars burning carbon or heavier elements in their cores (see Fig. 6.2). When neutrino cooling dominates, the evolution of the star accelerates. During the final stages of stellar evolution when the interior temperatures are $\gtrsim$ 10^8K the timescale for evolution is very short — 1000's of years down to fractions of a second. For this reason one does not expect to observe stars burning carbon or heavier elements — and in fact there is no firm evidence for such stars being observed (see, however, ref. 185).

Sato and Sato[186] were the first to point out the possible importance of the emission of light, weakly-interacting bosons from stars. Their results have been generalized to axions by a number of authors,[187-191] and lead to an astrophysical <u>lower</u> bound to f_{PQ} of about 10^8 GeV. That is, an upper bound to the coupling strength. I will now briefly summarize the basic physics of stellar axion emission and how such a bound comes about.

For most stars the dominant axion production process is Compton-like photoproduction (see Fig. 6.3): $\gamma + e^- \to e^- + a$, whose thermally-averaged cross section can be calculated easily from $\mathcal{L}_{axion}$:

$$\langle \sigma_c v \rangle \approx (\alpha/f_{PQ}^2)\ (3T/m_e)^2 \ , \qquad (6.6)$$

where T is the temperature at the production site. Axions produced in stars will stream right out, and so the energy loss due to axion emission is just

$$\mathcal{L}_a = \int_{star} \langle E_a \rangle n_e n_\gamma \langle \sigma_c v \rangle dV \ , \tag{6.7}$$

where $\langle E_a \rangle \approx 3T$ is the average energy of the emitted axion, n_e and n_γ are the number density of electrons and photons respectively, and the integral extends over the entire volume of the star.

It is conventional to discuss energy generation and loss in stars in terms of the energy generated or emitted per gram of material per sec ($\equiv \varepsilon$) where for axion emission via the Compton-like process,

$$\varepsilon_c \approx n_e n_\gamma \langle \sigma_c v \rangle \langle E_a \rangle / \rho \ . \tag{6.8a}$$

For reference, the nuclear energy generation rate at the center of the sun, $\varepsilon_N \approx$ few erg g^{-1} sec^{-1}, and the total mass and luminosity of the sun are 2.0×10^{33}g and 4×10^{33}erg sec^{-1} respectively.

The number density of electrons is related to the mass density ρ by:

$$n_e = (N_A/\mu_e)(\rho/g\ cm^{-3}),$$

where μ_e ($\mu_e^{-1} \equiv \sum_i (X_i Z_i / A_i)$; X_i = mass fraction of isotope A_i, Z_i) is average molecular weight per electron (1 for pure H, 2 for pure ^{4}He), and $N_A = 6.02 \times 10^{23}$ is Avagadro's number. Assuming local thermal equilibrium (a very good assumption), the number density of photons is as usual given by:

$$n_\gamma = 2\,\zeta(3)T^3/\pi^2 \,.$$

Pulling all these facts together it follows that

$$\varepsilon_c \approx 1\mu_e^{-1}\text{erg g}^{-1}\ \text{sec}^{-1}\ (T/10^7\text{K})^6(f_{PQ}/10^7\text{GeV})^{-2}. \qquad (6.8b)$$

For the Sun, $\mu_e \approx 1.15$, $T \approx 1.5 \times 10^7$K, and $\varepsilon_N \approx$ few erg g^{-1} sec^{-1}; thus for $f_{PQ} \lesssim$ few $\times$ 10^7GeV the energy loss due to axion emission would <u>exceed</u> the energy generated by nuclear reactions -- assuming that the Sun were not to adjust. The thermal relaxation time for the Sun (the time required for it to radiate away its thermal energy reserve) is $O(10^7\text{yrs})$; if ε_c were $> \varepsilon_N$, the Sun (or any star) would necessarily have to adjust itself to correct this condition on this timescale. The nuclear energy generation rate is very temperature dependent; to reachieve equilibrium T would necessarily increase. This in turn would increase the rate at which the Sun is liberating its free energy -- thereby <u>decreasing</u> its main sequence lifetime.

We know the age of the Sun from geological considerations: about 4.6 Byr. The evolution of the Sun can be accurately modeled,[192] yielding a calculated age which is consistent with 4.6 Byr to 10% or so. Therefore it is clear that f_{PQ} had better be greater than a few times 10^7 GeV in order not to upset this agreement. Applying similar reasoning to red giants (in this case the requirement being that they evolve slowly enough to be seen by astronomers), one obtains a slightly better lower bound to f_{PQ}

$$f_{PQ} \gtrsim 10^8 \text{ GeV (Red Giants)}$$

(obtained for red giants with $T \approx 10^8$K and $\varepsilon_N \approx 100$ g erg^{-1} sec^{-1}, by requiring $\varepsilon_c \lesssim 100$ g erg^{-1} sec^{-1}).

[Note that for axions more massive than a few $T_c \approx 3$–30 keV, their emission is suppressed by a factor of $(m_a/T)^{3/2} \exp(-m_a/T)$ -- the fraction of photons above threshold for axion production. For the original PQ axion, $f_{PQ} \approx 300$ GeV and $m_a \approx 200$ keV. Thus their emission from the Sun ($T_\odot \approx 1.6\times10^7$K) is suppressed by a factor of $\approx 10^{-64}$, making their effect on the Sun negligible. However, for red giants $T \approx 10^8$K and the suppression is only a factor of $\approx 10^{-8}$, resulting in ε_c (red giant) $\approx 10^7$ erg g^{-1} sec^{-1} >> ε_N (red giant) ≈ 100 erg g^{-1} sec^{-1}, thereby ruling them out; for more details see ref. 187.]

Let's examine the effect of axion emission on the Sun in more detail. As discussed earlier, the photon luminosity of the Sun is determined by photon diffusion. Just by analyzing the thermal transport of energy (i.e., photons) and assuming hydrostatic equilibrium, Chandrasekhar[193] has derived the so-called luminosity formula, valid for stars less massive than about 2 $M_\odot$:

$$\mathcal{L}_\gamma \propto \mu_e^7 G_N^7 M^5 T_c^{1/2} , \tag{6.9}$$

where T_c is the central temperature. [The luminosity formula is simple to derive. Hydrostatic equilibrium implies: $(M/\mu_e)kT \approx GM^2/R$. Radiative transfer implies that $\mathcal{L}_\gamma \approx R^3T^4/\tau_{diff}$, where $\tau_{diff} = R^2/\lambda$ is the diffusion time for a photon and $\lambda = (\kappa\rho)^{-1}$, κ = the opacity. For a star like the Sun the opacity $\kappa = \kappa_0\rho T^{-3.5}$ and $\rho \propto T^3\mu_e^{-3}M^{-2}$. Taken together, the equations of radiative transfer and hydrostatic equilibrium lead to Eqn. (6.9).]

In equilibrium, nuclear energy generation must provide the energy for both axion emission and photon emission:

$$Q_N = \mathcal{L}_a + \mathcal{L}_\gamma \,, \tag{6.10}$$

where

$$Q_N = \int_{star} \varepsilon_N \, dM \,,$$

$$\mathcal{L}_a = \int_{star} \varepsilon_c \, dM \,.$$

In general, $\varepsilon_N \propto \rho_c T_c^{\,n}$, where n depends upon T_c. The constancy of the entropy/baryon ($\rho/T^3 \approx$ const), implies that $\rho_c \propto T_c^{\,3}$ and thus $\varepsilon_N \propto T_c^{n+3}$. For stars like the Sun which have central temperatures $\approx 15 \times 10^6$K and burn Hydrogen via the p-p chain ($p + p \rightarrow D + e^+ + \nu$, $D + D \rightarrow {}^4He$) $n \approx 4$, so that $Q_N \propto T_c^7$. For slightly hotter stars ($T_c \gtrsim 20 \times 10^6$K) Hydrogen is burned via the CNO cycle and $n \approx 16$-20.

In the absence of axion emission the condition that $\mathcal{L}_\gamma$ ($\equiv \mathcal{L}_o$) = Q_N fixes the central temperature (and hence structure of the Sun), since $\mathcal{L}_\gamma \propto T_c^{1/2}$ and $Q_N \propto T_c^7$. Now consider the effect of axion emission on the Sun. Suppose $\mathcal{L}_a = \varepsilon \mathcal{L}_o$ (where $\varepsilon \ll 1$), then by using Eqns. (6.9 and 6.10) and $Q_N \propto T_c^7$, it follows that

$$\frac{\delta T_c}{T_c} \approx \frac{\varepsilon}{6.5} \,, \tag{6.11a}$$

$$\frac{\delta \mathcal{L}_\gamma}{\mathcal{L}_o} \approx \frac{\varepsilon}{13} \,. \tag{6.11b}$$

That is, for $\mathcal{L}_a = 1/2\mathcal{L}_o$ (corresponding to $\varepsilon_c \approx 1/2\ \varepsilon_N$), the central temperature increases by about 8%, the photon luminosity increases by about 4%, and the total luminosity increases by about 54%. Such changes would be very noticeable! Increasing the total luminosity by 54% would shorten the lifetime at a given epoch by about 30% or so. The predicted flux of the ^8B solar neutrinos (the ones which are detected in Davis's ^{37}Cl detector) is very temperature dependent: rate (^8B ν's) $\propto T_c^p$; $p \approx 13$ (the power p depends upon which quantities in the perturbed stellar model are being held constant. Holding the luminosity constant implies $p \approx 13$. In general $p \gtrsim 13$; for more details see ref. 192). An 8% increase in the central temperature would almost <u>triple</u> the predicted ^{37}Cl capture (SNU) rate in Davis's experiment.[192] Clearly, ε_c is large as even $\varepsilon_N/2$ is not possible.

