Fermi National Accelerator Laboratory

FERMILAB-Conf-86/146-A
October 1986

PARTICLE PHYSICS AND COSMOLOGY

Edward W. Kolb

Theoretical Astrophysics Group
Fermi National Accelerator Laboratory
Batavia, Illinois 60510

Abstract

This series of lectures is about the role of particle physics in physical processes that occurred in the very early stages of the big bang. Of particular interest is the role of particle physics in determining the evolution of the early Universe, and the effect of particle physics on the present structure of the Universe. The use of the big bang as a laboratory for placing limits on new particle physics theories will also be discussed. Section 1 reviews the standard cosmology, including primordial nucleosynthesis. Section 2 reviews the decoupling of weakly interacting particles in the early Universe, and discusses neutrino cosmology and the resulting limits that may be placed on the mass and lifetime of massive neutrinos. Section 3 discusses the evolution of the vacuum through phase transitions in the early Universe and the formation of topological defects in the transitions. Section 4 covers recent work on the generation of the baryon asymmetry by baryon-number violating reactions in Grand Unified Theories, and mentions some recent work on baryon number violation effects at the electroweak transition. Section 5 is devoted to theories of cosmic inflation. Finally, Section 6 is a discussion of the role of extra spatial dimensions in the evolution of the early Universe.

To appear in the Proceedings of the 1986 Theoretical Advanced Studies Institute, Santa Cruz, California, July 1986.

1 THE STANDARD COSMOLOGY

Before discussing the physical processes that occurred in the first second of the big bang, the basic Friedmann-Robertson-Walker (FRW) cosmology will be discussed. First, the physical observations that lead to the assumption of a homogeneous and isotropic space will be reviewed. Some implications of the Robertson-Walker metric for the red shift and the expansion of the Universe will be derived. With a simple assumption of a perfect fluid for the stress tensor, the Friedmann equation will be integrated to express the age of the Universe in terms of the expansion rate. The implication of the conservation of entropy will be illustrated by considering the decoupling of massless neutrinos and gravitons. Finally, the primordial production of the light elements will be discussed.

1.1 *Homogeneity and Isotropy of the Universe*

The distribution of matter (at least visible matter) in the Universe seems to be homogeneous and isotropic on sufficiently large scales. One indication that the distribution becomes homogeneous on large scales is the behavior of the two-point correlation function ξ. The two-point correlation function is defined as the probability of finding an object at a distance r from another object in a volume element δV [1]

$$\delta P = n\delta V[1 + \xi(r)], \tag{1.1}$$

where n is the number density of objects in the sample. For a uniform Poisson distribution $\xi = 0$. The magnitude of ξ is thus an indication of the departure of the distribution of galaxies from homogeneity. Several catalogs give a galaxy-galaxy correlation function consistent with a simple power law form of $\xi(r)$ given by $\xi(r) = (r/5h^{-1}\text{Mpc})^{-1.8}$, where h represents the uncertainty in the determination of Hubble's constant ($H_0 = 100\ h$ km sec^{-1}Mpc^{-1}), and 1 Mpc$=10^6$pc$=10^6 \times 3.1 \times 10^{18}$cm. At distances larger than $5h^{-1}$Mpc, the correlation function drops below unity, suggesting that a uniform distribution becomes a good approximation. Of course this one piece of evidence does not prove that the distribution is uniform. In the past few years there has been a growing amount of evidence that there is a rich structure in clusters, filament, bubbles, etc., on scales in excess of $5h^{-1}$Mpc, which serves to illustrate the fact that the two-point correlation function does not provide complete information on clustering. Nevertheless, the decrease in the galaxy-galaxy correlation function is evidence that on large scales the distribution of matter is uniform.

The microwave background radiation (MBR) is evidence that the Universe is spatially isotropic. The observations of the spectral nature of the MBR is shown in Figure 1.[2] Observations of the temperature of the MBR are consistent with a blackbody of $T = 2.72$K. Figure 1a is the MBR spectral flux as measured by Woody

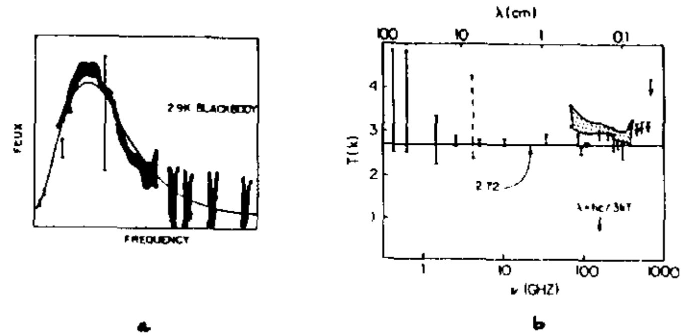

Figure 1: The spectrum of the microwave background radiation

and Richards (shaded area), along with microwave and optical results. Deviations from the Planck spectrum may be real, or they may represent systematic errors such as background subtractions. The equivalent blackbody temperature over a large frequency range is shown in Figure 1b (the dashed line is the original measurement of Penzias and Wilson). The best fit to all the measurements is a temperature of 2.72K.

A remarkable feature of the MBR is its high degree of isotropy. This isotropy is best illustrated by considering temperature differences in the background radiation as a function of the angular scale of the separation. The results of the observations are summarized in Figure 2.[3] But for a dipole moment to the radiation of $\Delta T/T \simeq 10^{-3}$, the MBR is isotropic on scales as small as 10". The dipole moment can be understood as the peculiar velocity of the earth with respect to the MBR. The isotropy on smaller scales indicates that when the MBR last scattered the distribution of matter was uniform.

It is easy to make an estimate of the distance to the last scattering surface. If we assume that the mean free path of the photons is determined by Compton scattering off free electrons, $\gamma + e \rightarrow \gamma + e$, the mean free path is given by $\lambda = (n_e\sigma_T)^{-1}$, where n_e is the number density of electrons and σ_T is the Thomson cross section ($\sigma_T = 6.65 \times 10^{-25}\text{cm}^2$). The number density of free electrons can be expressed in terms of the average number of electrons per nucleon, Y_e, and the electron ionization fraction, X_e, as $n_e = X_eY_en_N$, where n_N is the nucleon density. The nucleon density is usually determined by its contribution to the mass density $\rho_N = m_Nn_N$, where ρ_N is the mass density of nucleons, and m_N is the nucleon mass. The nucleon mass density in turn is usually expressed in terms of its ratio to a "critical density", given by

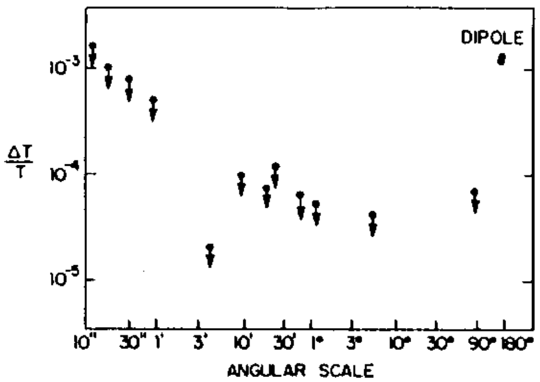

Figure 2: Limits on the anisotropy of the microwave background radiation

$$\rho_C = \frac{3H_0^2}{8\pi G} = 1.88 \times 10^{-29} h^2 \text{ g cm}^{-3}. \tag{1.2}$$

The fraction of the critical density in any species i is defined as Ω_i

$$\Omega_i = \rho_i/\rho_C. \tag{1.3}$$

The electron density is then $n_e = 1.12 \times 10^{-5} X_e Y_e \Omega_N h^2 \text{cm}^{-3}$. All observational evidence gives $\Omega_N h^2 \leq 1$. If we assume $X_e Y_e = 1$ (the maximal value), and ignore for the moment any change in n_e due to the expansion of the Universe, then the mean free path (distance to the last scattering surface) is $\lambda \geq 10^{29}$ cm. This calculation will be improved by taking into account the expansion of the Universe and by a better calculation of X_e. However the estimate made above serves to illustrate the main point: the surface of last scattering of the MBR is at a great distance and the distribution of matter on this large scale was isotropic when the MBR last scattered.

Finally, there is a strong theoretical prejudice for only considering spaces that are spatially homogeneous and isotropic. There will be only one undetermined function in the metric for a homogeneous and isotropic space. This allows for a real confrontation with the meager observational evidence. The philosophy taken here is to assume the simplest model and confront the data. A successful confrontation will result in the remarkable achievement of a simple model for the evolution of the Universe. A failure of this simple model would signal a breakdown either in the cosmological principle (that space is homogeneous and isotropic) or in the field equations of gravity.

1.2 The Robertson-Walker Metric

The metric for a space with homogeneous and isotropic spatial sections is given by [4]

$$ds^2 = dt^2 - R^2(t)\left\{\frac{dr^2}{1-kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\right\} \tag{1.4}$$

where (t, r, θ, ϕ) are coordinates, $R(t)$ is the cosmic scale factor, and $k = +1, -1$, or 0 for spaces of constant positive curvature, constant negative curvature, or zero spatial curvature. The coordinate r in Eq. 1.4 is dimensionless and scaled to $R(t)$, i.e., r ranges from 0 to 1.

The meaning of the cosmic scale factor $R(t)$ can be illustrated most easily by considering the space of constant positive curvature ($k = +1$), and by embedding the three-space into a four-dimensional euclidean space with coordinates x_1, x_2, x_3, and x_4. Under a coordinate transformation to "four-dimensional" spherical coordinates (R, χ, θ, ϕ) related to the four-dimensional cartesian coordinates by $x_1 = R\sin\chi\sin\theta\cos\phi$, $x_2 = R\sin\chi\sin\theta\sin\phi$, $x_3 = R\sin\chi\cos\theta$, $x_4 = R\cos\chi$, the Robertson-Walker metric takes the form

$$ds^2 = dt^2 - R^2(t)\left[d\chi^2 + \sin^2\chi(\sin^2\theta d\phi^2 + d\theta^2)\right]. \tag{1.5}$$

The above form explicitly illustrates the metric for $k = +1$ is that of a three-sphere, S^3, with radius given by $R(t)$. The volume of the S^3 is given by

$$V = \int_0^{2\pi}\int_0^{\pi}\int_0^{\pi} R^3 \sin^2\chi \sin\theta d\chi d\theta d\phi = 2\pi^2 R^3. \tag{1.6}$$

The radius of the S^3 today is larger than the Hubble radius, $R_H = H_0^{-1} = 9.24 \times 10^{27} h^{-1}$cm. The space has finite volume, but has no boundaries. For the $k = -1$ choice (space of constant negative curvature), the space is the 3-hyperboloid, Q^3, and the metric can be written in the form of Eq. 1.5 with $\sin\chi \to \sinh\chi$. The volume of the Q^3 is, of course, infinite since the range of χ is $-\infty$ to $+\infty$. For the $k = 0$ choice, the spatial metric is that of R^3, i.e., spatially flat. It also has infinite volume.

It should be noted that the assumption of homogeneity and isotropy only implies that the spatial metric is *locally* S^3, Q^3, or R^3, and the space can have different global properties. For instance, for the spatially flat case the global properties of the space might be that of the three-torus, T^3, rather than R^3. Such non-trivial topologies may be relevant in light of recent work on theories with extra dimensions, such as superstrings. In many such theories the internal space is compact, but has topological defects such as holes, handles, etc. If the internal space is not simply connetted, it is likely that the external space is also not simply connected, and the global properties of the space might be much different than the simple S^3, Q^3, or R^3.

Before considering the dynamics of expansion, it is possible to understand the effect of expansion on the red shift of light from distant galaxies. Suppose a photon is emitted from a source at coordinate $r = r_1$ at time t_1 and arrives at a detector at time t_0 at coordinate $r = 0$ (for simplicity consider propagation along $d\phi^2 = d\theta^2 = 0$). The massless photon will travel on a geodesic ($ds^2 = 0$), and the coordinate and time will be related by

$$\int_{t_1}^{t_0} \frac{dt}{R(t)} = \int_0^{r_1} \frac{dr}{(1-kr^2)^{1/2}} \equiv f(r_1). \tag{1.7}$$

A photon emitted at a time $t_1+\delta t_1$ will arrive at the detector at a time $t_0+\delta t_0$. The equation of motion will be the same as Eq. 1.7 with $t_1 \rightarrow t_1 + \delta t_1$ and $t_0 \rightarrow t_0 + \delta t_0$. Since $f(r_1)$ is constant (the source is fixed in the coordinate system)

$$\int_{t_1}^{t_0} \frac{dt}{R(t)} = \int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{dt}{R(t)}. \tag{1.8}$$

By simple rearrangement of the limits of integration

$$\int_{t_1}^{t_1+\delta t_1} \frac{dt}{R(t)} = \int_{t_0}^{t_0+\delta t_0} \frac{dt}{R(t)}. \tag{1.9}$$

If δt is sufficiently small, then R(t) will be constant over the integration time of Eq. 1.9, and

$$\frac{\delta t_1}{R(t_1)} = \frac{\delta t_0}{R(t_0)}. \tag{1.10}$$

If we consider δt_1 (δt_0) to be the time of successive waves of the emitted (detected) light, then δt_1 (δt_0) is the wavelength of the emitted (detected) light, and

$$\frac{\lambda_1}{\lambda_0} = \frac{R(t_1)}{R(t_0)}. \tag{1.11}$$

The red shift is usually defined in terms of z, by

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1}. \tag{1.12}$$

In terms of $R(t)$,

$$1 + z = \frac{R(t_0)}{R(t_1)}. \tag{1.13}$$

Any increase (decrease) in $R(t)$ leads to a red shift (blue shift) of the light from distant sources. The fact that today a red shift of light from distant sources is observed implies that the Universe is expanding.

Hubble's law may be found directly form the FRW metric without knowing anything about the dynamics of the expansion. Hubble's law relates the "luminosity distance" d_L to the red shift z. If a source has an absolute luminosity L (the energy per time produced by the source), the luminosity distance is defined in terms of the measured flux F (the energy per time per area measured by a detector) by

$$F = \frac{L}{4\pi d_L^2}. \tag{1.14}$$

If a source at co-moving coordinate $r = r_1$ emits light at time t_1, and a detector at co-moving coordinate $r = 0$ detects the light at $t = t_0$, conservation of energy ($T^{\mu\nu}{}_{;\nu} = 0$) implies

$$d_L = R^2(t_0)\frac{r_1}{R(t_1)} = R(t_0)r_1(1+z). \tag{1.15}$$

The dependence upon r_1 must be removed. The first step is to expand $R(t)$ in a power series

$$\frac{R(t)}{R(t_0)} = 1 + \frac{\dot{R}(t_0)}{R(t_0)}(t-t_0) - \frac{1}{2}\left(-\frac{\ddot{R}(t_0)}{\dot{R}^2(t_0)}R(t_0)\right)\frac{\dot{R}^2(t_0)}{R^2(t_0)}(t-t_0)^2 + \ldots, \tag{1.16}$$

or remembering $R(t_0)/R(t) = 1 + z$, Eq. 1.16 can be inverted for small $H_0(t_0 - t)$ to give (this analysis follows Weinberg [6])

$$z = H_0(t_0 - t) + \left(1 + \frac{q_0}{2}\right) H_0^2 (t_0 - t)^2 + \ldots \tag{1.17}$$

where

$$H_0 \equiv \frac{\dot{R}(t_0)}{R(t_0)} \tag{1.18}$$

$$q_0 \equiv \frac{-\ddot{R}(t_0)}{\dot{R}^2(t_0)} R(t_0). \tag{1.19}$$

Eq. 1.17 can be inverted to yield

$$(t_0 - t) = H_0^{-1}\left[z - \left(1 + \frac{q_0}{2}\right) z^2 + \ldots\right]. \tag{1.20}$$

It is also possible to expand $f(r_1)$ of Eq. 1.7 in a power series

$$\begin{aligned} f(r_1) &= r_1 + \frac{r_1^3}{6} + \dots \qquad (k=+1) \\ &= r_1 \qquad\qquad (k=0) \\ &= r_1 - \frac{r_1^3}{6} + \dots \qquad (k=-1). \end{aligned} \tag{1.21}$$

Using the expansion of Eq. 1.16 in Eq. 1.7 gives

$$r_1 = R^{-1}(t_0)\left[(t_0 - t_1) + \frac{1}{2}H_0(t_0 - t_1)^2 + \dots\right]. \tag{1.22}$$

Using the expression for $(t_0 - t_1)$ gives

$$r_1 = R(t_0)^{-1}H_0^{-1}\left[z - \frac{1}{2}(1+q_0)z^2 + \dots\right]. \tag{1.23}$$

Substituting this expression into Eq. 1.15 finally yields Hubble's law

$$H_0 d_L = z + \frac{1}{2}(1-q_0)z^2 + \dots \tag{1.24}$$

Note that the linear relation between d_L and z fails for $z \to 1$ if $q_0 \neq 1$. Note that Hubble's law was derived without explicitly solving the dynamics of the Einstein equations.