Thus far I have focused on Compton-like production of axions. At low temperatures (corresponding to stars of mass $\ll M_\odot$, as $T_c \propto M^{1/3}$), the bremsstrahlung process[194], $e^- + Z \to e^- + Z + a$, and the Primakoff process[187], $\gamma + Z \to Z + a$, become important (see Fig. 6.3), as their temperature dependence is milder than the T^6 for the Compton process,

$$\langle\sigma_B v\rangle \approx (2\times10^{-2}Z^2\alpha^2/f_{PQ}^2)\ (3T/m_e)^{3/2}, \tag{6.12a}$$

$$\langle\sigma_p v\rangle \approx (N^2/\pi^2)Z^2(\alpha^3/f_{PQ}^2), \tag{6.12b}$$

$$\varepsilon_B \approx (N_A/\mu_e)(N_A/\mu)\rho\ \langle E_a\rangle\langle\sigma_B v\rangle \tag{6.12c}$$

$$\approx 3\mu^{-1}\mu_e^{-1}(\rho/100\mathrm{gcm}^{-3})\ (T/10^7\mathrm{K})^{5/2}(f_{PQ}/10^7\mathrm{GeV})^{-2}\mathrm{erg\ g}^{-1}\ \mathrm{sec}^{-1},$$

$$\varepsilon_p \approx (N_A/\mu)\ n_\gamma\ \langle E_a\rangle\ \langle\sigma_p v\rangle \tag{6.12d}$$

$$\approx 1\mu^{-1}(N/6)^2\ (T/10^7K)^4(f_{PQ}/10^7GeV)^{-2}erg\ g^{-1}\ sec^{-1},$$

where $\mu^{-1} = \Sigma(X_i Z_i(Z_i+1)/A_i)$. Note the very mild temperature dependence of ε_B is in part due to the fact that $\varepsilon_B \propto n_Z$ rather than n_γ which itself $\propto T^3$. [For stars less massive than a few $M_\odot$, $n_Z/n_\gamma \approx 300\mu^{-4}(M/M_\odot)^{-2}$.] As emphasized by Dicus etal[87], the Primakoff process is particularly important because it provides a more model independent bound on f_{PQ}. In axion models where $\chi_e = 0$,[95] $\sigma_c = \sigma_B \approx 0$; whereas σ_p is directly related to $\Gamma(a \to 2\gamma)$ which always arises due to the triangle anomaly. For $N = 6$, the Primakoff and Compton processes are roughly comparable in the Sun; while for $N < 6$ the Compton process dominates as $\varepsilon_p/\varepsilon_c \propto N^2$.

For low mass stars ($M \lesssim M_\odot$) ε_N varies as T^n ($n \gg 4$). Thus such stars seem more vulnerable than the Sun to axion emission, and so one might hope to be able to derive a more stringent bound on f_{PQ}. In order to do this, however, one needs observable quantities which are affected by axion emission. Since stars much less massive than 1 $M_\odot$ have lifetimes which are much longer than the age of the Universe [$\tau \approx 10$ Byr $(M/M_\odot)^{-3 \text{ or } -4}$], their ages do not provide such a quantity. However, their surface temperatures and luminosities may provide the necessary observables.[96]

The Cooling of White Dwarfs and Neutron Stars

White dwarfs and neutron stars are born hot (white dwarfs – temperatures up to 10^9K, neutron stars – temperatures up to 10^{11}K). As they radiate neutrinos, photons (and axions) they cool:

$$-C_*(dT/dt) = \mathcal{L}_\nu + \mathcal{L}_\gamma + \mathcal{L}_a\ , \tag{6.13}$$

where C_* is the heat capacity of the star. In general, $\mathcal{L}_\gamma \propto T^4$ and $\mathcal{L}_\nu \propto T^8$, i.e., photon cooling is most important at low temperatures and neutrino cooling at high temperatures.[197,198] Iwamoto[191] has shown that for $f_{PQ} \lesssim$ few $\times 10^8$GeV, axion cooling can dominate neutrino cooling in some temperature regimes. Since $\mathcal{L}_a$ and $\mathcal{L}_\nu$ are not observables, the implications of this fact are not clear.

In order to make use of the cooling of white dwarfs and neutron stars to obtain bounds on f_{PQ}, one must calculate quantities which are observable, for example, temperature vs. age diagrams (the theoretical predictions are obtained by integrating Eqn. (6.13)). For white dwarfs, one finds that the observed age-temperature diagram is in conflict with theory for $f_{PQ} \lesssim 10^8$GeV[199] (see Fig. 6.4).

Axions and Cosmology

Perhaps the most intriguing aspects of the 'invisible' axion (i.e., axion with $f_{PQ} \gg 300$ GeV) are its cosmological implications. Recall that the vacuum expectation value of $\vec{\phi}$, $|\vec{\phi}| = \phi = f_{PQ}$, spontaneously breaks the PQ symmetry. At temperatures $T \gtrsim O(f_{PQ})$, finite temperature effects restore the PQ symmetry, so PQ symmetry breaking occurs at a temperature of $O(f_{PQ})$. When it does, θ is left unspecified as $V(\theta) = 0$. Thus θ will take on different values in different causally-distinct regions (i.e., regions of size $\simeq H^{-1} \simeq m_{pl}/f_{PQ}^2$). In the standard cosmology, this means that today the Universe is comprised of about $10^{59}(f_{PQ}/10^{10}\text{GeV})^3$ such regions. Of course, if inflation occurs after or during PQ symmetry breaking, the entire Universe is contained within one such region (where $\theta \simeq$ const).

In the non-inflationary case, as the Universe evolves θ smoothes itself on scales $\lesssim H^{-1}$ — the modes with $\lambda_{phys} \lesssim H^{-1}$ correspond to axions with momenta $p_a \approx \lambda^{-1}$ and redshift away (see Figs. 6.5 a,b). Thus at any given time the correlation length for θ is $O(H^{-1})$. [To be more precise, the equation of motion for the Fourier component $\delta_\theta(k)$ is: $\ddot{\delta}_\theta + 3H\dot{\delta}_\theta + (k^2/R^2)\delta_\theta = 0$. For modes with $\lambda_{phys} \approx R/k \lesssim H^{-1}$, $\delta_\theta \propto R^{-1}$.] When $T \approx O(\Lambda_{QCD})$ and $t \approx 10^{-5}$sec, the axion potential $V(\theta)$ develops minima at $\theta = n(2\pi/N)$, $n=0, 1, 2..., N-1$. I will now discuss the cosmological implications of the fact the value of θ in different horizon volumes is uncorrelated and not necessarily aligned with any of the minima — recall when PQ symmetry breaking occured the axion 'did not know' that the angles $\theta = n(2\pi/N)$ were special!

The Domain Wall Problem[82]

In models where $N > 1$, there are more than one degenerate vacua. This means that different horizon volumes will end up in different vacua (distinguished by $\theta = n(2\pi/N)$) when the axion potential develops at $t \approx 10^{-5}$sec and $T \approx O(\Lambda_{QCD})$. These different vacua will be separated by domain walls of thickness Δ and surface density σ, where

$$\Delta \approx m_a^{-1} \approx 10^{-2}\text{cm } (f_{PQ}/10^{10}\text{GeV}),$$

$$\sigma \approx f_\pi m_\pi f_{PQ} \approx 5\times10^{11}\text{g cm}^{-2} \ (f_{PQ}/10^{10}\text{GeV}).$$

As the Universe evolves and the horizon grows regions of space in the same vacuum state (specified by n) will coalesce (because of the surface tension in the walls the wall area tends to be minimized). However, one

is always left with the order of at least one domain wall per horizon volume, with the resulting energy density in walls

$$\rho_{wall} \simeq t^2\sigma/t^3 \simeq \sigma/t, \tag{6.14}$$

$$\simeq 15\text{g cm}^{-3}\ (f_{PQ}/10^{10}\text{GeV})(t/\text{sec})^{-1}$$

Because $\rho_{wall} \propto t^{-1}$ walls would quickly come to dominate the energy density of the Universe, causing the Universe to expand as: $R(t) \propto t$. In such a model Universe $T \propto R^{-1} \propto t^{-1}$ (rather than $t^{-1/2}$ for a radiation-dominated Universe or $t^{-2/3}$ for a matter-dominated Universe), and the Universe would reach a temperature of 3K at a very tender age: $t \simeq 10^3$ yr $(f_{PQ}/10^{10}\text{GeV})^{-1/2}$. A wall-dominated Universe is a very bad thing! [Zel'dovich, Kobzarev, and Okun first warned cosmologists of the potential dangers of domain walls,[200] and Sikivie[182] alerted us to the problem with respect to axions. Seckel[201] has discussed the remote possibility of wall-dominated inflation -- recall if $R(t) \propto t$, the horizon problem is solved as $d_H(t) \propto t\ \ell nt$.]

Clearly, one must arrange things to avoid the axion domain wall problem! Inflation during or after PQ symmetry breaking certainly does the trick, because all of the Universe we observe would be within one initial θ-domain (and necessarily, then in the same vacuum). It is also possible to construct axion models with $N = 1$ (see, e.g., J. Kim, ref. 180). One can also solve the problem by arranging things so that the different vacua are gauge-equivalent. This occurs in models where the Z_N symmetry coincides with or lies within the center of the gauge group.[202] In this case the N vacua are not distinct because they are gauge equivalent.

The Axion Energy Crisis[203] and the Axion-Dominated Universe

In the discussion above I rather slyly assumed that in each horizon volume $\theta(x)$ was at one of the minima. Of course, initially that is not the case! In general θ will be misaligned — remember, at the epoch of PQ symmetry breaking, when $\theta(x)$ was 'laid out' so to speak, $V(\theta)$ was flat, and so there was no reason for $\theta(x)$ to be near the minimum of $V(\theta)$. Once the potential starts to develop its minima (when $T \approx \Lambda_{QCD}$), θ will begin to oscillate due to its initial misalignment. As we shall see these coherent scalar field oscillations are equivalent to a condensate of very NR axions. An enormous energy is stored in the initial misalignment of $\theta(x)$, which is released when θ begins to oscillate and results in the creation of very many, very cold axions.