1.3 The Friedmann Equation

Before solving the Einstein equations for the evolution of the scale factor $R(t)$, it is necessary to make some assumptions about the dread right hand side of the Einstein equations. To be consistent with the symmetries of the metric, the stress tensor $T_{\mu\nu}$ should be diagonal, and by isotropy the non-zero spatial parts of the metric should be equal. The simplest realization of such a stress tensor is that of a perfect fluid characterized by an energy density ρ and a pressure p

$$T^{\mu}_{\ \nu} = \mathrm{diag}(\rho, -p, -p, -p). \tag{1.25}$$

The conservation of energy equations ($T^{\mu\nu}_{\ \ ;\nu} = 0$) gives

$$d(\rho R^3) = -p\,d(R^3). \tag{1.26}$$

For simple equations of state Eq. 1.26 gives

$$\begin{aligned} \text{RADIATION} \quad & (p = \tfrac{1}{3}\rho) \implies \rho \propto R^{-4} \\ \text{MATTER} \quad & (p = 0) \implies \rho \propto R^{-3} \\ \text{VACUUM ENERGY} \quad & (p = -\rho) \implies \rho \propto R^{0} \end{aligned} \tag{1.27}$$

The "early" Universe will be radiation dominated, and in the absence of vacuum energy, the "late" Universe will be matter dominated.

The dynamical equation that describes the evolution of $R(t)$ is found from the Einstein field equations. There are two independent equations from the field equations, one of them can be taken as Eq. 1.26, and the other is the Friedmann equation

$$\left(\frac{\dot{R}}{R}\right)^2 + \frac{k}{R^2} = \frac{8\pi G}{3}\rho. \tag{1.28}$$

With the definitions of ρ_C and Ω in Eqs. 1.2 and 1.3

$$\frac{k}{H^2R^2} = \Omega - 1. \tag{1.29}$$

Since $H^2R^2 \geq 0$, there is a correspondence between the sign of k, and the sign of $\Omega - 1$

$$\begin{aligned} k = +1 &\implies \Omega \geq 1 \qquad \text{CLOSED} \\ k = 0 &\implies \Omega = 1 \qquad \text{FLAT} \\ k = -1 &\implies \Omega \leq 1 \qquad \text{OPEN}. \end{aligned} \tag{1.30}$$

A different combination of the Einstein equations yields

$$3\ddot{R} = -4\pi G(\rho + 3p)R. \tag{1.31}$$

Since today $\dot{R} \geq 0$, if $\rho + 3p$ was always positive, then at some finite time in the past R must have been equal to zero. This time is defined as $t = 0$. At $R = 0$ there is a singularity, extrapolation past the singularity is not possible.

The Friedmann equation may be integrated to give the age of the Universe in terms of the expansion rate. Let sub-0 denote the value of quantities today. The energy density scales as $\rho/\rho_0 = (R/R_0)^{-3}$ for a matter-dominated (MD) Universe, and $\rho/\rho_0 = (R/R_0)^{-4}$ for a radiation-dominated (RD) Universe. The Friedmann equation becomes

$$\begin{aligned} \left(\frac{\dot{R}}{R_0}\right)^2 + \frac{k}{R_0^2} &= \frac{8\pi G}{3}\rho_0\frac{R_0}{R} \qquad \text{(MD)} \\ &= \frac{8\pi G}{3}\rho_0\left(\frac{R_0}{R}\right)^2 \quad \text{(RD)}. \end{aligned} \tag{1.32}$$

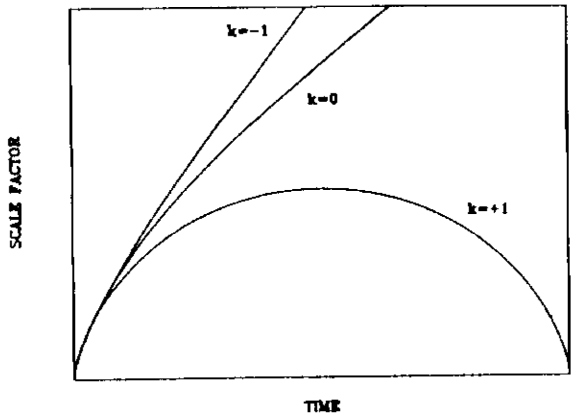

Figure 3: The evolution of R for closed ($k = +1$), open ($k = -1$), and flat ($k = 0$) cosmologies

Using $k/R_0^2 = H_0^2(\Omega_0 - 1)$, the time as a function of $R_0/R = 1 + z$ is given by

$$t = H_0^{-1}\int_0^{(1+z)^{-1}} \frac{dx}{[1 - \Omega_0 + \Omega_0 x^{-1}]^{1/2}} \quad \text{(MD)}$$
$$= H_0^{-1}\int_0^{(1+z)^{-1}} \frac{dx}{[1 - \Omega_0 + \Omega_0 x^{-2}]^{1/2}} \quad \text{(RD)}. \tag{1.33}$$

The age of the Universe is obviously a decreasing function of Ω. In the limit $\Omega \to 0, t \to H_0^{-1}(1+z)^{-1} = 9.78 \times 10^9(1+z)^{-1}h^{-1}$ years for both (RD) and (MD). If $\Omega = 1$, then

$$t = \frac{2}{3}(1+z)^{-3/2}H_0^{-1} \quad \text{(MD)}$$
$$= \frac{1}{2}(1+z)^{-2}H_0^{-1} \quad \text{(RD)}. \tag{1.34}$$

The present age for a matter-dominated $\Omega = 1$ Universe is $6.5 \times 10^9 h^{-1}$ years. This age is consistent with the lower end of estimates of the age of the Universe on the basis of stellar evolution and nucleocosmochronology if h is not too much larger than 1/2.

Eq. 1.33 can be integrated to give R as a function of time. For $k = +1$, R increases to a maximum, then decreases to zero. For $k = 0$ or $k = -1$, R increases without limit. The evolution of R is shown in Fig.3.

For many of the problems of interest in cosmology, it is only necessary to know the age of the "early" Universe, when it was radiation dominated and the curvature term in the Friedmann equation ($k/R^2 \propto R^{-2}$) was negligible compared to the energy density term ($(8\pi G/3)\rho \propto R^{-4}$ for radiation). The region of validity of these criteria will be quantified shortly. The second condition implies $\Omega = 1$ is a good approximation in the early Universe, and from Eq. 1.34

$$t = \frac{1}{2}H^{-1}$$
$$H^2 = \frac{8\pi G}{3}\rho_R, \tag{1.35}$$

where ρ_R is the energy density in radiation. The energy density and number density of a particle of mass m at temperature T is given by (for zero chemical potential)

$$\rho = \frac{g}{2\pi^2}\int_0^\infty \frac{(E^2 - m^2)^{1/2}}{1 \pm exp(E/T)}E^2 dE$$
$$n = \frac{g}{2\pi^2}\int_0^\infty \frac{(E^2 - m^2)^{1/2}}{1 \pm exp(E/T)}E dE, \tag{1.36}$$

where g is the number of spin states and the + (−) obtains for Fermi (Bose) statistics. In the relativistic limit ($T \gg m$)

$$\rho = \begin{cases} (\pi^2/30)gT^4 & \text{(BOSE)} \\ (\pi^2/30)(7/8)gT^4 & \text{(FERMI)} \end{cases}$$
$$n = \begin{cases} (\zeta(3)/\pi^2)gT^3 & \text{(BOSE)} \\ (\zeta(3)/\pi^2)(3/4)gT^3 & \text{(FERMI)}. \end{cases} \tag{1.37}$$

In the non-relativistic limit the energy density and the number density is the same for Bose and Fermi statistics

$$\rho = mn$$
$$n = g\left(\frac{mT}{2\pi}\right)^{3/2}\exp(-m/T). \tag{1.38}$$

The total radiation energy density can be expressed in terms of the photon temperature T as

$$\rho_R = \frac{\pi^2}{30}g_* T^4, \tag{1.39}$$

where g_* counts the effective massless degrees of freedom

$$g_* = \sum_{i=bosons} g_i\left(\frac{T_i}{T}\right)^4 + \frac{7}{8}\sum_{i=fermions} g_i\left(\frac{T_i}{T}\right)^4. \tag{1.40}$$

The relative factor of 7/8 accounts for the difference in Fermi and Bose statistics.

In terms of g_* and the Planck mass $m_{Pl} \equiv G^{-1/2} = 1.2 \times 10^{19}$GeV, the age and expansion rate of the early Universe is given by ($T_{\rm MeV} = T/1$ MeV)

$$\begin{aligned} t &= 0.3 g_*^{-1/2} \frac{m_{Pl}}{T^2} \simeq 1 \ {\rm sec} T_{\rm MeV}^{-2} \\ H &= 1.66 g_*^{1/2} \frac{T^2}{m_{Pl}}. \end{aligned} \tag{1.41}$$

1.4 Entropy

Throughout most of the history of the Universe (in particular the early Universe) the reaction rates of particles in the thermal bath, Γ_{int}, were much greater than the expansion rate, H, and local thermal equilibrium (LTE) should have been maintained. In this case the entropy per comoving volume element remains constant. The second law of thermodynamics states that

$$TdS = d(\rho V) + pdV \tag{1.42}$$

and the energy density and pressure are related by

$$\frac{\partial^2 S}{\partial T \partial V} = \frac{\partial^2 S}{\partial V \partial T} \tag{1.43}$$

which implies

$$T\frac{dp}{dT} = \rho + p. \tag{1.44}$$

Using the conservation of energy equation written in the form

$$\frac{d}{dt}\left[R^3(\rho + p)\right] = R^3\frac{dp}{dt} \tag{1.45}$$

in Eq. 1.44 results in the conserved quantity

$$\frac{d}{dt}\left\{\frac{R^3(\rho + p)}{T}\right\} = 0. \tag{1.46}$$

This conserved quantity is simply the entropy S.

It is useful to define an entropy density s

$$\begin{aligned} s &\equiv \frac{\rho + p}{T} \\ &= \frac{2\pi^2}{45} g_*' T^3, \end{aligned} \tag{1.47}$$

where

$$g_*' = \sum_{i={\rm bosons}} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{i={\rm fermions}} g_i \left(\frac{T_i}{T}\right)^3. \tag{1.48}$$

For most of the history of the Universe all particles had a common temperature, and g_*' can be replaced by g_*.

The conservation of S implies that $s \propto R^{-3}$, and that $g_* T^3 R^3$ is constant in the expansion. The factor of g_* enters because as the temperature of the Universe drops below the mass of a particle species, that species will disappear via annihilations (assuming it remains in equilibrium) and the entropy that was present in that species will be shared among the particles remaining in equilibrium. If g_* changes, T is not proportional to R^{-1}.

Massless particles that are decoupled from the heat bath will not share in the entropy released as the temperature drops below the mass threshold of a species, but rather the temperature of a massless decoupled species scales as $T \propto R^{-1}$. As an example consider a massless particle initially in LTE which decouples at time t_D, temperature T_D, and scale factor R_D. The phase-space distribution at decoupling is given by the equilibrium distribution

$$f(E, t_D) = [\exp(E/T_D) \pm 1]^{-1}. \tag{1.49}$$

After decoupling the energy of the massless particle is red-shifted by the expansion of the Universe $E(R) = E(R_D)(R_D/R)$. So at some time after decoupling the phase space density of a particle of energy E will be the phase space density of a particle of energy $E(R/R_D)$ at decoupling (since the phase space density is conserved)

$$\begin{aligned} f(E,t) &= f(E\frac{R}{R_D}, t_D) = \left[\exp\left(\frac{ER}{R_D T_D}\right) \pm 1\right] \\ &= [\exp(E/T) \pm 1]. \end{aligned} \tag{1.50}$$

Thus the distribution for massless particles is self-similar in expansion, with the temperature red-shifting as R^{-1}

$$T = T_D \frac{R_D}{R} \propto R^{-1} \qquad \text{DECOUPLED}, \tag{1.51}$$

not $\propto R^{-1} g_*^{-1/3}$ as for particles remaining in equilibrium.

The effect of decoupling is best illustrated by considering the decoupling of massless neutrinos. In the early Universe neutrinos are kept in equilibrium via reactions of the sort $\bar{\nu}\nu \leftrightarrow e^+e^- + \cdots$. The cross section is weak, given by $\sigma \simeq G_F^2 T^2$, where G_F is the Fermi constant. The number density of the massless particles is $n \simeq T^3$, so the interaction rate is

$$\Gamma_{int} = n\sigma|v| \simeq G_F^2 T^5. \tag{1.52}$$

The ratio of the interaction rate to the expansion rate is

$$\frac{\Gamma_{int}}{H} \simeq \frac{G_F^2 T^5}{T^2/m_{Pl}} \simeq 1\left(\frac{T}{1\text{MeV}}\right)^3. \tag{1.53}$$

At temperatures above 1 MeV, the interaction rate is greater than the expansion rate and the neutrinos are in equilibrium. At temperatures below 1 MeV the interaction rate is less than the expansion rate and neutrino interactions are too weak to keep them in equilibrium. Below 1 MeV the neutrino temperature T_ν scales as R^{-1}. Subsequent to neutrino decoupling the temperature drops below threshold for $e^\pm$ production and the entropy in the $e^\pm$ is transferred to the photons but not to the decoupled neutrinos. For $T \geq m_e$, g_* includes γ (g = 2) and $e^\pm$ (g = 4), for an effective $g_* = 11/2$. For $T \leq m_e$, only the photons are in equilibrium for an effective $g_* = 2$. Since $g_*(RT)^3$ is constant, RT is increased by the third-root of the ratio of g_* before $e^\pm$ annihilation (11/2) to g_* after $e^\pm$ annihilation (2). For the photons RT is increased by a factor of $(11/4)^{1/3}$, due to $e^\pm$ annihilation, while RT for the neutrinos is unaffected. Therefore today the ratio of T_γ and T_ν should be

$$\frac{T_\gamma}{T_\nu} = \left(\frac{11}{4}\right)^{1/3} = 1.4 \tag{1.54}$$

which gives $T_\nu = 1.9$ K. The addition of three two-component massless neutrinos at the above temperature results in a value of g_* today of

$$g_*(today) = 2 + \frac{7}{8} \times 2 \times 3 \times \left(\frac{4}{11}\right)^{4/3} = 3.36. \tag{1.55}$$

This results in a present energy density in massless particles and entropy density of

$$\begin{aligned} \rho_R &= \frac{\pi^2}{30} g_* T^4 = 7.56 \times 10^{-34}\text{g cm}^{-3} \\ s &= \frac{2\pi^2}{45} g'_* T^3 \simeq 2800\text{cm}^{-3}. \end{aligned} \tag{1.56}$$

Another example is the decoupling of gravitons. For particles with only gravitational strength interactions, the interaction rate should be $\Gamma_{int} = n\sigma|v| \simeq G^2T^5 \simeq T^5/m_{PL}^4$. This will become less than the expansion rate, $H \simeq T^2/m_{Pl}$, at temperatures less than m_{Pl}. If gravitons decouple at the Planck time, the contribution to g_* from particles we know [1] was 90.75. Therefore today the number density of gravitons should be smaller than the number density of photons by a factor of roughly (2/90.75).

[1] The particles we "know" are taken to be the three generations of quarks and leptons, the gauge particles of $SU_3 \times SU_2 \times U_1$, and the Higgs doublet of the Weinberg-Salam model.

Before concluding this section it is useful to mention three parameters that describe the Universe. The first parameter is the time of the decoupling of radiation and matter. Using the fact that the electron number density scales as $(R/R_0)^3$, the temperature scales as R_0/R, and the equilibrium ionization fraction of electrons found from the Saha equation is [6]

$$\frac{X_e^2}{1 - X_e} = \frac{(2\pi m_e T)^{3/2}}{(2\pi)^3 n_e} \exp(-B/T), \tag{1.57}$$

where m_e is the electron mass and B is the ionization potential of hydrogen, the red shift at decoupling (also referred to as recombination) is $1 + z_{rec} \simeq 1500$. This yields a temperature and time of decoupling of

$$\begin{aligned} T_{rec} &= T_0(1 + z_{rec}) = 4100\text{K} = 0.35\text{eV} \\ t_{rec} &= t_0(1 + z_{rec})^{-3/2} = 1.1 \times 10^5\ h^{-1} \text{ years} = 3.5 \times 10^{12}\ \text{h}^{-1} \text{ sec.} \end{aligned} \tag{1.58}$$

If we define ρ_M as the total energy density in "matter" (i.e., in non-relativistic particles), then today $\rho_M = 1.88 \times 10^{-29}\Omega_M\ h^2$ g cm^3, where Ω_M is the fraction of the critical density contributed by ρ_M. Using Eq. 1.56, and the fact that $\rho_R/\rho_M = R_0/R = 1 + z$, then the red shift, time, and temperature of equal matter and radiation energy densities is given by

$$\begin{aligned} 1 + z_{eq} &= 2.5 \times 10^4\ h^2\ \Omega_M \\ T_{eq} &= T_0(1 + z_{eq}) = 5.8\ h^2\ \Omega_M \text{ eV} \\ t_{eq} &= t_0(1 + z_{eq})^{-3/2} = 1.6 \times 10^3\ h^{-4}\ \Omega_M^{-3/2}\text{years}. \end{aligned} \tag{1.59}$$

The baryon number density is defined as

$$n_B = n_b - n_{\bar{b}} = 1.12 \times 10^{-5}\ \Omega_N\ h^2 \text{ cm}^{-3}, \tag{1.60}$$

where n_b ($n_{\bar{b}}$) is the baryon (antibaryon) number density, and it has been assumed that today $n_{\bar{b}} = 0$. The *baryon number* is defined as

$$B = \frac{n_B}{s} = 4 \times 10^{-9}\ \Omega_N\ h^2. \tag{1.61}$$

As long as the baryon number is conserved in particle interactions and entropy is conserved in the expansion of the Universe, B will remain constant.

1.5 Primordial Nucleosynthesis

The first step in understanding primordial nucleosynthesis is the application of nuclear statistical equilibrium (NSE). In kinetic equilibrium, the number density of a nucleus of mass number A is given by

^AZ	B_A
^{2}H	2.2 MeV
^{3}H	8.5 MeV
^{3}He	7.7 MeV
^{4}He	28.2 MeV

Table 1: The binding energies of some light nuclei.

$$n_A = g_A \left(\frac{m_A T}{2\pi}\right)^{3/2} \exp\left(\frac{\mu_A - m_A}{T}\right). \tag{1.62}$$

In chemical equilibrium the chemical potential of a nucleus with mass number A and charge Z is related to the neutron and proton chemical potentials by

$$\mu_A = Z\mu_p + (A - Z)\mu_n. \tag{1.63}$$

In kinetic equilibrium the neutron and proton number densities (n_n and n_p) are given by expressions like Eq. 1.62 with the neutron and proton chemical potentials and masses replacing μ_A and m_A. Therefore, in chemical equilibrium, [2]

$$\begin{aligned}\exp[\mu_A/T] &= \exp[(Z\mu_p + (A-Z)\mu_n)/T] \\ &= n_p^Z n_n^{A-Z}\left(\frac{2\pi}{m_N T}\right)^{3A/2} 2^{-A}\exp[(Zm_p + (A-Z)m_n)/T]\end{aligned} \tag{1.64}$$

With the definition of the binding energy

$$B_A \equiv Zm_p + (A - Z)m_n - m_A, \tag{1.65}$$

Eq. 1.62 becomes

$$n_A = g_A A^{3/2} 2^{-A}\left(\frac{2\pi}{m_N T}\right)^{3(A-1)/2} n_p^Z n_n^{A-Z}\exp(B_A/T). \tag{1.66}$$

A list of binding energies are given in Table 1.

Rather than the number density, it is useful to consider the mass fraction, which is defined as the fraction of the total baryon mass in any particular species

$$\begin{aligned} X_A &\equiv \frac{n_A A}{n_N} \\ \sum_i X_i &= 1. \end{aligned} \tag{1.67}$$

[2] In the pre-exponential factor the difference of the neutron and proton masses will not be important, and m_N will denote the nucleon mass.

In NSE the mass fraction of species A is given by

$$X_A = g_A A^{5/2} 2^{-A}\left[\frac{8}{\pi^3}\zeta^2(3)\frac{T}{m_N}\right]^{3(A-1)/2}\eta^{A-1}X_p^Z X_n^{A-Z}\exp(B_A/T), \tag{1.68}$$

where

$$\eta \equiv \frac{n_N}{n_\gamma} = 2.8\times 10^{-8}\,\Omega_N\, h^2. \tag{1.69}$$

The fact that the Universe is "hot" ($n_\gamma \gg n_N$; η is small) is crucial in primordial nucleosynthesis. After considering the "initial conditions" primordial nucleosynthesis will be considered in three steps.

- *Initial Conditions* ($T \gg 1$ MeV, $t \ll 1$ sec.): The initial conditions for primordial nucleosynthesis are no heavy elements (only protons and neutrons) and equal numbers of protons and neutrons. The lack of heavy elements is a result of the small value of η. Consider a simple system of ^{4}He, neutrons and protons with $n_n = n_p$. In NSE the ^{4}He mass fraction (X_4) is given by ($X_n = X_p = (1 - X_4)/2$):

$$\begin{aligned} X_4 &= 4^{5/2}2^{-4}\left[\frac{8}{\pi^3}\zeta^2(3)\right]^{9/2}\left(\frac{T}{m_N}\right)^{9/2}\eta^3\frac{1}{16}(1-X_4)^4\exp(B_4/T) \\ &= 6.2\times 10^{-44}T_{\rm MeV}^{9/2}\eta_9^3\exp(28.2/T_{\rm MeV})(1-X_4)^4, \end{aligned} \tag{1.70}$$

where $\eta_9 \equiv 10^9\eta$. A graph of X_4 as a function of T is shown in in Fig. 4. Until $T_{\rm MeV}$ drops below 0.3, the abundance of helium is small because η is small. [3]

The initial condition that $n_n = n_p$ is a result of the fact that the weak reactions that interconvert neutrons and protons are much faster than the expansion rate at this time. The six reaction that interconvert neutrons and protons are

$$n \longleftrightarrow pe\nu, \qquad \nu n \longleftrightarrow pe, \qquad en \longleftrightarrow p\nu. \tag{1.71}$$

The rate (per nucleon) of the above reactions are found by integrating the square of the amplitude for a given process, weighted by the available phase-space densities of particles (other than the initial nucleon), while enforcing four-momentum conservation. As an example, the rate for $pe \to \nu n$ is given by [7,8,5]

$$\lambda_{pe\to\nu n} = \int f_e(E_e)[1 - f_\nu(E_\nu)]|M|^2_{pe\to\nu n}\frac{\delta^4(p+e-\nu-n)}{(2\pi)^5}\frac{d^3p_e}{2E_e}\frac{d^3p_\nu}{2E_\nu}\frac{d^3p_n}{2E_n} \tag{1.72}$$

For all the above reactions

[3] The small abundance of ^{4}He at high temperature is sometimes incorrectly blamed on a deuterium "bottleneck." But the abundance of ^{4}He is small in NSE at high temperature, which has nothing to do with the binding energy of deuterium. The low binding energy of deuterium will be important only somewhat later.

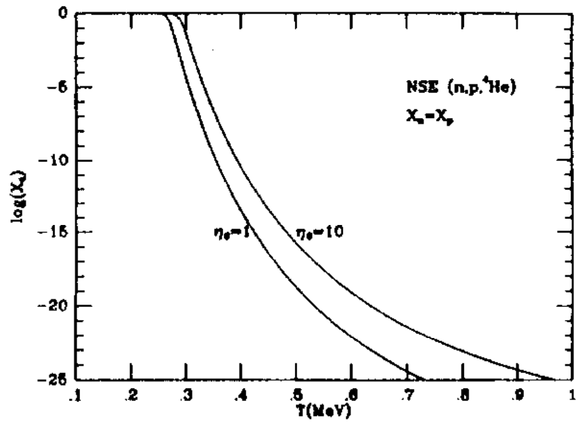

Figure 4: The NSE abundances of the n, p, ${}^4He$ system as a function of temperature.

$$|M|^2 \propto G_F^2(1+3g_A^2). \tag{1.73}$$

Since this is the same matrix element for neutron decay, it is convenient to write it in terms of the neutron lifetime τ.

The neutron-proton mass difference, $Q = m_n - m_p = 1.293$ MeV, and the electron mass determine the limits of integration in the rates. In terms of the dimensionless quantities $q = Q/m_e$, $\epsilon = E_e/m_e$, $z = m_e/T$, $z_\nu = m_e/T_\nu$,

$$\lambda_{pe\to\nu n} = (\tau\lambda_0)^{-1}\int_q^\infty d\epsilon \frac{\epsilon(\epsilon-q)^2(\epsilon^2-1)^{1/2}}{[1+\exp(\epsilon z)][1+\exp((q-\epsilon)z_\nu)]}, \tag{1.74}$$

where λ_0 simply represents a numerical factor from the phase space integral for neutron decay. In the high temperature and low temperature limits [8]

$$\lambda_{pe\to\nu n} \longrightarrow \begin{cases} 0 & T \ll Q,\ m_e \\ \frac{7}{60}\pi(1+3g_A^2)G_F^2T^5 \simeq G_F^2T^5 & T \gg Q,\ m_e. \end{cases} \tag{1.75}$$

In the high temperature limit, the weak rates are much greater than H, and the neutron proton ratio should obtain the equilibrium value

$$\frac{n}{p} \equiv \frac{n_n}{n_p} = \left(\frac{n_n}{n_p}\right)_{eq} = \exp(-Q/T). \tag{1.76}$$

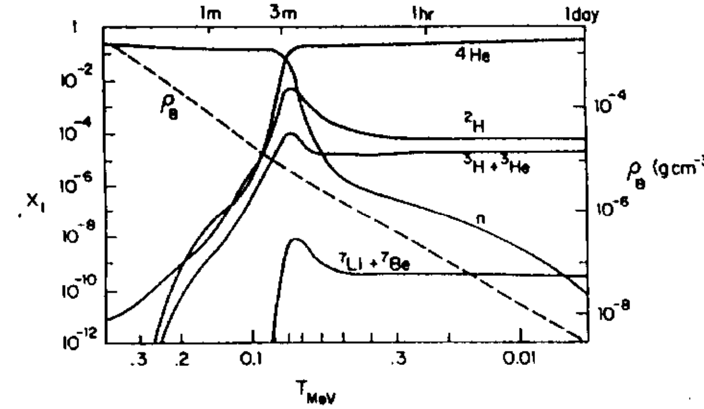

Figure 5: The development of primordial nucleosynthesis. The dashed line is the baryon density, and the solid lines are the mass fractions. The model is for 3 neutrinos, $\eta = 3\times10^9$, and $\tau_{1/2} = 10.6$ minutes.

using $(n/p)^0 = 1/7$.

At this point primordial nucleosynthesis ends when the nuclear rates become less than H. There are three reasons for the termination of primordial nucleosynthesis. The Coulomb barriers at low temperatures ($T \le 8\times10^8$ K) suppress the rates. The baryon density is low (see Fig. 5) which suppresses three-body initial states. Finally, there are no $A = 5$ or $A = 8$ stable nuclei to act as intermediate steps.

The development of primordial nucleosynthesis (the first day in the life of the Universe) is shown in Fig. 5. All numerical results presented here were obtained with aid of a computer program courtesy of Robert Wagoner [8].

Before discussing the agreement with the inferred values of the primordial abundances, it is useful to consider the sensitivity of the final abundances on the input parameters.

• τ: The weak rates are proportional to $(1+3g_A^2)$, which is usually determined by measurement of the neutron lifetime. Since $\lambda \propto \tau^{-1}$, an increase in τ results in a decrease of λ, which means the weak rates freeze out earlier. From Eq. 1.78, an increase in T_F leads to and increase of n/p, hence more ^{4}He.

• g_*: Since $H \propto g_*^{1/2}$, an increase in g_* leads to a faster expansion rate, which results in earlier freeze out and higher ^{4}He. This is used to study the effect of

Therefore, at high temperature ($T \gg Q$), $n_n = n_p$.

• *Step 1* ($t = 10^{-2}$sec., T = 10 MeV): For step 1, the energy density of the Universe is radiation dominated, and the $e^{\pm}$, γ, and 3 neutrinos give $g_* = 10.75$. The weak rates are much larger than H, so $n_n = n_p$, and $T_\nu = T_\gamma$. The heavy elements are in NSE, but they have very small abundances due to the fact that η is small. For example, if $\eta = 10^{-9}$

$$X_4 \simeq \eta^3 \left(\frac{T}{m_N}\right)^{9/2} \exp(B_4/T) = 2 \times 10^{-38}$$

$$X_2 \simeq \eta \left(\frac{T}{m_N}\right)^{3/2} \exp(B_2/T) = 1 \times 10^{-12}. \qquad (1.77)$$

• *Step 2* ($t \simeq 1$ sec., T = $T_F \simeq 1$ MeV): At about this time the weak rates freeze out (become smaller than H). When the weak rates freeze out the neutron-proton ratio is given by the equilibrium value,

$$\left(\frac{n}{p}\right)_{freeze-out} = \exp(-Q/T_F) \simeq \frac{1}{6}. \qquad (1.78)$$

after freeze-out the neutron-proton ratio is given by (τ is the neutron lifetime)

$$\left(\frac{n}{p}\right) \simeq \frac{1}{6}\exp(-t/\tau). \qquad (1.79)$$

After the neutrinos decouple, the annihilation of the $e^{\pm}$ increases the photon temperature relative to the neutrino temperature by a factor of $(11/4)^{1/3}$.