Within any given horizon volume (in the inflationary case, within the observable Universe) θ is constant, the higher spatial modes having been redshifted away. Denote the initial misalignment angle as θ_1. Taking θ to be spatially constant, its Lagrangian density is just

$$\mathcal{L} = R^3[1/2\ \dot{\theta}^2 - V(\theta)]f_{PQ}^2, \tag{6.15}$$

where $V(\theta) \approx 1/2\ m_a^2\theta^2$. The equation of motion for θ is just

$$\ddot{\theta} + 3H\dot{\theta} + m_a^2\theta = 0\ , \tag{6.16}$$

where $H = \dot{R}/R$ is as usual the expansion rate of the Universe. Recall that the axion mass is temperature dependent: $m_a(T) \to 0$ for $T \to \infty$, and $m_a(T) \to m_a$, <u>cf.</u> Eqn. (6.5), as $T \to 0$. The energy density associated with the axion field ρ_a is just

$$\rho_a = (1/2\ \dot{\theta}^2 + 1/2\ m_a^2\theta^2)f_{PQ}^2\ , \tag{6.17}$$

and from Eqn. (6.16) it follows that the evolution of ρ_a is governed by

$$\dot{\rho}_a = -3H\dot{\theta}^2 f_{PQ}^2 + (\dot{m}_a/m_a)\ m_a^2 f_{PQ}^2\theta^2. \tag{6.18}$$

Qualitatively the evolution of θ is as follows: (1) when $m_a \lesssim 3H$ (true for $T \gg \Lambda_{QCD}$), $\theta \approx \theta_1 \approx$ const; (2) when $m_a \gtrsim 3H$ (true for $T \ll \Lambda_{QCD}$) θ oscillates with angular frequency m_a. Once θ begins to oscillate $\dot{\theta}^2 f_{PQ}^2$ and $m_a^2 f_{PQ}^2\theta^2$ can be replaced by their cycle averages (as they are changing much more rapidly than m_a and ρ_a which change on a timescale of H^{-1}): $\langle\dot{\theta}^2 f_{PQ}^2\rangle = \langle m_a^2\theta^2 f_{PQ}^2\rangle = \rho_a$ (for more details see refs. 203, 204). Now Eqn. (6.18) becomes

$$\dot{\rho}_a = [(\dot{m}_a/m_a) - 3H]\rho_a, \tag{6.19}$$

which can be easily integrated:

$$\rho_a = \text{const} \times m_a/R^3\ , \tag{6.20}$$

which corresponds to $\theta \propto m_a^{-1/2}R^{-3/2}$. Note that once $m_a \approx$ const, $\rho_a \propto R^{-3}$ $\approx$ just like NR matter. Taking the onset of oscillation to begin when $m_i \equiv m_a(T_i) = 3H(T_i)$ it follows that

$$\rho_a = 1/2\ m_i m_a \theta_1^2 f_{PQ}^2 (R_i/R)^3 \tag{6.21}$$

for $T \lesssim T_i$. The evolution of θ is shown schematically in Fig. 6.6.

The axion oscillations begin when the Universe is radiation-dominated and $T_i \approx O(\mathrm{GeV})$. The expansion rate H is just

$$H \approx 1.66 g_*^{1/2} T^2/m_{pl} \ .$$

The finite temperature behaviour of the axion mass has been calculated by Gross, Pisarski and Yaffe[205]. In the limit of $T \gg \Lambda \equiv \Lambda_{QCD}$ their results imply:

$$m_a(T)/m_a = a\Lambda_{200}^{b}(\Lambda/T)^{c}[1-\ln(\Lambda/T)]^{d} \ , \tag{6.22}$$

where a,b,c,d depend upon the number of light ($m \ll T$) quark flavors N_f. For $N_f = 2-5$, $a = .035-.12$, $b = 1-(-1/2)$, $c = 3.83-4.33$, and $d = 1.02-1.74$. Using Eqn. (6.22) and $g_* \approx 60$, one can calculate T_i approximately[206]:

$$T_i \approx 1.2 \text{ GeV } (N/6)^{0.18}\Lambda_{200}^{0.7}(f_{PQ}/10^{12}\text{GeV})^{-0.18}, \tag{6.23a}$$

$$(f_{PQ} \lesssim 1.6 \times 10^{18} \text{ GeV } (N/6)\ \Lambda_{200}^{-2.3})$$

$$T_i \approx 160 \text{ MeV } (N/6)^{1/2}(f_{PQ}/10^{18}\text{GeV})^{-1/2} \ , \tag{6.23b}$$

$$(f_{PQ} \gtrsim 1.6 \times 10^{18} \text{ GeV } (N/6)\ \Lambda_{200}^{-2.3})$$

where $\Lambda = \Lambda_{200}$ 200 MeV.

Recall that the entropy per comoving volume, $S \equiv R^3 s = (2\pi^2/45) g_* T^3 R^3$, remains constant (in the absence of significant entropy

production). Using the entropy once again as a fiducial, we can calculate ρ_a (without reference to R and R_i):

$$\rho_a = (\rho_a/s)\ s, \tag{6.24a}$$

$$(\rho_a/s) = 0.71\ m_a f_{PQ}^2 \theta_1^2 / m_{pl} T_i\ . \tag{6.24b}$$

Using the fact that today $s \cong 1.7T^3$, and using $\gamma \equiv S(\text{today})/S_i$ to take into account any entropy production since $T = T_i$, we can calculate Ω_a, the fraction of critical density contributed by axions today,[206]

$$(\Omega_a h^2/T_{2.7}^3) = 1.0(N/6)^{0.83}\theta_1^2\gamma^{-1}\Lambda_{200}^{-0.7}(f_{PQ}/10^{12}\text{GeV})^{1.18}, \tag{6.25a}$$

$$(f_{PQ} \lesssim 1.6 \times 10^{18}\ \text{GeV}\ (N/6)\ \Lambda_{200}^{-2.3})$$

$$(\Omega_a h^2/T_{2.7}^3) = 3.5\times10^6(N/6)^{1/2}\theta_1^2\gamma^{-1}(f_{PQ}/10^{18}\text{GeV})^{1.5}, \tag{6.25b}$$

$$(f_{PQ} \gtrsim 1.6 \times 10^{18}\ \text{GeV}\ (N/6)\ \Lambda_{200}^{-2.3})$$

where the present photon temperature $T = 2.7\ T_{2.7}$ K.

Note that if $\theta_1 \cong O(1)$ and $\gamma \cong 1$, then $f_{PQ} \cong$ few $\times\ 10^{12}$ GeV corresponds to an axion-dominated Universe, while $f_{PQ} \gtrsim 10^{13}$ GeV corresponds to the 'over production' of axions ~~the~~ the so-called axion energy crisis.[203] If the Universe never inflated, or inflated before PQ symmetry breaking, then θ_1, the RMS value of $\theta(x)$, must be equal to $(\pi/N)/3^{1/2}$ which is of order unity.

The case where PQ symmetry breaking occurs before or during inflation is more interesting. In different fluctuation regions or bubbles (whichever is relevant) θ_1 will take on different values (see Fig. 6.7). The RMS average of θ_1 over many such domains must of course be $(\pi/N)/3^{1/2}$. But the RMS average of θ_1 is not what is relevant here; we live in but one of the bubbles or fluctuation regions. It is just as likely then that $|\theta_1|$ is between 0 and 0.1 as it is that it lies between 0.9 and 1.0. In that sense, Eqn. (6.25) serves to specify θ_1 in terms of f_{PQ}, h, $T_{2.7}$, Λ_{200}, and Υ. This point of view has been particularly emphasized by Pi[127] and in refs. 207, 208.

Since the number density of axions produced by the initial misalignment of θ is highly non-thermal, $n_a \approx O(f_{PQ}^2/m_{pl}T_i)T^3 \approx 10^5(f_{PQ}/10^{12}\text{GeV})^2T^3 >> n_\gamma$, one might wonder if the axions produced might thermalize, e.g., via processes like $a + e^\pm \to e^- + \gamma$. The rate for this process relative to the expansion rate is

$$\Gamma(a+e \to e+\gamma)/H \approx (m_e/f_{PQ})^2(\alpha/T^2)T^3/(T^2/m_{pl}) \quad (T \gtrsim m_e)$$

$$\approx (m_e/f_{PQ})^2(\alpha T^2/m_e^4)T^3/(T^2/m_{pl}) \quad (T \lesssim m_e)$$

For $T \lesssim O(\text{GeV})$, (Γ/H) is only $\gtrsim 1$ for $f_{PQ} \lesssim 10^7$ GeV, i.e., only for $f_{PQ} \lesssim 10^7$ GeV will the axions produced be able to thermalize. However, in this case n_a is not greater than n_γ in the first place! Thus, we conclude that thermalizing interactions cannot avert the axion energy crisis.

As mentioned earlier the energy density in these coherent axion oscillations correspond to a condensate of NR axions. It is amusing to

estimate their phase space number density. At the time of 'their birth', $T \approx T_i \approx O(\mathrm{GeV})$, the typical axion momentum $p_a \approx H \approx T_i^2/m_{pl}$ — higher momentum modes having been redshifted away. [Of course in the inflationary case $p_a \lll H$.] The momentum phase space volume they occupy then is:

$$V_p \approx T_i^6/m_{pl}^3 .$$

Their present number density is

$$n_a = \rho_a/m_a ,$$

$$\approx 3\times10^{10}\ (f_{PQ}/10^{13}\ \mathrm{GeV})^{2.2}\ \theta_1^2 \mathrm{cm}^{-3}.$$

When the temperature of the Universe was $\approx O(\mathrm{GeV})$ their number density was a factor of $(T_i/3K)^3 \approx 10^{37}$ or so larger. Their phase space density then must be greater than

$$\mathcal{N} \approx n_a(T_i/3K)^3(m_{pl}^3/T_i^6) , \tag{6.26}$$

$$\approx 10^{64}\theta_1^2(f_{PQ}/10^{13}\mathrm{GeV})^{2.7}!$$

Axions are indeed a Bose condensate.

Using Eqn. (5.38) we can calculate their 'freestreaming length'

$$\lambda_{FS} \approx 3 \times 10^{-5}\ \mathrm{Mpc} , \tag{6.27}$$

which is << 1 Mpc. Thus axions behave as (<u>very</u>) cold dark matter!

Isothermal Axion Perturbations

In an axion-dominated Universe ($\Omega_a \simeq 1$) which undergoes inflation after or during PQ symmetry breaking there will be in addition to the usual adiabatic perturbations in the axions (see Lecture 4), isothermal axion perturbations.[206,207,209] These isothermal fluctuations (more precisely 'isocurvature' fluctuations) correspond to fluctuations in the local axion-to-photon ratio (n_a/n_γ). They arise due to deSitter space induced quantum fluctuations in θ_1, the initial misalignment angle (which, at the classical level, is constant throughout the bubble or fluctuation region).