• *Step 3* ($t = 1 - 3$ minutes, $T = 0.3 - 0.1$ MeV): At this time the effective g_* is 3.36 for 3 neutrinos, and (n/p) has decreased from 1/6 to 1/7 due to neutron decay. At this time the NSE value of ^{4}He starts to rapidly approach one. However in the big bang the actual amount of ^{4}He cannot keep up with the NSE values since there are only trace amounts of ^{2}H, ^{3}H, and ^{3}He present, and these elements are intermediate steps in the fusion of ^{4}He. This two minute delay in the onset of ^{4}He obtaining the NSE value allows some more of the neutrons to decay. Once ^{4}He does obtain its NSE value, almost all of the available neutrons are processed into ^{4}He as it has by far the largest binding energy per nucleon of the light elements. If the densities of neutron and protons at this time are denoted by n_n^0 and n_p^0 and all the neutrons are processed to ^{4}He, then the final amount of ^{4}He is given by $n_4 = n_n^0/2$. In terms of the mass fraction of ^{4}He and the neutron-proton fraction

$$X_4 = \frac{4n_4}{n_n^0 + n_p^0} = \frac{2\,(n/p)^0}{(n/p)^0 + 1} \simeq 0.25, \qquad (1.80)$$

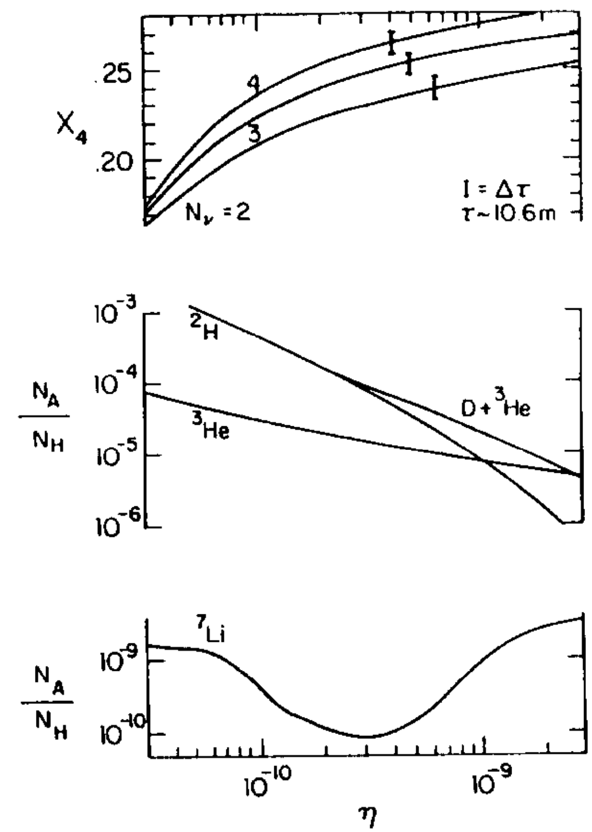

Figure 6: The primordial abundances of the light elements as a function of η

additional light neutrinos, since

$$\begin{aligned} g_* &= 2 + \frac{7}{8}(4 + 2 \times N_\nu) + \cdots \\ &= 10.75 + \cdots \qquad N_\nu = 3 \\ &= 12.50 + \cdots \qquad N_\nu = 4. \end{aligned} \tag{1.81}$$

• η: In NSE the abundances of the elements increases with η. An increase in η allows ^{4}He to be produced earlier when there were more neutrons, hence more ^{4}He will be produced. NSE will be obtained for a longer time, so there will be less ^{2}H and ^{3}He.

• Nuclear reaction rates: For the reactions of interest in primordial nucleosynthesis the uncertainties in the nuclear reaction cross sections are not important.

The primordial mass fraction of ^{4}He and the abundances of ^{2}H≡D, ^{3}He, and ^{7}Li relative to hydrogen are given in Fig. 6.[9] The effect of the uncertainty in the neutron lifetime and the number of light neutrinos on X_4 is indicated. The determination of the primordial abundances from present observations is very difficult. In order of decreasing reliability [9]

$$\begin{aligned} X_4 &= 0.22 - 0.26 \\ \frac{D}{H} &\geq 10^{-5} \\ 8 \times 10^{-5} &\geq \frac{D + {}^3He}{H} \\ 2 \times 10^{-10} &\geq \frac{{}^7Li}{H} \geq 1 \times 10^{-10}. \end{aligned} \tag{1.82}$$

All these abundances are consistent if $N_\nu \leq 4$. This consistency is the strongest evidence that the standard Friedmann-Robertson-Walker cosmology can be extrapolated back as far as 1 second after the bang when the temperature was 1 MeV. Having successfully gone back 15 billion years, the next sections will take us back the final second.

1.6 References

1. P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, 1980).
2. P. L. Richards, in *Inner Space/Outer Space* (University of Chicago Press, Chicago, 1986), E. W. Kolb, M. S. Turner, D. Lindley, K. Olive, and D. Seckel, eds.
3. D. T. Wilkinson, in *Inner Space/Outer Space* (University of Chicago Press, Chicago, 1986), E. W. Kolb, M. S. Turner, D. Lindley, K. Olive, and D. Seckel, eds.
4. L. D. Landau and E. M. Lifshitz, *Classical Theory of Fields* (Pergamon Press, Oxford, 1975).
5. S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
6. P. J. E. Peebles, *Physical Cosmology* (Princeton University Press, Princeton, 1971).
7. R. A. Alpher, J. W. Follin, and R. C. Herman, Phys. Rev. **92**, 1347 (1953).
8. R. V. Wagoner, W. A. Fowler, F. Hoyle, Astrophys. J. **148**, 3 (1967).
9. J.Yang, M. S. Turner, G. Steigman, D. N. Schramm, and K. A. Olive, Astrophys. J. **281**, 493 (1984).

2 NEUTRINO COSMOLOGY

Before considering the survival of any particular particle, the general framework for considering the decoupling of particles will be developed in detail. The general results will be used to study in detail the decoupling of massless and massive neutrinos. [4] This study will allow a combination of cosmological and astrophysical limits to be placed on the properties of massive neutrinos. Neutrinos are only an example of the application of cosmology and astrophysics to limit the properties of elementary particles. Some brief comments on other particles will be made, and the possibility of detecting these fossil particles will be discussed.

2.1 Freeze Out

Consider a particle ψ of mass M that is stable, and present in equal numbers with its antiparticle $\bar{\psi}$. Let f denote the phase-space density of ψ. The evolution of f is determined by the Boltzmann equation, which can be written in the form

$$\hat{\mathbf{L}}[f] = \mathbf{C}[f], \tag{2.1}$$

where $\mathbf{C}$ is the collision operator and $\hat{\mathbf{L}}$ is the Liouville operator. The collision operator depends on interactions at a point, so it should be independent of the geometry. The Liouville operator, however, depends on derivatives, and will be sensitive to the geometry. The non-relativistic form of the Liouville operator is

$$\hat{\mathbf{L}} = \frac{\partial}{\partial t} + \vec{v}\cdot\vec{\nabla}_r + \frac{\vec{F}}{M}\cdot\vec{\nabla}_v. \tag{2.2}$$

The general relativistic generalization of this operator is

$$\hat{\mathbf{L}} = p^\alpha \frac{\partial}{\partial x^\alpha} - \Gamma^\alpha_{\beta\gamma} p^\beta p^\gamma \frac{\partial}{\partial p^\alpha}. \tag{2.3}$$

For the FRW metric, $\hat{\mathbf{L}}$ simplifies considerably, and the Boltzmann equation becomes

$$E\frac{\partial f}{\partial t} - \frac{\dot{R}}{R}|\vec{p}|^2\frac{\partial f}{\partial E} = \mathbf{C}[f(E,t)]. \tag{2.4}$$

Using the definition of the number density in terms of the phase space density

$$n = (2\pi)^{-3}\int f(E)d^3p, \tag{2.5}$$

[4]The terms quenching, decoupling, and freeze out will be used interchangeably.

after integration by parts the Boltzmann equation may finally be expressed in the form

$$\frac{dn}{dt} + 3\frac{\dot{R}}{R}n = (2\pi^3)^{-1}\int \mathbf{C}[f]\frac{d^3p}{E}. \tag{2.6}$$

Consider the contribution to the collision integral from the process $\psi\bar{\psi} \to \gamma\gamma$, where "$\gamma$' will represent a generic massless particle. The right hand side of Eq. 2.6 is [6]

$$(2\pi)^{-3}\int \mathbf{C}[f]\frac{d^3p}{E}$$

$$= \Lambda^{\psi\bar{\psi}}_{\gamma_1\gamma_2}(2\pi)^{-6}\left[f_\gamma(E_{\gamma_1})f_\gamma(E_{\gamma_2})|M|^2_{\gamma\gamma\to\psi\bar{\psi}} - f_\psi(E_\psi)f_\psi(E_{\bar{\psi}})|M|^2_{\bar{\psi}\psi\to\gamma\gamma}\right], \tag{2.7}$$

where

$$\Lambda^{ab\cdots}_{ij\cdots} \equiv \int \frac{d^3p_i}{2E_i}\frac{d^3p_j}{2E_j}\cdots\frac{d^3p_a}{2E_a}\frac{d^3p_b}{2E_b}\delta^4(p_i + p_j + \cdots - p_a - p_b - \cdots). \tag{2.8}$$

Eq. 2.7 can be written in an extremely useful form with the help of two assumptions. The first assumption is T (or CP) invariance of the matrix element

$$|M|^2_{\psi\bar{\psi}\to\gamma\gamma} = |M|^2_{\gamma\gamma\to\psi\bar{\psi}}. \tag{2.9}$$

The second assumption is that the massless particles (γ's) are in equilibrium $f_\gamma(E) = \exp(-E/T)$. This assumption, together with the conservation of energy ($E_{\gamma_1} + E_{\gamma_2} = E_\psi + E_{\bar{\psi}}$) from δ^4, allows the γ phase space density to be expressed in terms of the ψ phase space density

$$\begin{aligned} f_\gamma(E_{\gamma_1})f_\gamma(E_{\gamma_2}) &= \exp(-E_{\gamma_1}/T)\exp(-E_{\gamma_2}/T) \\ &= \exp(-E_\psi/T)\exp(-E_{\bar{\psi}}/T) \\ &= f^{eq}_\psi(E_\psi)f^{eq}_\psi(E_{\bar{\psi}}), \end{aligned} \tag{2.10}$$

where

$$f^{eq}_\psi(E) \equiv \exp(-E/T). \tag{2.11}$$

Since the phase space densities depend only upon E_ψ and $E_{\bar{\psi}}$, the phase space integrals over p_{γ_1} and p_{γ_2} can be done, yielding $\langle|v|\sigma_{\psi\bar{\psi}\to\gamma\gamma}\rangle$. The integrals over p_ψ and $p_{\bar{\psi}}$ yield either n_ψ or n^{eq}_ψ. The Boltzmann equation then becomes

[6]In the absence of Bose condensation or Fermi degeneracy, Maxwell-Boltzmann statistics should be a good approximation, and will be employed unless otherwise indicated.

$$\dot{n_\psi} + 3\frac{\dot{R}}{R}n_\psi = \left[\left(n_\psi^{eq}\right)^2 - n_\psi^2\right]\langle|v|\sigma_{\psi\bar{\psi}\to\gamma\gamma}\rangle. \tag{2.12}$$

It is straightforward to include other processes in the same manner. Including all possible final states for $\psi\bar{\psi}$ annihilation results in

$$\dot{n_\psi} + 3\frac{\dot{R}}{R}n_\psi = \left[\left(n_\psi^{eq}\right)^2 - n_\psi^2\right]\langle|v|\sigma_A\rangle, \tag{2.13}$$

where σ_A is the total $\psi\bar{\psi}$ annihilation cross section. In most cases the non-relativistic form of σ_A will have a simple dependence on the temperature, and the temperature dependence can be parameterized as

$$\langle|v|\sigma_A\rangle = \sigma_0\left(\frac{M}{T}\right)^{-n}. \tag{2.14}$$

The terms in Eq. 2.13 have a simple explanation. The term proportional to $-n_\psi^2$ represents the decrease of ψ due to $\psi\bar{\psi}$ annihilation. The term proportional to $(n_\psi^{eq})^2$ represents the increase of ψ due to collisions of the γ's in the thermal bath. The Boltzmann factor in n_ψ^{eq} in the NR limit is reflects the fact that at $T \ll M$, it becomes exponentially unlikely that a collision of two γ's will have sufficient energy to create a $\psi\bar{\psi}$ pair. Note that if the creation and annihilation rates are fast enough (greater than H), n_ψ will be driven to its equilibrium value.

It is convenient to express Eq. 2.13 in terms of the dimensionless quantities $Y_\psi \equiv n_\psi/s$, $x \equiv M/T$

$$\frac{dY}{dx} = -0.26g_*^{1/2}m_{Pl}Mx^{-2}\langle|v|\sigma_A\rangle\left(Y^2 - Y_{eq}^2\right). \tag{2.15}$$

In the non-relativistic (NR) limit ($x \gg 3$) and in the extreme relativistic (ER) limit ($x \ll 3$) Y_{eq} has the limiting forms

$$Y_{eq} = \begin{cases} (\kappa 45 g_\psi/2\pi^4 g_*) & x \ll 3 \\ (45g_\psi/2\pi^4 g_*)(\pi/8)^{1/2}x^{3/2}\exp(-x) & x \gg 3, \end{cases} \tag{2.16}$$

where g_ψ is the number of spin degrees of freedom for ψ, and κ is $\zeta(3)$ $(3\zeta(3)/4)$ for Bose (Fermi) statistics.

In general there are no closed-form solutions of Eq. 2.15. However some approximate solutions may be obtained quite easily. Y will track the equilibrium value until freeze out. Freeze out will occur when the density of ψ becomes so small that the rate of ψ annihilation ($\Gamma_{annihilation} = n_\psi\langle|v|\sigma_A\rangle$) becomes less than H, and the temperature becomes so small that the rate of ψ creation ($\Gamma_{creation} = n_\psi^{eq}\langle|v|\sigma_A\rangle$) becomes less than H. After this time ψ's are neither created nor destroyed, and

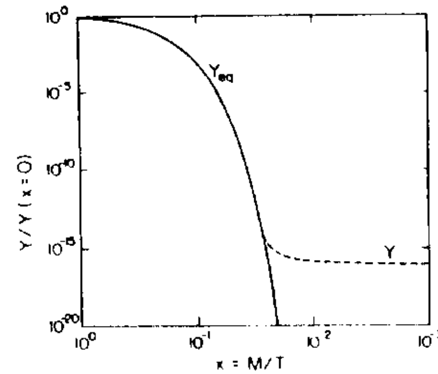

Figure 7: The evolution and freeze out of a massive particle

$n_\psi \propto R^{-3}$. If the entropy is also conserved, $s \propto R^{-3}$, so the value of Y after freeze-out, Y_∞, will be constant in the expansion. The typical behavior of Y is shown in Fig. 7.[1]

The value of x at freeze out, x_f, can be found by equating $\Gamma_{creation}$ and H [6]

$$x_f \simeq \ln\left\{(n+1)\frac{45}{2\pi^4}\left(\frac{\pi}{8}\right)^{1/2}0.26g_\psi g_*^{-1}m_{Pl}M\sigma_0\right\}$$
$$-(n+\frac{1}{2})\ln\left[\ln\left\{(n+1)\frac{45}{2\pi^4}\left(\frac{\pi}{8}\right)^{1/2}0.26g_\psi g_*^{-1}m_{Pl}M\sigma_0\right\}\right] + \cdots \tag{2.17}$$

Substituting x_f into the NR expression for Y_{eq} gives [7]

$$Y_\infty \simeq \frac{(n+1)g_\psi}{0.26g_*^{1/2}m_{Pl}M\sigma_0}x_f^{n+1}. \tag{2.18}$$

As an example, consider the annihilation of nucleons and antinucleons in the early Universe. For $N\bar{N}$, the annihilation cross section can be parameterized as $\sigma_0 = cm_\pi^{-2}$, where m_π is the pion mass and c is a constant. Since the annihilation

[6] In the following, it will be assumed that decoupling occurs when the ψ is non-relativistic.

[7] There is of course an ambiguity in the definition of x_f. Freeze out could be defined when Γ = any number $\times$ H. The constant has been chosen to give the best fit to the numerical results.[2]

is exothermic, $n = -1/2$. If $g_* = 15$ is used, $x_f \simeq 40 + \ln c$ and $Y_\infty = 6 \times 10^{-19} c^{-1}$. Today, $Y_\infty = 4 \times 10^{-9} \Omega_B h^2$, so the above result is wrong by a factor of 10^{10}. The calculation is correct, and in fact agrees quite well with the numerical result. The discrepancy between the prediction and the observation implies that the nucleon system does not satisfy one of the assumptions. If there is an asymmetry in the number of nucleons and the number of antinucleons the above formalism is incorrect. The lesson is that there must have been such an excess of nucleons relative to antinucleons before $N\bar{N}$ annihilation. The annihilation shut off when the antinucleons were used up, and the nucleons we observe today were the ones that could not find antinucleons to annihilate.