Expand the fluctuations in θ_1, in a Fourier series

$$(\delta\theta_1/\theta_1) = (2\pi)^{-3}\int d^3k\,\delta_\theta(k)e^{-ikx}. \tag{6.28}$$

After the axion oscillations begin ($T \lesssim \Lambda_{QCD}$), fluctuations in θ_1 will result in fluctuations in n_a since $n_a \propto \theta_1(x)^2$. In particular,

$$\delta(k) = 2\,\delta_\theta(k)\ , \tag{6.29}$$

where

$$\delta n_a/n_a = (2\pi)^{-3}\int d^3k\,\delta(k)e^{-ikx} \tag{6.30}$$

The deSitter space produced quantum fluctuations in the axion field $a(\equiv \theta f_{PQ})$ are characterized by

$$|\delta_a(k)|^2 = H^2/2k^3\ , \tag{6.31}$$

and so the fluctuations in θ_1 are given by

$$|\delta_\theta(k)|^2 = |\delta_a(k)|^2/f^2\theta_1^2 \ , \tag{6.32}$$

$$= H^2/(2\theta_1^2 f_{gal}^2 k^3) \ .$$

Here H is the expansion rate during inflation, and f_{gal} is the value of $|\vec{\phi}|$ when the scales of interest (galaxies, etc.) go outside the horizon during the inflationary epoch. If PQ symmetry breaking is occurring during inflation (e.g., if $|\vec{\phi}|$ is also the scalar field responsible for inflation, as in Pi's model[127]; see Fig. 6.8), then f_{gal} is not necessarily equal to f_{PQ}. Bringing everything together we have[207]

$$(\delta\rho_a/\rho_a)_{gal} \cong |\delta(k)|^2 k^3/(2\pi)^3 \ , \tag{6.33}$$

$$\cong H/(2\pi^{3/2} f_{gal}\theta_1) \ ,$$

$$\cong 0.1\alpha^{1/2}(H/f_{PQ})(f_{PQ}/f_{gal})(f_{PQ}/10^{12}\text{GeV})^{.59} \ ,$$

$$(f_{PQ} \leq 1.6 \times 10^{18} \text{ GeV } (N/6)\ \Lambda_{200}^{\cong 2.3})$$

$$\cong 170\beta^{1/2}(H/f_{PQ})(f_{PQ}/f_{gal})(f_{PQ}/10^{18}\text{GeV})^{.75} \ ,$$

$$(f_{PQ} \geq 1.6 \times 10^{18} \text{ GeV } (N/6)\ \Lambda_{200}^{\cong 2.3})$$

where $\alpha \cong T_{2.7}^{3}\ (N/6)^{.83} h^{-2}\ \gamma^{\cong 1} \Lambda_{200}^{\cong .7}$ and $\beta = T_{2.7}^{3}\ (N/6)^{1/2} h^{-2} \gamma^{-1}$.

In order to be relevant for galaxy formation $(\delta\rho_a/\rho_a)$ must be of the order of $\approx 10^{-4}$. [These isocurvature perturbations also lead to anisotropies in the μ-wave background. The observed isotropy of the μ-wave background constraints $(\delta\rho_a/\rho_a)$ to be less than a few × 10^{-4}; for more details, see refs. 207, 210.] If the reheating after inflation is 'good', $T_{RH}^2 \approx H\, m_{pl}$. In order that the PQ symmetry not be restored, T_{RH} must be $\lesssim f_{PQ}$. These two facts imply that $(H/f_{PQ}) \lesssim (f_{PQ}/m_{pl})$. If reheating is good and $f_{PQ} \approx 10^{12} - 10^{13}$ GeV (θ_1 order unity), then $(\delta\rho_a/\rho_a) \approx 10^{-7}$ — far too small to be of relevance for galaxy formation. However, if reheating is poor in which case H/f_{PQ} can be $\gtrsim (f_{PQ}/m_{pl})$, or if $f_{PQ} >> 10^{13}$ GeV (corresponding to $\theta_1 << 1$), $\delta\rho_a/\rho_a \approx 10^{-4}$ can easily be achieved. In Pi's model[127] $f_{PQ} \approx 10^{18}$ GeV and $f_{gal} \approx f_{PQ}/230$, and $(\delta\rho_a/\rho_a) \approx 10^{-4}$.

The spectrum of isothermal axion perturbations is flatter than the adiabatic one[207] (see Fig. 6.9) and so in the presence of isothermal axion perturbations galaxies form slightly later, and almost all scales go non-linear together. Whether or not the resulting structure is qualitatively different (than for adiabatic perturbations) is yet to be determined.

Detection of 'Invisible' Axions

The axion in models with $f_{PQ} >> 300$ GeV was originally dubbed the 'invisible' axion because of the weakness of its interactions (which $\propto f_{PQ}^{-1}$). However, Sikivie[211] has devised several experiments to detect the cosmic reservoir of axions which exists in an axion-dominated Universe. In an axion-dominated Universe, the halos of galaxies should be comprised of axions. The contribution of the halo of our galaxy to the

mass density at our position in the galaxy is about 5×10^{-25}g cm^{-3} (ref. 212), which corresponds to an axion number density of[242]

$$n_a(\text{halo}) \approx 5\times10^{14}\ N^{-1}\ (f_{PQ}/10^{13}\text{GeV})\text{cm}^{-3}.$$

Sikivie's technique exploits the axion's coupling to two photons to convert a handful of the many halo axions into photons in the presence of an inhomogeneous magnetic field. In a resonant detector of volume $(30\ \text{cm})^3$ he calculates that 30 or so axions are converted to photons per second. Several experiments based on his clever idea are presently being built. Moody etal.[243] have proposed a variation on Sikivie's idea which exploits the axion's coupling to electron spins.

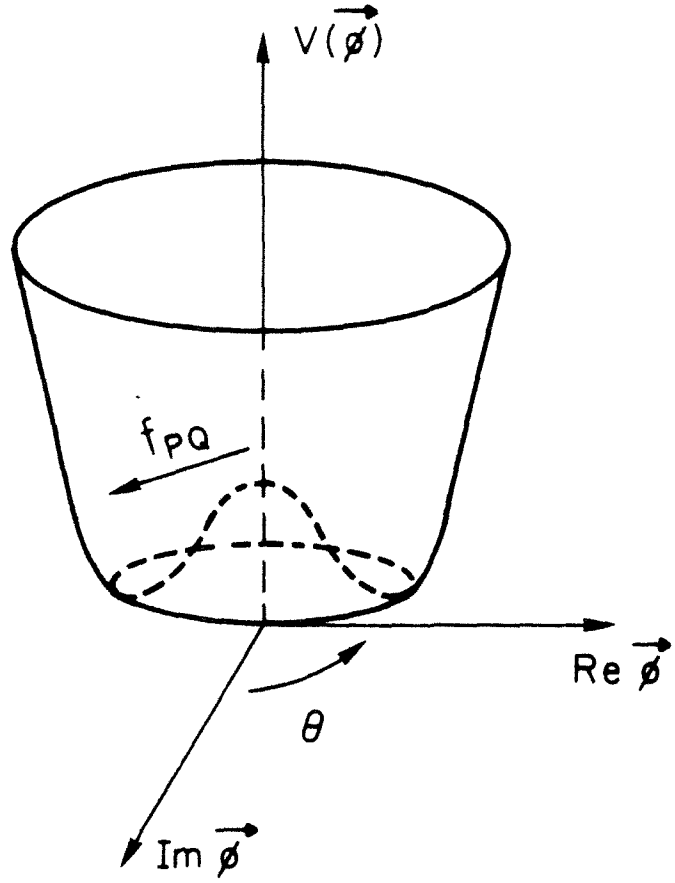


Fig. 6.1a – The potential $V(\vec{\phi})$ after PQ symmetry breaking. The vacuum expected value of $|\vec{\phi}| = f_{PQ}$ breaks the PQ symmetry. Until $T \lesssim \Lambda_{QCD}$ the θ-direction is flat.

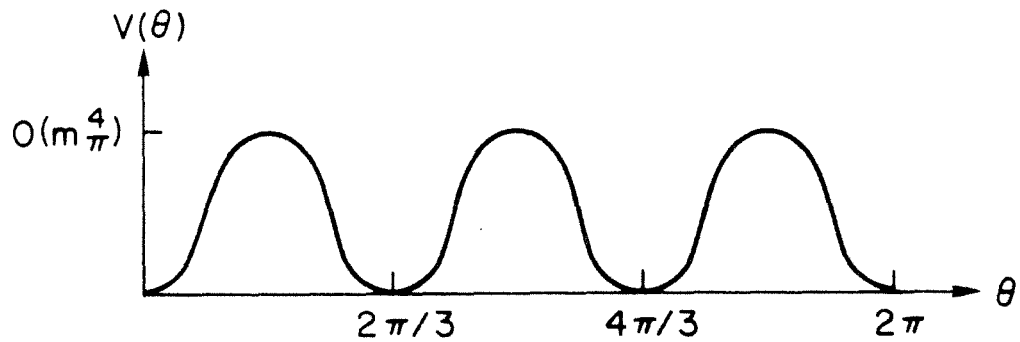


Fig. 6.1b – The axion potential $V(\theta)$ after instanton effects develop a mass for the axion ($T \lesssim \Lambda_{QCD}$). In general $V(\theta)$ has N degenerate vacua at $\theta = 0, 2\pi/N, \ldots, (N-1)2\pi/N$. Here $N = 3$.