2.2 Light Stable Neutrinos ($M \leq 1$ MeV)

Light neutrinos decoupled when the temperature was about 1 MeV. After neutrino decoupling, $e^\pm$ annihilation increased the temperature of the photons relative to the neutrinos by a factor of $(11/4)^{1/3}$. The present number density of each light neutrino species should be

$$\begin{aligned} n_\nu &= \frac{3}{4}\frac{\zeta(3)}{\pi^2} g_\nu T_\nu^3 = \frac{3}{4}\frac{g_\nu}{2}\frac{4}{11} n_\gamma \\ &= 109 \frac{g_\nu}{2}\ \mathrm{cm}^3. \end{aligned} \tag{2.19}$$

If the light neutrino has a mass greater than $T_\nu \simeq 1.9$ K $\simeq 1.6 \times 10^{-4}$ eV, then the present energy density of the neutrino would be $\rho_\nu = M n_\nu$, which would contribute to Ω an amount

$$\Omega_\nu h^2 = 1.03 \times 10^{-2} \left(\frac{M}{\mathrm{eV}}\right)\left(\frac{g_\nu}{2}\right). \tag{2.20}$$

Since there is an observational limit on the maximum value of Ω, there is a maximum value of M

$$M \leq 96.8\mathrm{eV}\left(\frac{2}{g_\nu}\right)\left(\Omega_\nu \mathrm{h}^2\right)_{\max}. \tag{2.21}$$

This limit is known as the Cowsik - McClelland [3] bound. [§]

[§] In the original paper of Cowsik and McClelland, they assumed two four-component neutrinos with the same mass ($g_\nu = 4$), $h = 1/2$, and $T_\nu = 2.7$ K, which gives $M \leq 8$ eV for their assumed upper limit of $\Omega = 3.8$.

2.3 Heavy Stable Neutrinos ($M_Z \geq M \geq 1$ MeV)

If neutrinos are NR at decoupling, Eq. 2.18 gives the final abundance. Neutrino annihilation proceeds through Z exchange to final states $i\bar{i}$, where $i = \nu_L$, e, μ, τ, u, d, s, $\cdots$ Here ν_L denotes a light neutrino. For $T \leq M \leq M_Z$, the annihilation cross section depends upon whether ν is a Dirac or a Majorana particle

$$\left(|v|\sigma_A\right)_{Dirac} = \frac{G_F^2 M^2}{2\pi}\sum_i (1 - z_i^2)^{1/2}\left[(C_{V_i}^2 + C_{V_i}^2) + \frac{1}{2} z_i^2 (C_{V_i}^2 - C_{A_i}^2)\right] \tag{2.22}$$

$$\left(|v|\sigma_A\right)_{Majorana} = \frac{G_F^2 M^2}{2\pi}\sum_i (1 - z_i^2)^{1/2}\left[(C_{V_i}^2 + C_{V_i}^2) 8\beta_i^2/3 + C_{A_i}^2 2 z_i^2\right], \tag{2.23}$$

where $z_i = m_i/M$, β is the relative velocity, and C_V and C_A are given in terms of the weak isospin, electric charge, and the Weinberg angle by $C_A = j_3$, $C_V = j_3 - 2q\sin^2\theta_W$.

In the Dirac case, $\sigma_0 = cG_F^2 M^2/2\pi$, $n = 0, g_* = 60$, and c is a constant ≈ 5. The value of x_f and Y_∞ is

$$\begin{aligned} x_f &\simeq 25 + 3\ln M_{\mathrm{GeV}} + \ln c \\ Y_\infty &\simeq 10^{-8} g_\nu M_{\mathrm{GeV}}^{-3}\left(1 + \frac{3}{25}\ln M_{\mathrm{GeV}} + \frac{1}{25}\ln c\right), \end{aligned} \tag{2.24}$$

where $M_{\mathrm{GeV}} \equiv (M/1\ \mathrm{GeV})$. If $g_\nu = 2$, the present neutrino energy density, and the contribution to Ω would be

$$\begin{aligned} \rho_\nu &\simeq 5.7 \times 10^4 M_{\mathrm{GeV}}^{-2}\left(1 + \frac{3}{25}\ln M_{\mathrm{GeV}}\right)\ \mathrm{eV\ cm^{-3}} \\ \Omega_\nu h^2 &\simeq 5.4 M_{\mathrm{GeV}}^{-2}\left(1 + \frac{3}{25}\ln M_{\mathrm{GeV}}\right). \end{aligned} \tag{2.25}$$

For the Majorana case, $\Omega_\nu h^2$ is similar to Eq. 2.25 with $5.4 \to 18$. The limit on M that results from Eq. 2.25 is usually referred to as the "Lee - Weinberg" bound (although it was discovered simultaneously by several people).[4]

The contribution to $\Omega_\nu h^2$ as a function of M is shown in Fig. 8 for Dirac and Majorana neutrinos. The exact limits on M depend on the value of $(\Omega_\nu h^2)_{max}$. If $h = 1/2$ and $\Omega_\nu \leq 0.9$, then $M \leq 21.8$ eV for light neutrinos, or $M \geq 5$ GeV (9 GeV) for heavy Dirac (Majorana) neutrinos.

The above limits have been found assuming $M \leq M_Z$ and the chemical potential of the neutrinos are zero. It is trivial to find the limit if $M \geq M_Z$, or if the chemical potential is large.

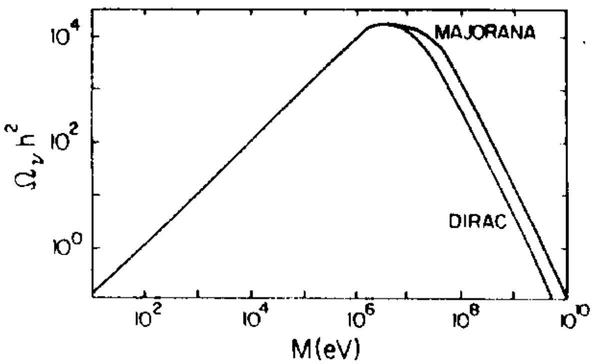

Figure 8: The contribution to $\Omega_\nu h^2$ from a neutrino of mass M

2.4 Heavy Unstable Neutrinos

The above limits on the mass of neutrinos can be evaded if the neutrinos are unstable.[8] The energy density of massive particles decreases in the expansion as R^{-3}, while the energy density of massless particles decreases as R^{-4}, which leads to $\rho_M/\rho_R = (1+z)$. If a massive neutrino decayed at a redshift z_D into *massless* particles,[9] the contribution of the massless decay products to Ωh^2, denoted $\Omega_D h^2$, would be smaller than the contribution to Ωh^2 if the neutrino had not decayed (denoted as $\Omega_\nu h^2$) by a factor of $1+z_D$. In terms of the neutrino mass M

$$\begin{aligned}\Omega_D h^2 &= \Omega_\nu h^2(1+z_D)^{-1}\\ &= \begin{cases} 1.03\times 10^{-2} M_{eV}(1+z_D)^{-1} & \text{light neutrinos}\\ 5.4 M_{GeV}^{-2}(1+z_D)^{-1} & \text{heavy Dirac neutrinos}\\ 18 M_{GeV}^{-2}(1+z_D)^{-1} & \text{heavy Majorana neutrinos.}\end{cases}\end{aligned} \qquad (2.26)$$

The requirement that $\Omega_D h^2$ is less than some maximum value places a limit on $1+z_D$, of (again picking $h = 1/2$, $\Omega_D \le 0.9$)

[9] It is assumed that the decay is instantaneous. Integrating over an exponential decay probability does not significantly change the results.

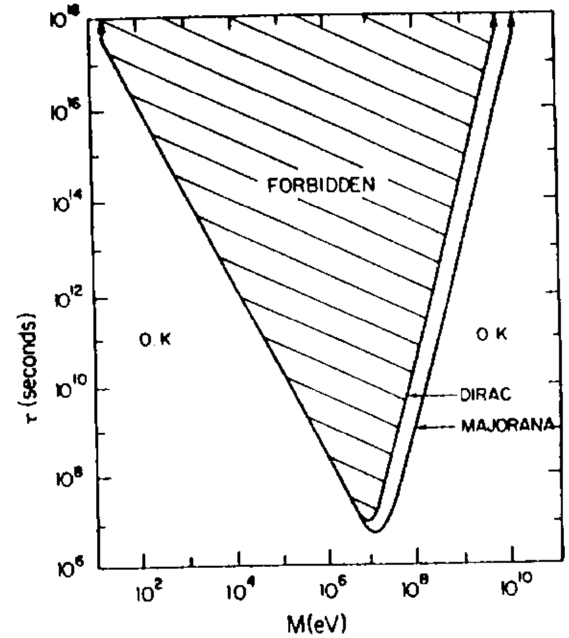

Figure 9: Limit on the neutrino lifetime from the energy density of the Universe

$$1+z_D \ge \begin{cases} (M/21.8\text{eV}) & 21.8\text{eV} \le M \le 1\text{ MeV}\\ (5\text{GeV}/M)^2 & 1\text{ MeV} \le M \le 5\text{ GeV}\\ (18\text{GeV}/M)^2 & \text{MeV} \le M \le 18\text{ GeV}.\end{cases} \qquad (2.27)$$

The limit on z_D can be converted to a limit on the age of the Universe at decay (i.e., the neutrino lifetime).[6] This lifetime is shown in Fig. 9.

The limits in Fig. 9 obtain for any decay mode of the neutrino, even if the decay products are "invisible", e.g., light neutrinos. However if the neutrino decay products include "visible" particles, such as γ, $e^\pm$, pions, etc., much better limits can be placed. The limits will depend on the epoch of decay. Decay at five different epochs will be considered.

Before discussing the limits, it is useful to calculate the time at which the energy density of the massive neutrino would dominate the energy density in photons. The energy density in photons is $\rho_\gamma = (\pi^2/15)T^4$, and if the neutrinos are NR, their energy density is $\rho_\nu = Y_\infty M s$. The energy densities are equal at $T \simeq 3Y_\infty M$, using $g_*' \simeq 4$. For heavy neutrinos Y_∞ is given by Eq. 2.24, and for light neutrinos, $Y_\infty = 135\zeta(3)/44\pi^4 = 0.04$, again using $g_*' \simeq 4$. Therefore the neutrino energy density will exceed the photon energy density at $T/M \le 0.1$ for light neutrinos, and $T/M \le 6 \times 10^{-8} M_{\rm GeV}^{-3}$ for heavy neutrinos. Using $t \simeq 1$ sec$/T_{\rm MeV}^2$ for the age of the Universe, the age of matter domination by the massive neutrinos is given by

$$t(\text{sec.}) \simeq \begin{cases} 10^{14}(M/1\text{eV})^{-2} & \text{light neutrinos} \\ 10^{8} M_{\rm GeV}^{-4} & \text{heavy neutrinos.} \end{cases} \tag{2.28}$$

• $t_U \simeq 3 \times 10^{17}$sec. $\le \tau$: If the lifetime is greater than the age of the Universe, the decay products will contribute to the background photon flux. The line will be narrow, since the cold neutrinos have a small velocity characterized by the velocity dispersion of galaxies, $\langle v^2 \rangle^{1/2} \simeq 10^{-3}$. The photon flux from ν decay will be

$$\begin{aligned} \mathcal{F} &= n_\nu t_U \tau^{-1} 10^3 \\ &= \begin{cases} 10^{33}\tau_{\rm sec}^{-1} \text{ erg/erg cm}^2\text{sec} & n_\nu \simeq 100 \text{ cm}^{-3} \\ 10^{23} M_{\rm GeV}^{-3} \tau_{\rm sec}^{-1} \text{ erg/erg cm}^2\text{sec} & n_\nu \simeq 10^{-8} M_{\rm GeV}^{-3} \text{ cm}^{-3}. \end{cases} \end{aligned} \tag{2.29}$$

The observed photon background is shown in Fig. 10. A very rough limit of

$$\mathcal{F} \le \left(\frac{1\text{MeV}}{E}\right) \text{erg/erg cm}^2\text{sec}, \tag{2.30}$$

can be placed on the contribution of neutrino decay products to the photon background. Requiring that the contribution of Eq. 2.29 is less than this limit gives the limit on τ in this lifetime range of ($M_{\rm eV} = M/$eV) [6]

$$\tau_{\rm sec} \ge \begin{cases} 10^{27} M_{\rm eV} & \text{light neutrinos} \\ 10^{26} M_{\rm GeV}^{-2} & \text{heavy neutrinos.} \end{cases} \tag{2.31}$$

• $t_{rec} \simeq 2 \times 10^{12}$sec. $\le \tau \le t_U$: If the neutrino decays after recombination, but before t_U, the photons will not scatter and should appear in the photon background. If $T \ge M$ at the time of neutrino decay, the Universe would not yet be dominated by the massive neutrino, and the massless decay products will have an energy density less than that in the MBR, and escape detection. Thus, any lifetime in the range $t_{rec} \le \tau \le t_U$ is forbidden if $T \le M$ at decay. [7]

The forbidden region in (τ, M) where decay of the neutrino would result in a photon flux in excess of the observed limit is indicated as region **A** in Fig. 11. Also shown in Fig. 11 is the region marked MDU, which is the disallowed region from Fig. 9.

• $t_{therm} \simeq 10^6$sec. $\le \tau \le t_{rec}$: If the neutrino decays before recombination (and $M \ge T$), the photons from the decay can scatter with electrons, which can, in

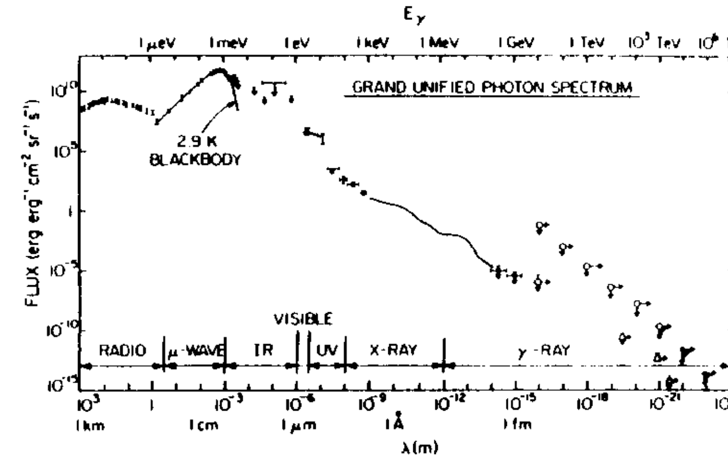

Figure 10: The photon background (graph courtesy of M. S. Turner)

turn, scatter with the photons in the MBR, leading to unacceptable distortions in the spectrum. However, if the neutrinos decay early enough, the initial distortions in the MBR can be re-thermalized. The time for the relaxation to a thermal spectrum is determined by the cross section for additional photon production through $\gamma + e \to \gamma + \gamma + e$. The cross section for the double Compton process is smaller than the cross section for single Compton process, so $t_{therm} \simeq 10^6$sec. $\le t_{rec}$. The forbidden region in (τ, M) where distortions of the MBR will result is indicated as region **B** in Fig. 11. [7]

• $t_{end\ nucleo} \simeq 3$ min. $\le \tau \le t_{therm}$: If the neutrino dominates the Universe and decays before t_{therm}, the present MBR is due to the photons produced in neutrino decay. The photons from neutrino decay increase the entropy of the Universe (the neutrino was "out of equilibrium" at decay). This increase in entropy after nucleosynthesis means that η during nucleosynthesis was larger than η today, and the success of primordial nucleosynthesis is lost. It is also possible that the high-energy photons from neutrino decay destroy the light elements produced in primordial nucleosynthesis.[8] The forbidden region in (τ, M) where the entropy produced in neutrino decay decreases η below acceptable values is indicated as region **C** in Fig. 11.[9]

• $t_{begin\ nucleo} \simeq 1$sec. $\le \tau \le t_{end\ nucleo}$: If the neutrino lifetime is longer than about 1 sec., the neutrino can dominate the mass of the Universe during nucleosyn-

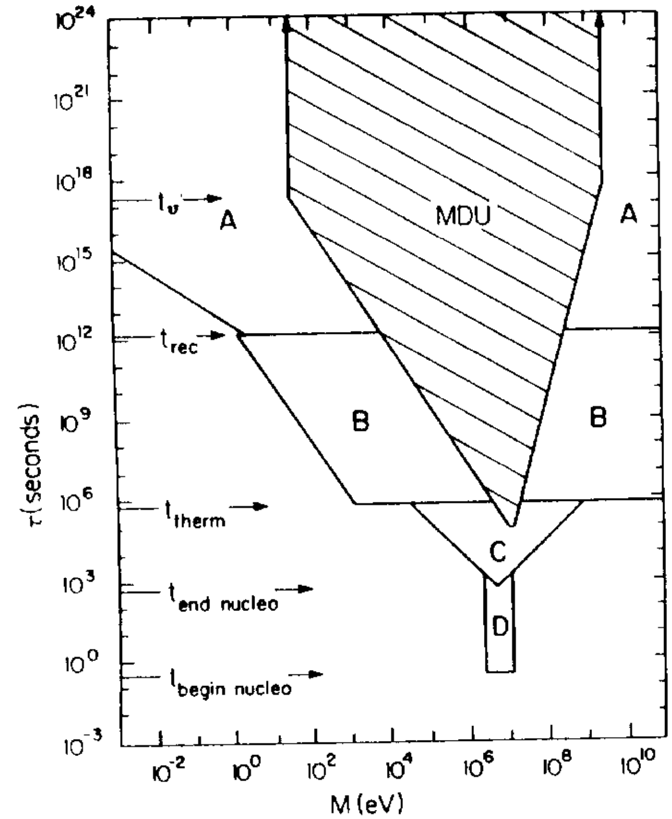

Figure 11: Cosmological limits on the mass and lifetime of neutrino decay to visible modes

thesis, leading to an increase in ^{4}He production. The forbidden region in $(\tau,\ M)$ which would lead to an overproduction of primordial ^{4}He is indicated as region D in Fig. 11.[10]

Neutrino decay into visible modes would also have other "astrophysical" effects.[Type II supernovae are a source of neutrinos. They explode with a frequency of about $f_{SN} = 1$ per galaxy per century, releasing about 10^{53}ergs of energy in neutrinos of energy $E_\nu \simeq 10$ MeV (i.e., about $N_\nu = 10^{57}$ neutrinos). If the neutrino decàys to photons with a lifetime greater than the age of the Universe, they would contribute to the γ-ray background flux

$$F_\gamma = N_\nu f_{SN} n_G t_U \frac{t_U}{\tau'} = 60 \frac{t_U}{\tau'} \mathrm{cm}^{-2}\, \mathrm{s}^{-1}\, \mathrm{sr}^{-1}, \tag{2.32}$$

where τ' is the lifetime in the rest frame of the Universe, $\tau' = \tau E/M$, and n_G is the number density of galaxies. If F_γ from neutrino decay is less than observed ($F_\gamma(observed) \leq 10^{-3}$ cm^{-2} s^{-1} sr^{-1}), the lifetime must satisfy $\tau' \geq 10^{22}$ sec., or $\tau \geq 10^{16} M_{eV}$ sec. This limit has assumed that the neutrino is light enough to be produced in the explosion ($M \leq 10$ MeV), and decays outside of the exploding star ($\tau' \geq 10^3$sec., or $\tau \geq 10^{-3} M_{eV}$ sec).[11]

The same argument may be used for white dwarfs.[6] They occur with a frequency of about 1 per year per galaxy, releasing about 10^{55} neutrinos of energy 100 keV. In order that the decay products of the neutrino give an X-ray flux smaller than the observed flux of 10^{-1} cm^{-2} s^{-1} sr^{-1}, requires $\tau \geq 3 \times 10^{16} M_{eV}$ sec.. The forbidden region in $(\tau,\ M)$ is shown in Fig. 12.

The combination of cosmological and astrophysical limits provide information and limits on neutrino properties that are not accessible to accelerators.

2.5 Conclusions

The decoupling and survival of neutrinos have been considered in detail, and limits on the mass and lifetime of massive neutrinos have been derived. Many of the techniques developed for neutrinos can (and have) been applied to other particles. There are some quite general comments that can be made. First, x_f is pretty much model independent, and is usually in the range 25-40. The dependence of x_f on the mass, cross section, etc., is only logarithmetic. Y_∞, on the other hand, is proportional to M^{-1} and σ_0^{-1}. For fixed M, Y_∞ increases as σ_0 decreases. This is obvious, since as the cross section decreases, the particles will be less efficient in annihilation. The dependence on M, however, can be more complicated. If σ_0 is independent of M, $Y_\infty \propto M^{-1}$. In many cases, however, σ_0 will depend on M. There are examples where $\sigma_0 \propto M^{-2}$. In this case, $Y_\infty \propto (M\sigma_0)^{-1} \propto M$. There are also examples (such as massive neutrinos) where $\sigma_0 \propto M^{+2}$. In this case, $Y_\infty \propto (M\sigma_0)^{-1} \propto M^{-3}$.

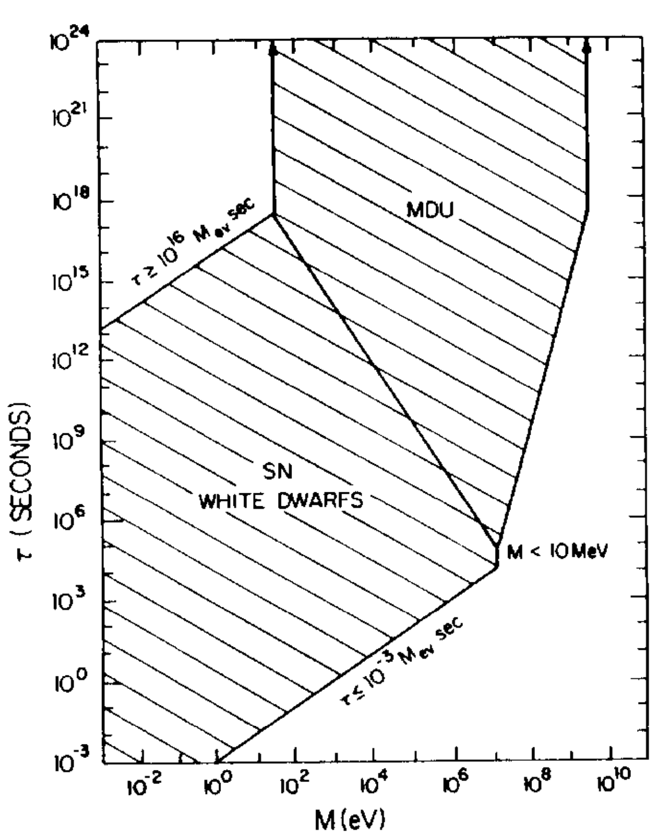

Figure 12: Astrophysical limits on the mass and lifetime of neutrino decay to visible modes

With the assumption that σ_0 is independent of M, that the fossil particle is stable, and that it gives a definite contribution to Ω, the value of σ_0 is almost uniquely determined. Consider a Majorana particle Ξ of mass M. If it freezes out when $g_* \simeq 75$, with $\sigma_0 = a \times 10^{-34}$ cm^2, then

$$x_f = 37\left[1 + \frac{1}{37}\ln M_{GeV} + \frac{1}{37}\ln a\right] = 37b^{1/2}$$

$$Y_\infty = \frac{1.8x_f^2}{m_{Pl}M\sigma_0} = 8 \times 10^{10} M_{GeV}^{-1} a^{-1} b. \qquad (2.33)$$

The contribution to Ωh^2 from Ξ is

$$\Omega_\Xi = \frac{MY_\infty s}{\rho_C} = 0.225a^{-1}bh^{-2}, \qquad (2.34)$$

which is *independent* of M. If $a = b = 1$, $h = 1/2$, then $\Omega_\Xi \simeq 0.9$. Since we know from nucleosynthesis that the contribution to Ω from nucleons is about 0.1, $\Omega_\Xi = 0.9$ could be the dark matter necessary if $\Omega_{TOTAL} = 1$. A cross section of this magnitude is "weak", and could be relevant for a variety of proposed particles, such as the ones from the supersymmetric zoo.

As shown above, if $\Omega_\Xi = 0.9$ the annihilation cross section is determined. The annihilation cross section also determines the rate of annihilation of Ξ in the *present* Universe. The possibility of present annihilation of Ξ's in the galactic halo, in the sun, and in the earth has been explored. It may be possible to detect the annihilation products, either as photons from annihilation in the halo, or as neutrinos from annihilation in the sun or earth. The annihilation cross section is also related to the scattering cross section of Ξ as it passes through matter. A cross section of 10^{-34} cm^2 may be large enough to detect cosmic-ray Ξ's by bolometric or other low threshold detectors.

2.6 References

1. Freeze out has been considered by many people. For a review, see G. Steigman, *Ann. Rev. Nucl. Part. Sci.* **29**, 313 (1979); for a discussion of the numerical work, see S. Wolfram, *Phys. Lett* **82B**, 65 (1979); the derivation from the Boltzmann equation follows E. W. Kolb and S. Wolfram, *Nucl. Phys.* B172, 224 (1980).

2. R. J. Scherrer and M. S. Turner, Fermilab Preprint (1986).

3. R. Cowsik and J. McClelland, *Phys. Rev. Lett.* **29**, 669 (1972).

4. B. W. Lee and S. Weinberg, *Phys. Rev. Lett* **39**, 165 (1977); P. Hut, *Phys. Lett.* **69B**, 85 (1977); K. Sato and M. Kobayashi, *Prog. Theor. Phys.* **58**, 1775 (1977); M. I. Vysotskii, A. D. Dolgov, and Ya. B. Zel'dovich, *JETP Lett.* **26**, 188 (1977).

5. D. A. Dicus, E. W. Kolb, and V. L. Teplitz, *Phys. Rev. Lett* **39**, 169 (1977).

6. R. Cowsik, *Phys. Rev. Lett.* **39**, 784 (1977).

7. D. A. Dicus, E. W. Kolb, and V. L. Teplitz, *Astrophys. J.* **221**, 327 (1978).

8. D. Lindley, *Mon. Not. R. Astr. Soc.* **188**, 15 (1979).

9. D. A. Dicus, E. W. Kolb, V. L. Teplitz, and R. V. Wagoner, *Phys. Rev. D* **17**, 1529 (1978).

10. E. W. Kolb and R. J. Scherrer Phys. Rev. D **25**, 1481 (1982).

11. S. W. Falk and D. N. Schramm, *Phys. Lett.* **79B**, 511 (1978).

3 THE EVOLUTION OF THE VACUUM

One of the most important tools in building particle physics models is the use of spontaneous symmetry breaking (SSB). The proposal that there are underlying symmetries of nature that are not manifest in the vacuum is a crucial link in the unification of forces. Of particular interest for cosmology is the expectation that at the high temperatures of the big bang, a symmetry that is broken today, will be restored, and that there is a phase transition to the broken state. The possibility that topological defects will be produced in the transition is the subject of this section. The possibility that the Universe will undergo inflation in a phase transition will be the subject of section 5.

Before discussing the creation of topological defects in the phase transition, some general aspects of high-temperature restoration of symmetry and the development of the phase transition will be reviewed.

3.1 High Temperature Symmetry Restoration

To study temperature effects, consider a real scalar field described by the Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - V(\phi)$$

$$V(\phi) = -\frac{1}{2}M^2\phi^2 + \frac{1}{4}\lambda\phi^4. \qquad (3.1)$$

This potential is shown in Fig. 13. The minima of the potential (determined by the condition $\partial V/\partial\phi = 0$), and the value of the potential at the minima are given by

$$\langle\phi\rangle = \pm\sqrt{\frac{M^2}{\lambda}}$$

$$V(\langle\phi\rangle) = -\frac{M^4}{4\lambda}. \qquad (3.2)$$

Presumably, the ground state of the system is either $+\langle\phi\rangle$ or $-\langle\phi\rangle$ and the reflection symmetry $\phi \leftrightarrow -\phi$ present in the Lagrangian is not respected by the vacuum state. When a symmetry of the Lagrangian is not respected by the vacuum, the symmetry is said to be spontaneously broken.

From the definition of the stress tensor in terms of the Lagrangian

$$T_{\mu\nu} = -\partial_\mu\phi\partial_\nu\phi - \mathcal{L}g_{\mu\nu}, \qquad (3.3)$$

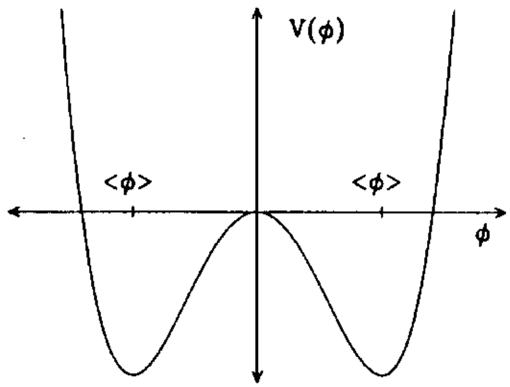

Figure 13: An example of the potential for a model with SSB

the energy density of the vacuum is

$$\langle T_{00}\rangle = \rho_V = -\mathcal{L} = V(\phi) = -\frac{M^4}{4\lambda}. \tag{3.4}$$

The contribution of the vacuum energy to the total energy density today must be smaller than the critical density $\rho_C = 1.88 \times 10^{-29} h^2$ g cm$^{-3} \simeq 10^{-46}$ GeV^4. Since this number is so small, it is tempting to require $\rho_V = 0$. This can be accomplished by adding to the Lagrangian a constant factor of $+M^4/4\lambda$. This constant term will not affect the equations of motion, and the sole effect will be to cancel the present vacuum energy.

There are several ways to understand the phenomena of high-temperature symmetry restoration. The most physical way is to express the effective finite-temperature mass of ϕ as the zero-temperature mass, $-M^2$, and a plasma mass, $M_{plasma} \simeq a\lambda T^2$, where a is a constant of order unity. If $M_T^2 = -M^2 + M_{plasma}^2 \leq 0$, the minimum of the potential will be at $\phi \neq 0$ (SSB), while if $M_T^2 = -M^2 + M_{plasma}^2 \geq 0$, the effective mass term will be positive and the minimum of the potential will be at $\phi = 0$ (symmetry restored). There is a critical temperature, $T_c = M/(a\lambda)^{1/2}$ above which $\langle\phi\rangle = 0$.[1]

A more rigorous approach to symmetry restoration is to account for the effect of the ambient background gas in the calculation of the higher-order quantum corrections to the classical potential. The finite temperature potential will include a temperature-dependent term that represents the free energy of ϕ particles at temperature T. To one loop, the full potential is[2]

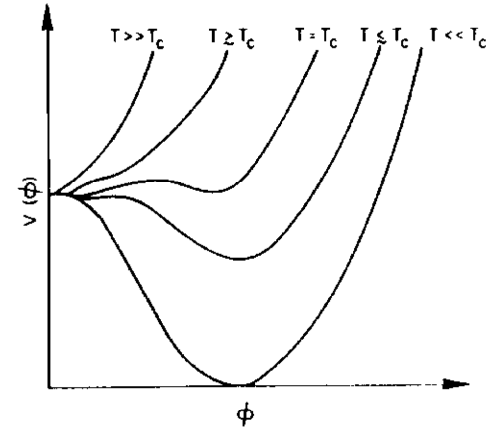

Figure 14: The temperature dependence of $V_T(\phi)$

$$V_T(\phi) = V(\phi) + \frac{T^4}{2\pi^2}\int_0^\infty dx x^2 \ln\left[1 - \exp[-(x^2 + \mu^2/T^2)^{1/2}]\right], \tag{3.5}$$

where $V(\phi)$ is the zero-temperature one-loop potential, and $\mu^2 = -M^2 + 3\lambda\phi^2$. At high temperature, Eq. 3.5 has the expansion

$$V_T(\phi) = V(\phi) + \frac{\pi^2}{90}T^4 + \frac{\lambda}{8}T^2\phi^2 + \cdots \tag{3.6}$$

The term proportional to T^4 is minus the pressure of a spinless boson, which should be the leading contribution to the free energy, and the second term is the "plasma" mass term for ϕ. Eq. 3.5 has a critical temperature, $T_c = 2M/\lambda^{1/2}$, above which the symmetry is restored.

The temperature dependence of $V_T(\phi)$ is shown in Fig. 14. The phase transition from the symmetric to the broken phase can be either first order or higher order. If at T_c there is a barrier between $\phi = 0$ and the SSB minimum $\phi = \sigma$, the change in ϕ will be discontinuous, signalling a first order transition. If no barrier is present at T_c, the change in ϕ will be continuous, signalling a higher order transition.