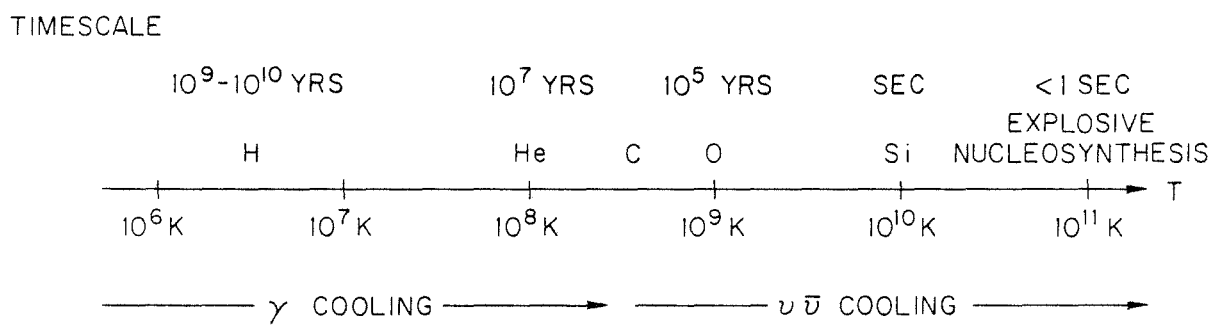


Fig. 6.2 – Timescales for and stages of stellar nucleosynthesis for core temperatures of $10^6 - 10^{11}$ K. Evolution timescale is determined by the cooling rate. At $T \lesssim$ few 10^8 K photon cooling dominates and the timescale is long; for $T \gtrsim$ few 10^8 K neutrino cooling dominates and the timescale is short.

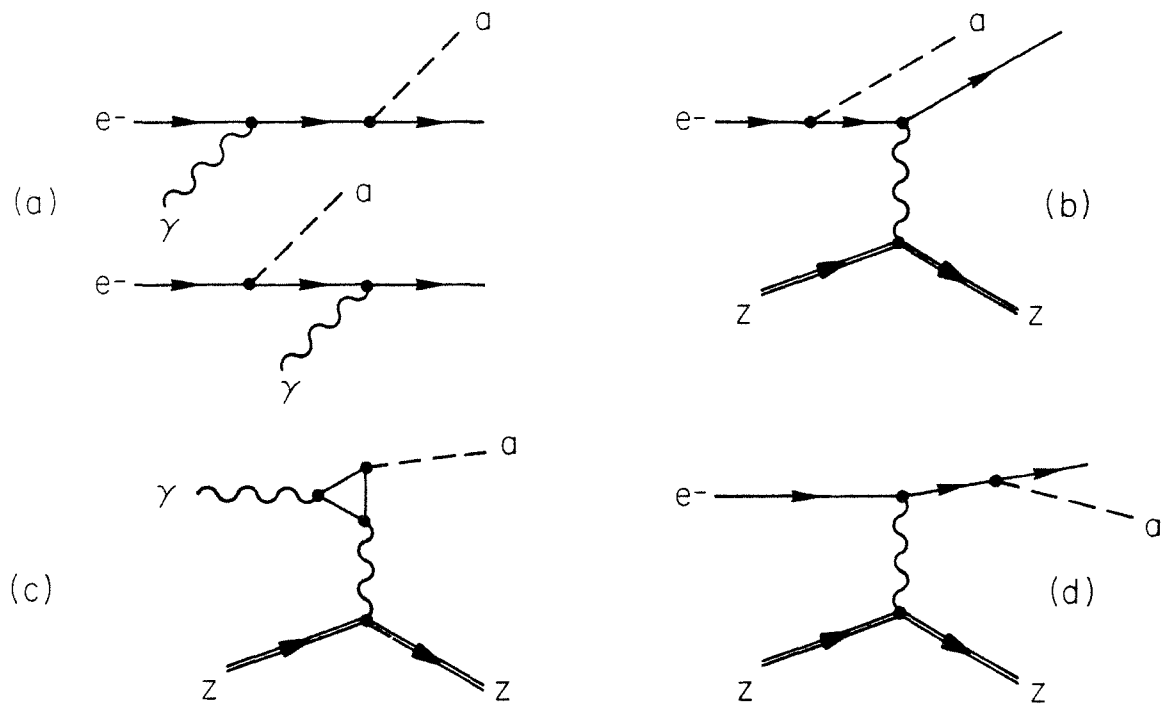


Fig. 6.3 – Feynman diagrams for stellar axion production: (a) Compton-like photoproduction, $\gamma + e^- \to a + e^-$; (b) and (d) bremsstrahlung, $Z + e^- \to Z + e^- + a$; (c) Primakoff process, $\gamma + Z \to Z + a$.

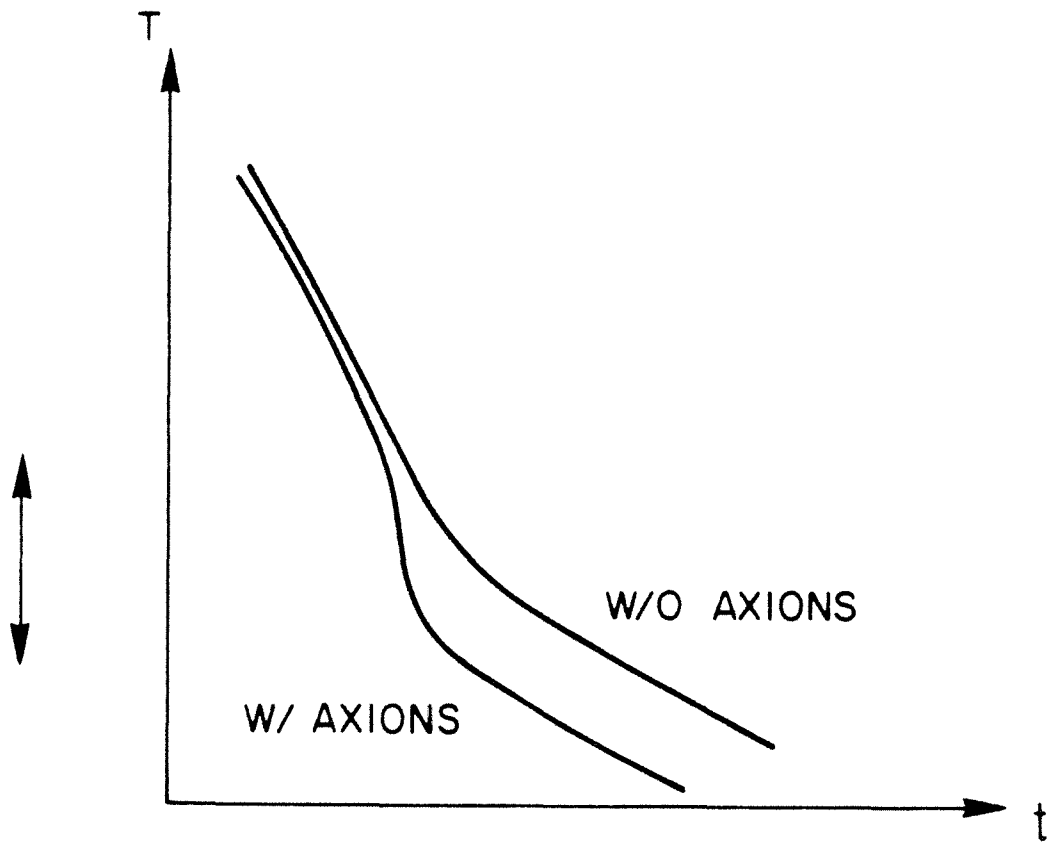


Fig. 6.4 – Schematic cooling curves (temperature as a function of age) for white dwarfs and neutron stars. At high temperatures neutrino cooling dominates ($\mathcal{L} \propto T^8$), while at low temperatures photon cooling dominates ($\mathcal{L} \propto T^4$). The effect of axion cooling is shown schematically; during the temperature range indicated axion emission is the dominant cooling process, and during that epoch the cooling process is hastened, resulting in a star which has a lower surface temperature for a specified age. The key point is that while $\mathcal{L}_a$ is <u>not</u> an observable, temperature as a function of age is potentially observable.

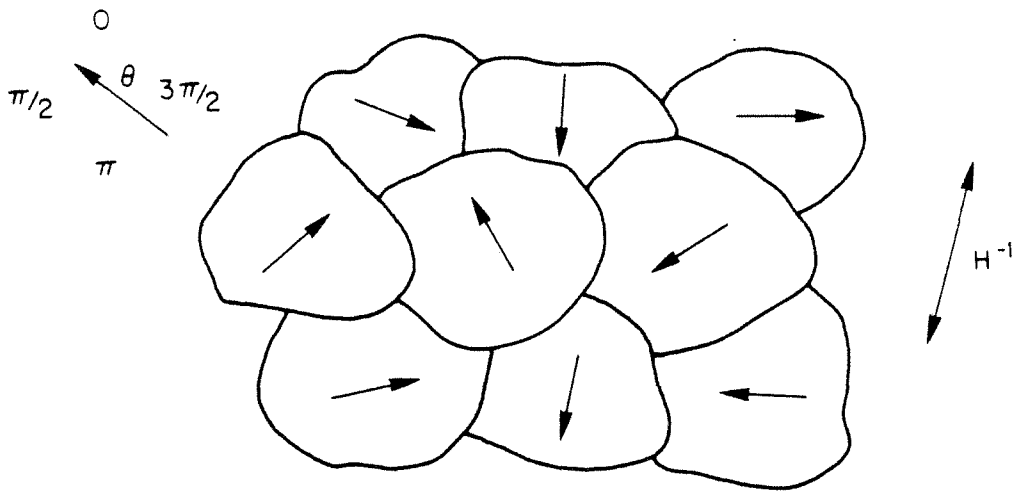


Fig. 6.5a – The spatial distribution of the axion angle θ (indicated by arrows) for $f_{PQ} \gtrsim T \gtrsim \Lambda_{QCD}$. Because the expansion redshifts all modes with wavelengths $\lesssim H^{-1}$, θ should always be uniform on scales $\lesssim H^{-1}$. On scales larger than H^{-1}, θ is uncorrelated.