In general, at some temperature $T \leq T_c$, the $\phi = 0$ phase is a metastable phase, and this phase will be terminated by the decay of the false vacuum by quantum or thermal tunneling. Here, quantum tunneling will refer the zero-temperature part of the tunneling rate.

The quantum tunneling occurs by the nucleation of bubbles of the new phase. The probability for bubble nucleation is calculated [3] by solving the *Euclidean* equation of motion [10]

$$\Box_E\phi - V'(\phi) = \frac{d^2\phi}{dt^2} + \nabla^2\phi - V'(\phi) = 0 \qquad (3.7)$$

(where $V' \equiv dV/d\phi$) with boundary conditions $\phi = 0$ at $\vec{x}^2 + t^2 = \infty$. The probability of bubble nucleation per unit volume per unit time is

$$\Gamma = A\exp(-S_E) \qquad (3.8)$$

where S_E is the Euclidean action for the solution of Eq. 3.7

$$S_E(\phi) = \int d^4x \left[\frac{1}{2}\left(\frac{d\phi}{dt}\right)^2 + \frac{1}{2}(\nabla\phi)^2 + V(\phi)\right]. \qquad (3.9)$$

The calculation of the constant A is quite complicated, but for most applications a guess of A on dimensional grounds will suffice.

Of the many possible solutions to Eq. 3.7, the one with least action is the most important. The least action solution has $O(4)$ symmetry, and the Euclidean equation of motion becomes

$$\frac{d^2\phi}{dr^2} + \frac{3}{r}\frac{d\phi}{dr} - V'(\phi) = 0, \qquad (3.10)$$

with boundary conditions $\phi = 0$ at $r^2 = \vec{x}^2 + t^2 = \infty$ and $d\phi/dr = 0$ at $r = 0$. In general solutions to this equation can not be found. However in the "thin-wall" approximation, where the difference in energy between the metastable and true vacua are small compared to the height of the barrier, the "damping" term proportional to $d\phi/dr$ can be neglected. The solution for S_E is then simply

$$S_E = \int_0^\sigma d\phi\sqrt{2V(\phi)}. \qquad (3.11)$$

The tunneling rate at finite temperature[4] can be found following the above procedure, remembering that field theory at finite temperature is equivalent to Euclidean field theory with the time periodic with period T^{-1}. The finite-temperature tunneling rate is found by solving the equation of motion (only considering the least-action solution, which in this case has $O(3)$ symmetry)

[10] The tunneling rate is associated with a classical motion in imaginary time because the decay rate is related to the imaginary part of the energy. This is because the wave function oscillates in the classically allowed region, but is exponentially damped in the classically forbidden region.

$$\frac{d^2\phi}{ds^2} + \frac{2}{s}\frac{d\phi}{ds} - V_T'(\phi) = 0, \qquad (3.12)$$

where $s = \vec{x}^2$. The finite-temperature tunneling rate is

$$\Gamma_T = A\frac{S_3}{T}\exp(-S_3/T), \qquad (3.13)$$

where S_3 is the three-dimensional action of the solution of Eq. 3.12

$$S_3 = \int d^3x\left[\frac{1}{2}(\nabla\phi)^2 + V_T(\phi)\right]. \qquad (3.14)$$

3.2 Domain Walls

The simple model of the previous section can be used to demonstrate domain walls.[5] The Lagrangian can be written in the form

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{4}\lambda(\phi^2 - \langle\phi\rangle^2)^2; \qquad \langle\phi\rangle^2 \equiv \sigma^2 = \frac{M^2}{\lambda}. \qquad (3.15)$$

The Z_2 symmetry of the Lagrangian is broken when ϕ obtains a vacuum expectation value $\phi = +\sigma$ or $\phi = -\sigma$. Imagine that space is divided into two regions. In one region of space $\phi = +\sigma$, and in the other region of space $\phi = -\sigma$. The transition region between the two vacua is called a domain wall. Domain walls should be produced, for instance, in the nucleation of bubbles. The bubbles of true vacuum will be either $\phi = +\sigma$ or $\phi = -\sigma$, with equal probability.

Imagine a wall in the $x - y$ plane at $z = 0$. At $z = -\infty$, $\phi = -\sigma$, and at $z = +\infty$, $\phi = +\sigma$. The equation of motion for ϕ is

$$\Box\phi + \lambda\phi(\phi^2 - \sigma^2) = 0. \qquad (3.16)$$

The minimum energy solution to the equation of motion, subject to the boundary conditions above, is

$$\phi_W(z) = \sigma\tanh(z/\Delta) \qquad (3.17)$$

where Δ is the "thickness" of the wall, given by $\Delta = (\lambda/2)^{1/2}\sigma^{-1}$. This solution is illustrated in Fig. 15.

The finite, but non-zero, thickness of the wall is easy to understand. The terms contributing to the energy include a gradient term and a potential energy term. The gradient term is minimized by making the wall as thick as possible, and the potential term is minimized by making the wall as thin as possible, i.e., by minimizing the distance over which ϕ is away from $\pm\sigma$. The balance between these terms results in a wall of thickness Δ.

The stress tensor of Eq. 3.3 with $\phi = \phi_W$ is

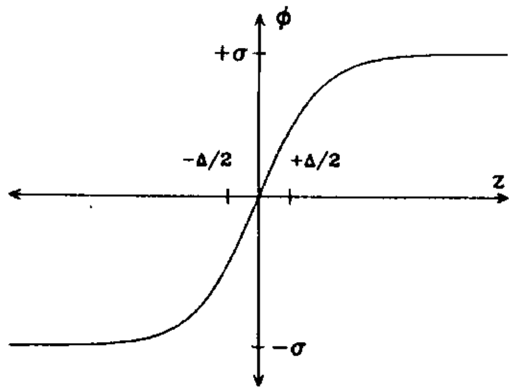

Figure 15: The solution for an infinite wall in the $x - y$ plane

$$T_\mu{}^\nu = \frac{\lambda}{2}\sigma^4 \cosh^{-4}(z/\Delta)\,\mathrm{diag}(1,1,1,0). \tag{3.18}$$

The energy density in the wall as a function of z is shown in Fig. 16.

From the stress tensor it is possible to define a surface tension for the wall, $\eta = \int T_0{}^0 dz = (4/3)(\lambda/2)^{1/2}\sigma^3$. It is also obvious from the stress tensor that since the (ii) component is equal to the (00) component, the gravitational interaction of the infinite wall will be non-Newtonian. This can lead to some strange interactions. For instance, two infinite walls in the $x - y$ plane will *repel* each other. This strange gravitational behavior only obtains for infinite and straight walls. The gravitational field at large distances from a spherical wall of radius R, would be that of a massive particle of mass $m \simeq R^2\sigma$.

The existence of domain walls can be ruled out today simply on the grounds of their contribution to the total mass of the Universe. A domain wall with $R \simeq R_{\rm horizon} \simeq H_0^{-1} \simeq 10^{28}$cm would contribute a mass of $M_{\rm wall} = \eta R^2_{\rm wall} = 10^{60}$grams. This would be about a factor of 10^5 larger than the total mass within $R_{\rm horizon}$.

The simple model of this section had domain walls because of the existence of disconnected vacuum states. The general condition for the existence of domain walls in the symmetry breaking $\mathcal{G} \to \mathcal{H}$ is that $\Pi_0(\mathcal{M}) \neq I$, where $\mathcal{M}$ is the manifold of equivalent vacuum states $\mathcal{M} \equiv \mathcal{G}/\mathcal{H}$, and Π_0 is the homotopy group that counts disconnected components. In the above example, $\mathcal{G} = Z_2$, $\mathcal{H} = I$, $\mathcal{M} = Z_2$, and $\Pi_0(\mathcal{M}) = Z_2 \neq I$.

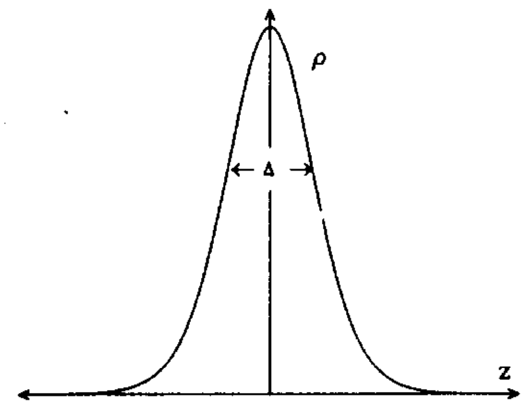

Figure 16: The energy density in the wall as a function of z

3.3 Cosmic Strings

A simple model that demonstrates the existence of cosmic strings is a gauge version of the model of the previous section. For a review of strings, see Refs. [6,7]. The Lagrangian of the model contains a U_1 gauge field, A_μ, in addition to the complex Higgs field, ϕ,

$$\mathcal{L} = D_\mu\phi D^\mu\phi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}\lambda(\phi^\dagger\phi - \langle\phi\rangle^2)^2; \qquad \langle\phi\rangle^2 = \sigma\exp(i\theta) \tag{3.19}$$

Again, $\sigma^2 = M^2/\lambda$, and

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu \\ D_\mu\phi &= \partial_\mu\phi - ieA_\mu\phi. \end{aligned} \tag{3.20}$$

Since there is a local gauge symmetry, $\theta = \theta(\vec{x})$, can be position dependent. Since ϕ is single valued, the total $\Delta\theta$ around any closed path must be an integer multiple of 2π. Imagine such a closed path with $\Delta\theta = 2\pi$. As the path is shrunk to a point (and no singularities are encountered), $\Delta\theta$ cannot change from $\Delta\theta = 2\pi$ to $\Delta\theta = 0$. There must therefore be one point contained within the path where the phase θ is undefined, i.e., $\langle\phi\rangle = 0$. The region of false vacuum within the path is part of a tube of false vacuum. These tubes of false vacuum either must be closed or infinite in length, otherwise it would be possible to deform the path around the tube, and contract it to a point without encountering the tube of false vacuum.

It will turn out that these tubes of false vacuum have a characteristic transverse dimension far smaller than its length, so they appear as one-dimensional objects called "strings." [11]

The string solution to the Lagrangian in Eq. 3.19 was first found by Nielsen and Olesen.[8] At large distances from an infinite string in the z-direction,

$$\begin{aligned} \phi &\longrightarrow \sigma \exp(in\theta) \\ A_\mu &\longrightarrow -ie^{-1}\partial_\mu [\ln(\phi/\sigma)], \end{aligned} \quad (3.21)$$

where θ is the angle in the $x-y$ plane. Note this choice of A_μ and ϕ, is a finite energy solution, since at large distances from the string, $D_\mu\phi \to 0$ and $F_{\mu\nu} \to 0$.

For an infinite string in the z-direction, the stress tensor takes the form

$$T_\mu{}^\nu = \mu\delta(x)\delta(y)\mathrm{diag}(1,0,0,1), \quad (3.22)$$

where μ is the mass per unit length of the string (string tension) given by $\mu \simeq \sigma^2$.

Far from a string loop of radius R, the gravitational field of the string is that of a particle of mass $M_{string} = \mu R_{string}$. For a string that stretches across the present horizon, the mass would be $M_{string} = 10^{18}(\sigma/\mathrm{GeV})^2$ grams. Cosmic string networks may have very interesting astrophysical consequences, including acting as seeds for the formation of large-scale structure.

String solutions will be present in the symmetry breaking $\mathcal{G} \to \mathcal{H}$, if the manifold of degenerate vacuum states $\mathcal{M} = \mathcal{G}/\mathcal{H}$ contains unshrinkable loops, i.e., if the mapping of $\mathcal{M}$ onto the circle is non-trivial. This is formally expressed by the statement that string solutions exist if $\Pi_1(\mathcal{M}) \neq I$. In the above example $\mathcal{G} = U_1$ was broken, $\mathcal{M}$ is a circle, and $\Pi_1(\mathcal{M}) = Z$, the set of integers.

3.4 Magnetic Monopoles

Domain walls are two-dimensional topological defects, and strings are one-dimensional defects. Zero-dimensional defects appear in theories with SSB as magnetic monopoles.[9,10] For a simple model that illustrates the existence of magnetic monopoles, consider an SO_3 gauge theory with a Higgs triplet field ϕ^a

$$\mathcal{L} = \frac{1}{2}D_\mu\phi^a D^\mu\phi^a - \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} - \frac{1}{4}\lambda(\phi^a\phi^a - (\phi)^2)^2; \qquad (\phi)^2 = \sigma\hat{\sigma}, \quad (3.23)$$

where $\sigma\hat{\sigma}$ is an isovector in the SO_3 space of magnitude σ and direction $\hat{\sigma}$ ($\hat{\sigma}$ is a unit isovector). Here

$$\begin{aligned} F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - e\epsilon_{abc}A^b_\mu A^c_\nu, \\ D_\mu\phi^a &= \partial_\mu\phi^a - e\epsilon_{abc}A^b_\mu\phi^c. \end{aligned} \quad (3.24)$$

[11] There should be no confusion between the (cosmic) strings considered here, and superstrings.

Since the theory has a local gauge symmetry, σ is a constant, but $\hat{\sigma}$ can be a function of $\vec{x}$. Imagine a configuration in which at one point $\phi^a = \sigma(0,0,1)$, at another point $\phi^a = \sigma(0,1,0)$, at another point $\phi^a = \sigma(1,0,0)$, and so forth. The lowest-energy configuration has ϕ^a = constant, and the x-dependence of ϕ^a can in general be gauged away. However there are configurations that cannot be deformed into a configuration of constant $\hat{\sigma}$ by a finite-energy transformation. An example of such a configuration is the "hedgehog" configuration, in which $\hat{\sigma} = \hat{r}$, where $\hat{r}$ is the unit vector in the radial direction. But for the obvious angular dependence, the solution is spherically symmetric at $r \to \infty$

$$\begin{aligned} \phi^a(r,t) &\to \sigma\hat{r} \\ A^a_\mu(r,t) &\to \epsilon_{\mu ab}\hat{r}_b/er. \end{aligned} \quad (3.25)$$

The magnetic field at $r \to \infty$ corresponding to the hedgehog solution is

$$B^a_i = \frac{1}{2}\epsilon_{ijk}F^a_{jk} = \frac{\hat{r}_i\hat{r}^a}{er^2}, \quad (3.26)$$

which is the magnetic field of a magnetic charge of $g = 1/e$. The mass of the field configuration is $M_{monopole} \simeq \sigma/e$.

There have been many experiments to look for magnetic monopoles. The limit on the average number density of magnetic monopoles in the Universe depends upon the properties of the monopoles (mass, charge, proton decay catalysis, etc.). If magnetic monopoles exist, they would have a multitude of astrophysical consequences.

Monopoles will be present in the symmetry breaking $\mathcal{G} \to \mathcal{H}$, if the manifold of degenerate vacuum states contains unshrinkable surfaces, i.e., if the mapping of $\mathcal{M}$ onto the two-sphere is non-trivial. This is formally expressed by the statement that monopole solutions exist if $\Pi_2(\mathcal{M}) \neq I$. In the above example $\mathcal{G} = SO_3$, $\mathcal{H} = U_1$ and $\Pi_2(\mathcal{M})$ is the set of even integers.

3.5 The Kibble Mechanism

The existence of the above topological defects is a prediction of many gauge theories with SSB. They are inherently non-perturbative, and cannot be produced in high energy collisions. The only place they can be produced is in phase transitions in the early Universe. Although monopoles, strings, and domain walls are topologically stable, they are, of course, not the minimum energy solution. However the production of the defects in the phase transition seems unavoidable. The mechanism for the production of the defects is known as the Kibble mechanism.[8]

The Kibble mechanism is based upon the fact that in the phase transition the correlation length is limited by the particle horizon. The particle horizon is the maximum distance over which a massless particle could propagate from the time of the bang. Imagine that a particle is emitted at coordinates ($t=0$, $r=r_H$, $\theta=0$, $\phi=0$) and is detected at the origin of the coordinate system at coordinates ($t=t$, $r=0$, $\theta=0$, $\phi=0$). As before, r_H is given by

$$\int_0^t \frac{dt'}{R(t')} = \int_0^{r_H} \frac{dr}{(1-kr^2)^{1/2}} \simeq r_H, \tag{3.27}$$

The coordinate r_H by itself is just a label. The proper distance to the horizon is given by $d_H = R(t)r_H$, so

$$d_H = R(t)\int_0^t \frac{dt'}{R(t')}. \tag{3.28}$$

If $R \propto t^n$ $(n > 1)$, then $d_H = (1-n)^{-1}t$.