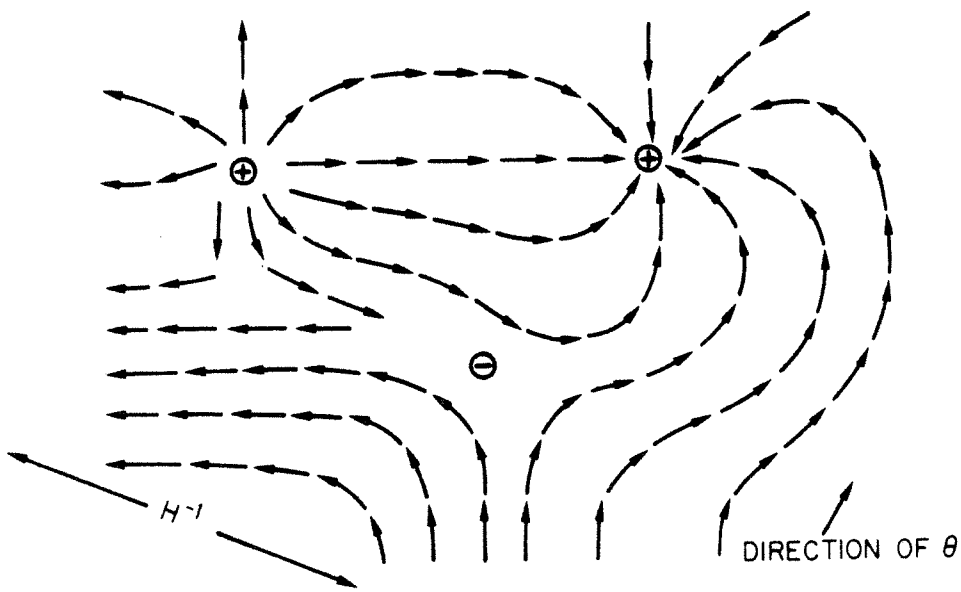


Fig. 6.5b – More accurate picture of the spatial distribution of the axion angle (indicated by arrows). Because the mapping of θ into physical space is non-trivial, θ cannot continuously smooth itself on scales $\lesssim H^{-1}$. In the process topological knots form (vortices in 2-dim, strings in 3-dim), indicated by ⊕'s and ⊖'s. On scales $\lesssim H^{-1}$ these vortices and antivortices can annihilate, always leaving of order 1 per horizon volume. In between the vortices θ is uniform on scales $< H^{-1}$.

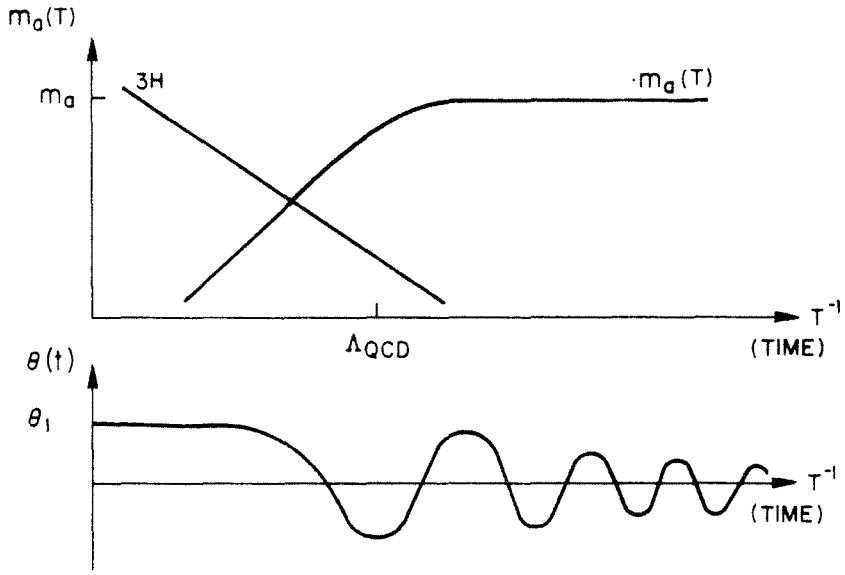


Fig. 6.6 – Evolution of the axion mass and angle θ. When the axion mass ≈ 3H, θ begins to oscillate with frequency m_a and $\theta \propto m_a^{-1/2} R^{-3/2}$.

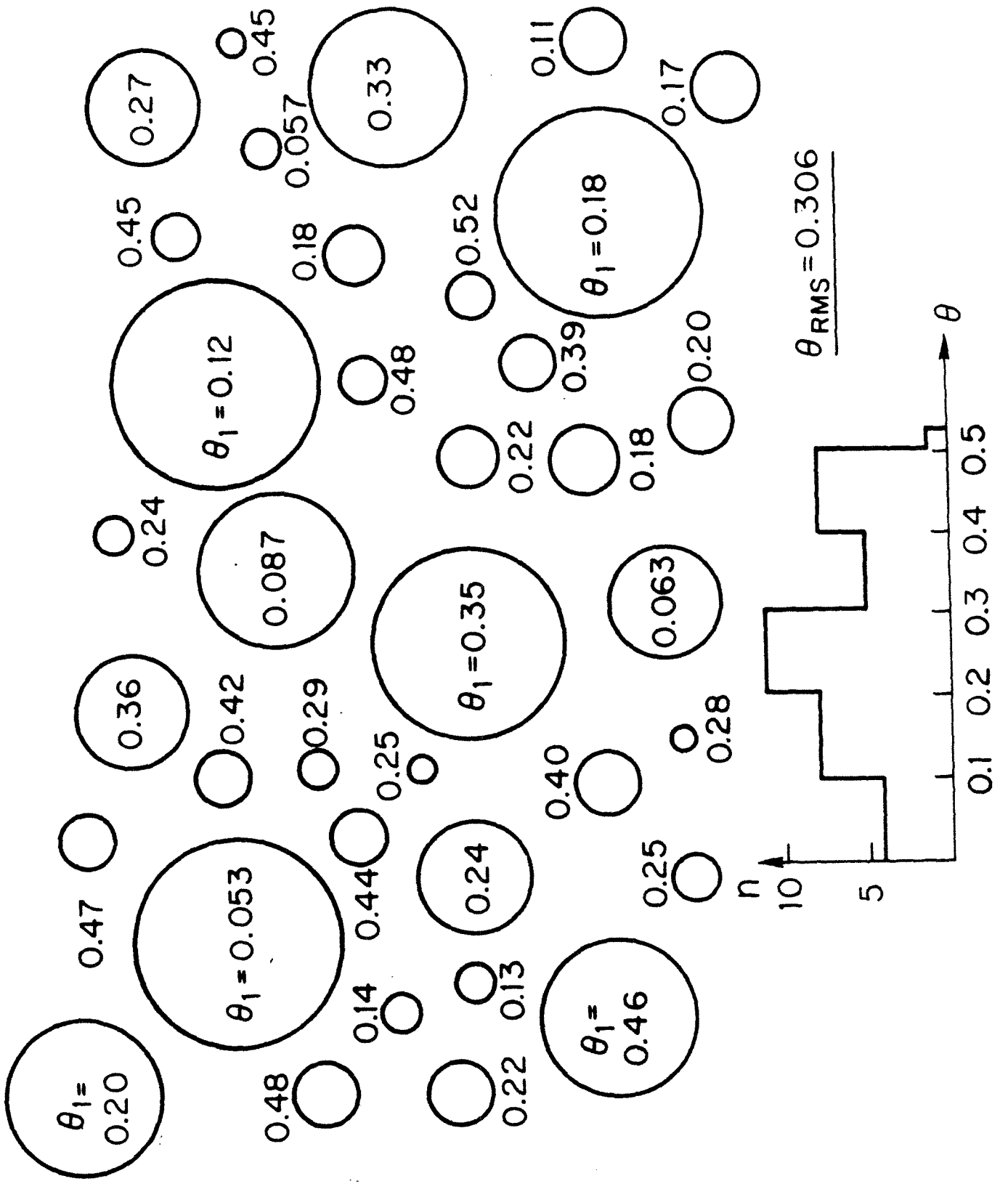


Fig. 6.7 – Distribution of θ_1 in an inflationary Universe (for N = 6). The RMS value of θ_1 (averaged over all bubbles) should be $(\pi/N)/\sqrt{3} \approx 0.302$; however, since we live in only one bubble that fact is not very relevant. [Angles were generated using the random number function of an HP-15 calculator.]

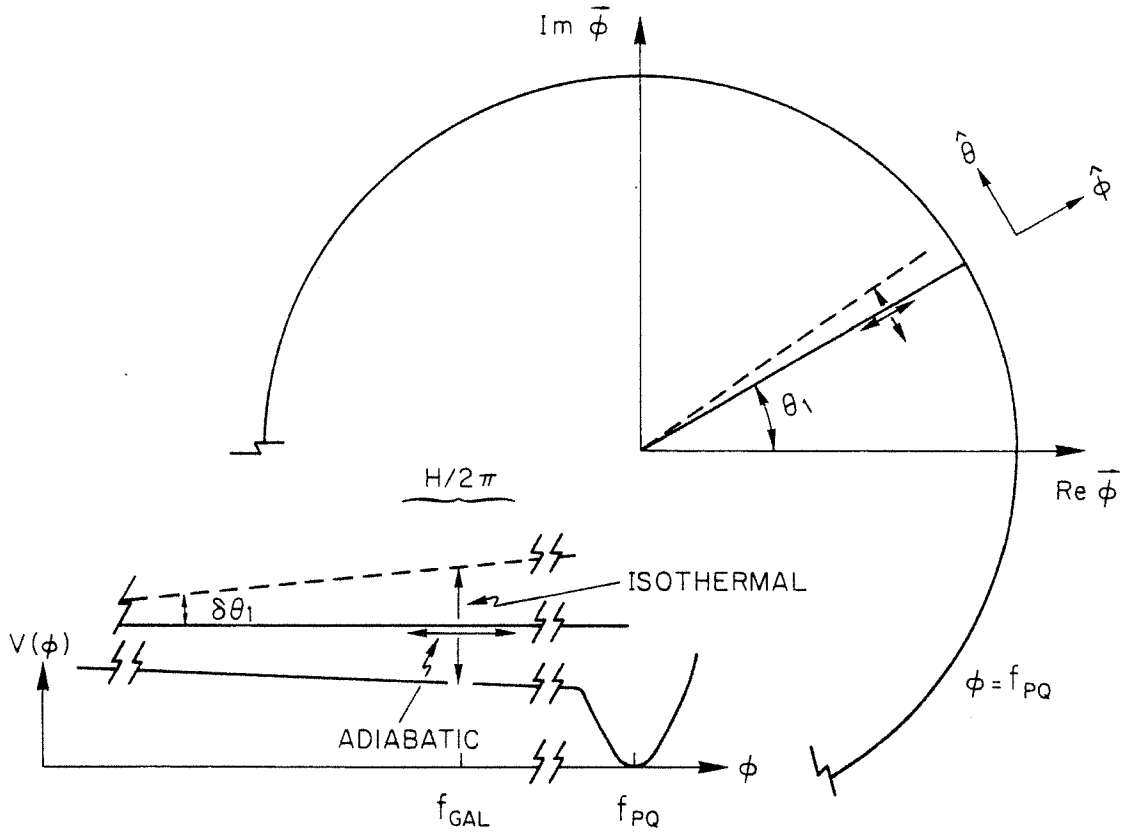


Fig. 6.8 – In Pi's model of inflation[27] PQ symmetry breaking is induced by the field responsible for inflation, $|\vec{\phi}| = \phi$. During inflation the field ϕ rolls in the direction θ_1; galactic-sized perturbations leave the horizon when $f = f_{GAL} < f_{PQ}$. DeSitter-space-produced quantum fluctuations in $\vec{\phi}$ can be decomposed into the $\hat{\phi}$ and $\hat{\theta}$ directions; each have magnitude of $O(H/2\pi)$. The fluctuations in the $\hat{\phi}$ direction lead to the usual adiabatic density perturbations, while the fluctuations in the $\hat{\theta}$ direction lead to 'isothermal' axion perturbations. The orthogonality of the two modes (adiabatic and isothermal) is particularly clear in this figure.

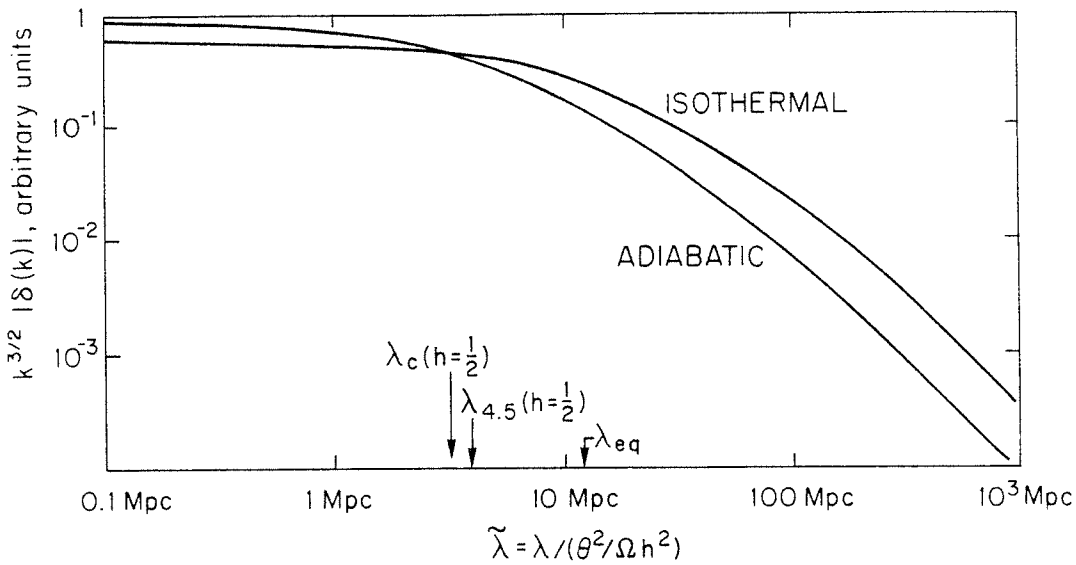


Fig. 6.9 – Spectrum of density perturbations, $\delta\rho/\rho \propto k^{3/2}|\delta(k)|$, at the epoch of matter domination in an axion-dominated Universe for both the adiabatic and isothermal modes. The isothermal spectrum is flatter on scales less than O(10 Mpc), implying that a multitude of scales should collapse almost simultaneously. When the two spectra are normalized to the same value on $\lambda \approx 7h^{-1}$Mpc, the isothermal spectrum has more power on large scales and less on small scales, implying larger microwave anisotropies on large angular scales ($\theta \gg 1°$) and later galaxy formation. For further discussion see ref. 207.

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. 1,2). 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 coauthored 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. M. B. Green and J. H. Schwarz, Nucl. Phys. B181, 502 (1981); B198, 252 (1982); B198, 441 (1982); Phys. Lett. 109B, 444 (1982); M. B. Green, J. H. Schwarz, and L. Brink, Nucl. Phys. B198, 474 (1982); J. H. Schwarz, Phys. Rep. 89, 223 (1982); M. B. Green and J. H. Schwarz, Phys. Lett. 149B, 117 (1984); B 151B, 21 (1985); Nucl. Phys. B243, 475 (1984); L. Alvarez-Gaume and E. Witten, Nucl. Phys. B243, 475 (1984); G. Chapline and N. Manton, Phys. Lett. 120B, 105 (1983); D. Gross, J. Harvey, E. Martinec, and R. Rohm, Phys. Rev. Lett. 54, 502 (1984); P. Candelas, G. Horowitz, A. Strominger, and E. Witten, Nucl. Phys. B, in press (1985); M. B. Green, Nature 314, 409 (1985); E. Kolb, D. Seckel, and M. S. Turner, Nature 314, 415 (1985); K. Huang and S. Weinberg, Phys. Rev. Lett. 25, 895 (1970).
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); C. Kounnas etal., Grand Unification With and Without Supersymmetry (World Scientific, Singapore, 1984); Q. Shafi and C. Wetterich, Phys. Lett. 129B, 387 (1983); P. Candelas and S. Weinberg, Nucl. Phys. B237, 397 (1984); R. Abbott, S. Barr, and S. Ellis, Phys. Rev. D30, 720 (1984); E. Kolb, D. Lindley, and D. Seckel, Phys. Rev. D30, 1205 (1984).

3. B. A. Peterson etal. Astrophys. J. 260, L27 (1982); D. Koo, Astron. J. 90, 418 (1985); S. Djorgovski, H. Spinrad, P. McCarthy, and M. Strauss, Astrophys. J. (Lett), in press (1985).

4. J. Peterson, P. Richards, and T. Timusk, Phys. Rev. Lett. 55, 332 (1985); G. F. Smoot etal., Astrophys. J. 291, L23 (1985); D. Meyer and M. Jura, Astrophys. J. 276, L1 (1984); D. Woody and P. Richards, Astrophys. J. 248, 18 (1981); D. Wilkinson, in Inner Space/Outer Space, eds. E. Kolb etal. (Univ. of Chicago Press, Chicago, 1985).

5. S. Weinberg, Gravitation and Cosmology (Wiley: NY, 1972); Physical Cosmology, eds. R. Balian, J. Audouze, and D. N. Schramm (North-Holland: Amsterdam, 1980).

6. R. Buta and G. deVaucouleurs, Astrophys. J. 266, 1 (1983); A. Sandage and G. A. Tammann, Astrophys. J. 256, 339 (1982), Nature 307, 326 (1984); W. D. Arnett, D. Branch, and J. Wheeler, Nature 314, 337 (1985).

7. J. Huchra, and A. Sandage and G. Tammann, in Inner Space/Outer Space, ed. E. Kolb etal. (Univ. of Chicago Press, Chicago, 1985); J. Gunn and B. Oke, Astrophys. J. 195, 255 (1975); J. Kristian, A. Sandage, and J. Westphal, Astrophys. J. 221, 383 (1978).

8. G. Gamow, Phys. Rev. 70, 572 (1946).

9. R. A. Alpher, J. W. Follin, and R. C. Herman, Phys. Rev. 92, 1347 (1953).

10. P. J. E. Peebles, Astrophys. J. 146, 542 (1966); 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. 281, 493 (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. 141B, 337 (1984).

29. R. J. Scherrer and M. S. Turner, Phys. Rev. D in press (1985).

30. S. Wolfram, Phys. Lett. 82B, 65 (1979); G. Steigman, Ann. Rev. Nucl. Part. Sci. 29, 313 (1979).

31. B. Lee and S. Weinberg, Phys. Rev. Lett. 39, 169 (1977); 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. Segrè 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, Astrophys. J. 286, 397 (1984).

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. 286, 702 (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 (1983-85).

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. 286, 216 (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, 265 (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.

133. P. J. E. Peebles, Astrophys. J. 243 L119 (1981); M. L. Wilson and J. Silk, Astrophys. J. 243, 14 (1981); 244, L37 (1981).

134. N. Vittorio and J. Silk, Astrophys. J. 285, L39 (1984); Phys. Rev. Lett. 54, 2269 (1985).

135. J. R. Bond and G. Efstathiou, Astrophys. J. 285, L44 (1984).

136. L. Abbott and M. Wise, Astrophys. J. 282, L47 (1984).

137. P. J. E. Peebles, Large Scale Structure of the Universe (Princeton Univ. Press, Princeton, 1980).

138. M. Davis and P. J. E. Peebles, Astrophys. J. 267, 465 (1983); also, see ref. 137.

139. N. Bahcall and R. Soneira, Astrophys. J. 270, 20 (1983); M. Hauser and P. J. E. Peebles, Astrophys. J. 185, 757 (1973).

140. N. Kaiser, Astrophys. J. 284, L9 (1984).

141. A. Szalay and D. N. Schramm, Nature 314, 718 (1985).

142. S. Gregory and L. Thompson, Astrophys. J. 222, 784 (1978); S. Gregory, L. Thompson, and W. Tifft, Astrophys. J. 243, 411 (1981); M. Tarenghi, W. Tifft, G. Chincarini, H. Rood, and L. Thompson, Astrophys. J. 234, 793 (1979); 235, 724 (1980).

143. R. P. Kirshner, A. Oemler, P. L. Schechter, and S. A. Shectman, Astrophys. J. 248, L57 (1981).

144. See, e.g., E. T. Vishniac, in Proceedings of Inner Space/Outer Space, eds. E. Kolb etal. (Univ. of Chicago Press, Chicago, 1985).

145. S. Faber and J. Gallagher, Ann. Rev. Astron. Astrophys. 17, 135 (1979).

146. V. Rubin etal., Astrophys. J. 225, L107 (1978); 261, 439 (1980); A. Bosma, Astron. J. 86, 1721; 86, 1825 (1981).

147. J. Bahcall, Astrophys. J. 287, 926 (1984).

148. Proceedings of IAU Symposium 117, eds, J. Kormendy and E. Turner (1985).

149. See pp. 280-284 of ref. 137.

150. M. Davis and P. J. E. Peebles, Astrophys. J. 267, 465 (1983).

151. B. Carr, J. Bond, and W. Arnett, Astrophys. J. 277, 445 (1984); C. J. Hogan, Mon. Not. r. Astron. Soc. 188, 781 (1979); A. Doroshkevich, Ya B. Zel'dovich, and I. Novikov, Sov. Astron. 11, 231 (1967).

152. J. M. Bardeen, Phys. Rev. D22, 1882 (1980).

153. W. Press and E. T. Vishniac, Astrophys. J. 239, 1 (1980).

154. J. Silk, Astrophys. J. 151, 459 (1968); also, see ref. 5, chapter 15.

155. B. J. Carr, Astrophys. J. 201, 1 (1975); 206, 8 (1976).

156. S. W. Hawking, Nature 248, 30 (1974).

157. E. R. Harrison, Phys. Rev. D1, 2726 (1970); Ya B. Zel'dovich, Mon. Not. r. Astron. Soc. 160, 1p (1972).

158. For a review of the pancake scenario, see, Doroshkevich etal., Ann. NY Acad. Sci. 375, 32 (1980); J. Bond and A. Szalay, Ann. NY Acad. Sci. 422, 82 (1984).

159. S. D. M. White, in Proceedings of Inner Space/Outer Space, eds, E. Kolb etal. (Univ. of Chicago Press, Chicago, 1985) and references therein.

160. C. Frenk, S. White, and M. Davis, Astrophys. J. 271, 417 (1983); J. Centrella and A. Melott, Nature 305, 196 (1983).

161. P. R. Shapiro, C. Struck-Marcell, and A. Melott, Astrophys. J. 275, 413 (1983).

162. J. R. Bond, J. Centrella, A. S. Szalay and J. R. Wilson, Mon. Not. r. Astron. Soc. 210, 515 (1984).

163. M. Davis, G. Efstathiou, C. Frenk, and S. White, Astrophys. J. in press (1985); also see ref. 159.

164. A. Melott etal., Phys. Rev. Lett. 51, 935 (1983).

165. G. Blumenthal, S. Faber, J. Primack, and M. Rees, Nature 311, 517 (1984).

166. M. S. Turner, G. Steigman, and L. Krauss, Phys. Rev. Lett. 52, 2090 (1984); M. S. Turner, Phys. Rev. D31, 1212 (1985).

167. P. J. E. Peebles, Astrophys. J. 284, 439 (1984).

168. D. Dicus, E. Kolb, and V. Teplitz, Astrophys. J. 221, 327 (1978); G. Gelmini, D. Schramm, and J. Valle, Phys. Lett. 146B, 311 (1984); M. Fukugita and T. Yanagida, Phys. Lett. 144B, 386 (1984); K. Olive, D. Seckel, and E. T. Vishniac, Astrophys. J., in press (1985).

169. A. Vilenkin, Phys. Rev. Lett. 53, 1016 (1984); but also see M. S. Turner, Phys. Rev. Lett. 54, 252 (1984).

170. N. Kaiser, in Proceedings of Inner Space/Outer Space, eds. E. Kolb etal. (Univ. of Chicago Press, Chicago, 1985).

171. A. Vilenkin, in Proceedings of Inner Space/Outer Space, eds. E. Kolb etal. (Univ. of Chicago Press, Chicago, 1985); Phys. Rep. in press (1985); Ya B. Zel'dovich, Mon. Not. r. Astron. Soc. 192, 663 (1980); A. Vilenkin, Phys. Rev. Lett. 46, 1169, 1496(E) (1981); A. Vilenkin and Q. Shafi, Phys. Rev. Lett. 51, 1716 (1983); N. Turok and D. Schramm, Nature 312, 598 (1984); N. Turok, Phys. Lett. 126B (1983); Nucl. Phys. B242, 520 (1984); T. Vachaspati and A. Vilenkin, Phys. Rev. D30, 2036 (1984); A. Albrecht and N. Turok, in press (1985); R. Scherrer and J. Frieman, Phys. Rev. D, in press (1985).

172. J. P. Ostriker and L. Cowie, Astrophys. J. 243, 427 (1980); C. Hogan, Mon. Not. r. Astron. Soc. 202, 1101 (1983); S. Ikeuchi, Pub. Astr. Soc. Japan 33, 211 (1981).

173. P. Sikivie, Phys. Rev. Lett. 51, 1415 (1983).

174. B. J. Carr, Nucl. Phys. B252, 81 (1985) and references therein.

175. G. 't Hooft, Phys. Rev. Lett. 37, 8 (1976); Phys. Rev. D14, 3432 (1976); R. Jackiw and C. Rabbi, Phys. Rev. Lett. 37, 172 (1976); C. G. Callan, R. Dashen, and D. Gross, Phys. Lett. 63B, 334 (1976); A. Belavin, A. Polyakov, A. Schwartz, and Y. Tyuplin, Phys. Lett. 59B, 85 (1975).

176. R. Crewther, P. DiVecchia, G. Veneziano, and E. Witten, Phys. Lett. 88B, 123 (1979); V. Baluni, Phys. Rev. D19, 2227 (1979).

177. W. Dress etal., Phys. Rev. D15, 9 (1977); also see N. Ramsey, Phys. Rep. 43c, 409 (1977).

178. R. Peccei and H. Quinn, Phys. Rev. Lett. 38, 1440 (1977); Phys. Rev. D16, 1791 (1977).

179. S. Weinberg, Phys. Rev. Lett. 40, 223 (1978); F. Wilczek, Phys. Rev. Lett. 40, 279 (1978).

180. J. Kim, Phys. Rev. Lett. 43, 103 (1979); also see, M. Shifman, A. Vainshtein, and V. Zakharov, Nucl. Phys. B166, 493 (1980).

181. M. Dine, W. Fischler, and M. Srednicki, Phys. Lett. 104B, 199 (1981).

182. P. Sikivie, Phys. Rev. Lett. 48, 1156 (1982).

183. P. Sikivie, "Axions in Cosmology", in Internationale Univeersitätswochen fur Kernphysik der Karl-Franzens-Universität XXith (Schladming, Austria 1982).

184. M. Srednicki, "Axions Coupling to Matter", UCSB preprint (1985); D. B. Kaplan, Nucl. Phys. B, in press (1985).

185. R. Stothers, Astrophys. J. 155, 935 (1969); Phys. Rev. Lett. 24, 538 (1970).

186. K. Sato and H. Sato, Prog. Theor. Phys. 54, 1564 (1975).

187. D. Dicus, E. Kolb, and V. Teplitz, Phys. Rev. D18, 1329 (1978); D22, 839 (1980).

188. M. I. Vysotsskii, Ya B. Zel'doviich, M. Khlopov, and V. Chechetkin, JETP Lett. 27, 502 (1978).

189. K. Sato, Prog. Theor. Phys. 60, 1942 (1978).

190. M. Fukugita, S. Watamura, and M. Yoshimura, Phys. Rev. Lett. 48, 1522 (1982); Phys. Rev. D26, 1840 (1982).

191. N. Iwamoto, Phys. Rev. Lett. 53, 1198 (1984).

192. J. Bahcall etal., Rev. Mod. Phys. 54, 767 (1982).

193. S. Chandrasekhar, An Introduction to Stellar Structure (Univ. of Chicago Press: Chicago, 1939); as Chandrasekhar discusses, it was Eddington who first derived the luminosity formula.

194. L. L. Krauss, J. Moody and F. Wilczek, Phys. Lett. 144B, 391 (1984).

195. E.g., see J. Kim ref. 180, and M. Shifman etal. ref. 180.

196. M. S. Turner, in preparation (1985); also see, D. Dearborn, D. Schramm, and G. Steigman, in preparation (1985).

197. S. Tsurata, Phys. Rep. 56, 237 (1979); K. Van Riper and D. Lamb, Astrophys. J. 244, L13 (1981).

198. See, e.g., D. Lamb and H. van Horn, Astrophys. J. 200, 306 (1975).

199. P. Schinder and M. S. Turner, in preparation (1985). Similar work is being done for neutron stars age-temperature diagrams.

200. Ya B. Zel'dovich, I. Y. Kobzarev, and L. B. Okun, Sov. Phys. JETP 40, 1 (1975).

201. D. Seckel, Phys. Rev. D in press (1985).

202. See, e.g., G. Lazarides and Q. Shafi Phys. Lett. 115B, 21 (1982).

203. J. Preskill, M. Wise, and F. Wilczek, Phys. Lett. 120B, 127 (1983); L. Abbott and P. Sikivie, Phys. Lett. 120B, 133 (1983); M. Dine and W. Fischler, Phys. Lett. 120B, 137 (1983).

204. M. S. Turner, Phys. Rev. D28, 1243 (1983).

205. D. Gross, R. Pisarski, and L. Yaffe, Rev. Mod. Phys. 53, 43 (1981).

206. P. Steinhardt and M. Turner, Phys. Lett. 129B, 51 (1983).

207. D. Seckel and M. S. Turner, Phys. Rev. D in press (1985).

208. A. Linde, Lebedev Phys. Inst. preprint 65 (1985).

209. M. S. Turner, F. Wilczek, and A. Zee, Phys. Lett. 125B, 35, 519(E) (1983); M. Axenides, R. Brandenberger, and M. S. Turner, Phys. Lett. 128B, 178 (1983).

210. J. R. Bond and G. Efstathiou, Stanford Univ. preprint (1985).

211. P. Sikivie, Phys. Rev. Lett. 51, 1415 (1983).

212. M. S. Turner, Fermilab peprint (1985).

213. L. L. Krauss, J. Moody, F. Wilczek, and D. Morris, ITP/UCSB preprint (1985).