The correlation length in the phase transition sets the maximum distance over which the Higgs field can be correlated. In general, the calculation of the correlation length depends upon the details of the transition. However, the fact that the horizon is finite in the standard cosmology implies that at the phase transition ($t=t_c$, $T=T_c$), the Higgs field must be uncorrelated on scales greater than the horizon, so the horizon acts as an effective correlation length.

Imagine that at the phase transition the Higgs field is uncorrelated on scales greater than $\xi = d_H$. The initial random nature of $\langle\phi\rangle$ is damped (remember E_{min} occurs for $\langle\phi\rangle$ = constant). However there are Higgs configurations that are topologically stable and will be frozen in as topological defects.

Consider monopoles as an example of the freezing in of topological defects.[11] The direction of the isovector Higgs field is random on scales greater than ξ. The probability that a random orientation of $\langle\phi\rangle$ will have a hedgehog structure is about 0.1, so there should be about one monopole (or antimonopole) per 10 horizon volumes, $n_M = 0.1d_H^3 \simeq 0.1(m_{Pl}/T_c^2)^3$, using the age of a radiation-dominated Universe $t = m_{Pl}/T^2$. The entropy density at T_c is $s \simeq T_c^3$, so the monopole-entropy ratio is $n_M/s \simeq 0.1(T_c/m_{Pl})^3$. Since monopole-antimonopole annihilation is not important, if entropy is not created after monopole production, the above monopole-entropy ratio should obtain today. For $T_c = 10^{15}$GeV, $m_M = 10^{16}$ GeV as expected in grand unified theories, $n_M/s \simeq 10^{-13}$, which gives the present energy density in magnetic monopoles $\rho_{monopoles} \simeq 10^{11}\rho_C$. Obviously some mechanism must suppress monopole production, enhance monopole annihilation, or increase entropy. An increase in entropy would also dilute the abundance of strings and domain walls. It is possible that monopoles were diluted to a level accessible to observation, or that strings were produced after the dilution of monopoles. Detection of monopoles or strings would provide unique information about both particle physics and cosmology. In complicated gauge theories with several symmetry breaking steps there are often interesting hybrid creatures, such as domain walls bounded by strings, strings terminated by monopoles, monopoles with strings through them, etc. They all have unique signatures, and observation of them would provide information about the steps of symmetry breaking.

3.6 References

1. D. A. Kirzhnits and A. D. Linde, Sov. Phys. JETP **40**, 628 (1974).
2. L. Dolan and R. Jackiw, Phys. Rev. D. **9**, 3320 (1974); S. Weinberg, Phys. Rev. D. **9**, 3357 (1974).
3. S. Coleman, Phys. Rev. D **15**, 2929 (1977); C. Callan and S. Coleman, Phys. Rev. D **16**, 1762 (1977); S. Coleman and F. De Luccia, Phys. Rev. D **21**, 3305 (1980).
4. A. Linde Nucl. Phys. **B216**, 421 (1983).
5. Ya. B. Zel'dovich, I. Yu. Kobzarev, and L. B. Okun, Sov. Phys. JETP **40**, 1 (1975).
6. T. W. B. Kibble, J. Phys. **A 9**, 1387 (1976). This is an excellent paper that is required reading in the subject.
7. A. Vilenkin, Phys. Rep **121**, 263 (1985). This is a detailed and well-written review that contains many details not included here.
8. H. B. Nielsen and P. Olesen, Nucl. Phys. **B61**, 45 (1973).
9. G. 't Hooft, Nucl. Phys. **B79**, 276 (1974); A. M. Polyakov, JETP Lett. **20**, 194 (1974).
10. J. Preskill, Ann. Rev. Nucl. Part. Sci. **34**, 461 (1984).
11. J. Preskill, Phys. Rev. Lett. **43**, 1365 (1979).

4 BARYOGENESIS

One of the most successful applications of particle physics to cosmology is the use of grand unified theories (GUTs) to explain the origin of the baryon asymmetry. Before making some general comments on the generation of the asymmetry, the evidence for a baryon asymmetry will be reviewed. Finally, some new developments in baryogenesis concerning non-perturbative electroweak effects will be discussed

4.1 Evidence for a Baryon Asymmetry

Antimatter is rare on earth. It exists in "large" quantities only in the antiproton accumulators at Fermilab and CERN. Antimatter is also rare in the solar system. The fact that Neil Armstrong survived his "one small step" is evidence that the moon is matter. Planetary probes have visited seven of the nine (ten?) planets, and have shown that the solar system is matter.

Cosmic rays provide a sample of the entire galaxy (at least). Antiprotons are seen in cosmic rays at about the 10^{-4} level compared to protons. These cosmic rays are usually assumed to be secondary particles, and not primary particles from an antimatter source. The flux of antimatter nuclei is also below the 10^{-4} level compared to nuclei, and there is no clean event that signals detection of an antinucleus. Cosmic rays are solid evidence that there is a galactic asymmetry between baryons and antibaryons, and that this asymmetry is maximal, i.e., everything is baryons, and there aren't any antibaryons in the galaxy.

Our evidence on larger scales is somewhat more uncertain. There are clusters of galaxies containing intercluster gas. If both matter galaxies and antimatter galaxies exist in the same cluster there would result a large amount of annihilation which would contribute a large γ-ray flux. The absence of such a large γ-ray flux is evidence that clusters of galaxies are either all baryons or all antibaryons. There is basically no information on scales larger than clusters of galaxies ($\simeq 10^{14}\ M_\odot$).

If antimatter exists in appreciable quantities, it must be separated on scales larger than about $10^{14}\ M_\odot$. However this separation must be done before $T \simeq 1$ GeV, or as shown in Sec.2, the nucleons and antinucleons would have annihilated below acceptable limits. The existence of horizons in the standard model make such a separation impossible, since the mass in baryons contained within the horizon is less than a solar mass at $T = 1$ GeV.

The most reasonable conclusion is that at $T \geq 1$ GeV, there was an asymmetry between the number of baryons and the number of antibaryons.[1] This asymmetry is characterized by the baryon number discussed in Sec. 1, $B = 4 \times 10^{-9}\Omega_N h^2$. Although the baryon asymmetry is maximal today (i.e., no antimatter) at $T \geq 1$ GeV, the temperature was high enough to create $N\bar{N}$ pairs, and $B \simeq 10^{-9}$ means that for every billion antibaryons there were a billion and one baryons. The goal of this section is to explain how such a curious (but crucial) number could arise in a Universe with symmetric, $B = 0$, or even better, random, initial conditions.

particle		final state	branching ratio	B
X	$\rightarrow$	qq	r	2/3
X	$\rightarrow$	$\bar{q}\bar{l}$	$1-r$	-1/3
$\bar{X}$	$\rightarrow$	$\bar{q}\bar{q}$	$\bar{r}$	-2/3
$\bar{X}$	$\rightarrow$	ql	$1-\bar{r}$	1/3

Table 2: Final states and branching ratios

4.2 The Basic Picture

There are three basic ingredients necessary to generate a non-zero B from an initial symmetric state.[2] • *Baryon Number Violation:* There must obviously be a violation of baryon number. If baryon number is conserved in all interactions, the present baryon asymmetry must simply reflect the initial conditions. • *C and CP Violation:* Since B is odd under C and CP, they both must be violated to generate a non-zero B. • *Non-Equilibrium Conditions:* In chemical equilibrium the entropy is maximal when the chemical potentials associated with all non-conserved quantum numbers vanish. If baryon number is not conserved, if the Universe ever obtains chemical equilibrium, B must be zero.

To illustrate the simple model, consider a particle X which decays to final states with different baryon number (hence violating baryon number) with branching ratios given in Table 2.[3,4,5] Note that CP is violated if $r \neq \bar{r}$, but CPT is conserved, since the total rate for X decay is the same as the total rate for $\bar{X}$ decay.

Imagine a box containing equal amounts of X and $\bar{X}$. The baryon number produced by the decay of the X's is proportional to $B_X = r(2/3) + (1-r)(-1/3)$, and the baryon number produced by the $\bar{X}$'s is proportional to $B_{\bar{X}} = \bar{r}(-2/3) + (1-\bar{r})(1/3)$. The net asymmetry produced by an equal amount of X and $\bar{X}$'s is proportional to $\epsilon = B_X + B_{\bar{X}} = r - \bar{r}$. The baryon number vanishes, of course, if CP is conserved ($r = \bar{r}$). If there are no further baryon number violating reactions (equilibrium is not maintained) a net baryon asymmetry will result. This extremely simple picture illustrates the basic idea. The remainder of this section involves various refinements on this picture.[6]

First, consider the possibility of B, C, and CP violation, and non-equilibrium conditions in the early Universe.

• *B Violation:* The existence of baryon number violation seems to be a generic feature of GUTs. If strong and electroweak interactions are unified, quarks and leptons typically appear as elements of a common irreducible representation of the gauge group, and gauge bosons can mediate interactions that violate B. The limit on the stability of the proton, $\tau \geq 10^{31}$ years, implies that these bosons should have masses in excess of 10^{14}GeV or so. The weaker coupling allows Higgs bosons that mediate baryon-number violation to have a somewhat smaller mass. In both cases the large mass of the intermediate bosons is responsible for the feebleness of baryon-number violation today. This suppression should have been overcome at the extremely high temperatures in the big bang, and interactions that violate baryon number should be just as common as other interactions. The requirement of B violation is easy to arrange in GUTs. A gauge or Higgs boson that violates baryon number in its decay will be generically denoted as X.

• C and CP *Violation*: C is maximally violated in the weak interactions, so C violation in the decay of the X bosons should not be a fundamental problem. [12] CP violation is observed in the kaon system, with typical dimensionless strength of 10^{-3}. Since its origin is not well understood, it is easy to say it should also appear in the X-boson system as well.

To explicitly see how CP violation enters, consider a system with two massive particles, X and Y, with baryon number violating decays. The generalization of ϵ defined above is

$$\begin{aligned} \epsilon_X &= \sum_f B_f \frac{\Gamma(X \to f) - \Gamma(\bar{X} \to \bar{f})}{\Gamma_X} \\ \epsilon_Y &= \sum_f B_f \frac{\Gamma(Y \to f) - \Gamma(\bar{Y} \to \bar{f})}{\Gamma_Y}, \end{aligned} \tag{4.1}$$

where the sum is over all final states f with baryon number B_f, and Γ_X (Γ_Y) is the total X (Y) decay width. For simplicity, assume there are but two final states for X and Y decay, and that the interaction Lagrangian is given by

$$\mathcal{L} = g_1 X i_2^\dagger i_1 + g_2 X i_4^\dagger i_3 + g_3 Y i_1^\dagger i_3 + g_4 Y i_2^\dagger i_4 + \text{h.c.} \tag{4.2}$$

The lowest order processes for X and Y decay are shown in Fig. 17. The lowest order processes cannot contribute to ϵ, as $\Gamma(X \to \bar{i}_1 i_2) = |g_1|^2 I_X = \Gamma(\bar{X} \to i_1 \bar{i}_2) = |g_1^*|^2 I_X$ where $I_X = I_{\bar{X}}$ represents the phase space factors. The first non-zero contribution to ϵ comes from the interference of the lowest order contribution in Fig. 17, and the one-loop contributions shown in Fig. 18. The interference terms contribute to X decay terms given by

$$\begin{aligned} \Gamma(X \to \bar{i}_1 i_2) &= g_1 g_2^* g_3 g_4^* I_{XY} + (g_1 g_2^* g_3 g_4^* I_{XY})^* \\ \Gamma(\bar{X} \to i_1 \bar{i}_2) &= g_1^* g_2 g_3^* g_4 I_{XY} + (g_1^* g_2 g_3^* g_4 I_{XY})^*, \end{aligned} \tag{4.3}$$

[12]There is however the possibility that during some intermediate stage of symmetry breaking C is a good symmetry. For instance in the pattern $SO_{10} \to SU_4 \times SU_{2L} \times SU_{2R} \to SU_3 \times SU_{2L} \times U_1$, C invariance obtains during the first two stages of symmetry, and baryogenesis must await the final stage of symmetry breaking.

where I_{AB} now includes kinematic factors from integration over the internal momentum loop due to B exchange in A decay. If intermediate particles in the loop are kinematically allowed to propagate on shell, I_{AB} will be complex.

The difference between $X \to \bar{i}_1 i_2$ and $\bar{X} \to i_1 \bar{i}_2$ is given by

$$\begin{aligned} \Gamma(X \to \bar{i}_1 i_2) - \Gamma(\bar{X} \to i_1 \bar{i}_2) &= 2i I_{XY} \mathrm{Im}(g_1 g_2^* g_3 g_4^*) + 2i I_{XY}^* \mathrm{Im}(g_1^* g_2 g_3^* g_4) \\ &= 4 \mathrm{Im} I_{XY} \mathrm{Im}(g_1^* g_2 g_3^* g_4). \end{aligned} \tag{4.4}$$

With a similar calculation for the other decay mode

$$\epsilon_X = \frac{4}{\Gamma_X} \mathrm{Im} I_{XY} \mathrm{Im}(g_1^* g_2 g_3^* g_4) \left[(B_{i_4} - B_{i_3}) - (B_{i_2} - B_{i_1}) \right]. \tag{4.5}$$

For Y decay $\epsilon_Y = -\epsilon_X$, with $X \leftrightarrow Y$ everywhere.

There are several lessons to be learned from this tedious exercise. There must be *two* baryon number violating bosons, with masses greater than the sum of the masses in the internal loops. CP violation is manifest as complex coupling constants. The X and Y particles in the above example must not be degenerate, or the baryon number produced in X decay will cancel the baryon number produced in Y decay.

• *Non-Equilibrium Conditions*: The non-equilibrium conditions are provided by the expansion of the Universe. If the expansion rate is faster than particle interaction rates, non-equilibrium can result. Assume that at the planck time X and $\bar{X}$ are present in equilibrium, $n_X = n_{\bar{X}} = n_\gamma$. In LTE $n_X = n_{\bar{X}} = n_\gamma$ for $T \geq m_X$, and $\ll n_\gamma$ for $T \leq m_X$. However LTE will be obtained only if the X interaction rates are greater than H. If LTE is not maintained, n_X will remain equal to n_γ and there will be an excess of X relative to the equilibrium abundance. This is illustrated in Fig. 19.

The conditions necessary for a departure from equilibrium can be quantified. The decay rate of the X, denoted by Γ_D, the inverse decay rate (i.e., X production), denoted as Γ_{ID}, and the two-body scattering rate, denoted as Γ_S, and the expansion rate, denoted as H, are given by

$$\begin{aligned} \Gamma_D &= \alpha m_X \begin{cases} m_X/T & m_X \leq T \\ 1 & m_X \geq T \end{cases} \\ \Gamma_{ID} &= \Gamma_D \exp(-m_X/T) \\ \Gamma_S &= n\sigma \simeq T^3 \alpha^2 \frac{T^2}{(T^2 + m_X^2)^2} \\ H &\sim g_*^{1/2} T^2 / m_{Pl}. \end{aligned} \tag{4.6}$$

A comparison of the rates are given in Fig. 20. Note that all the reaction rates depend upon m_X, while H is independent of m_X. It is necessary to determine

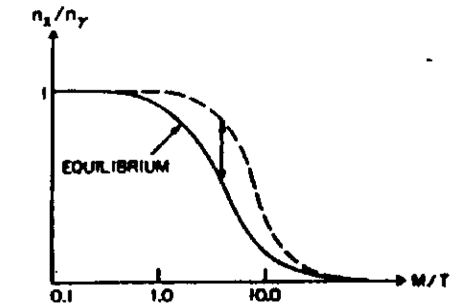

Figure 19: Departure from equilibrium denoted as the arrow between equilibrium (solid) and actual (dashed) ratio of n_X/n_γ

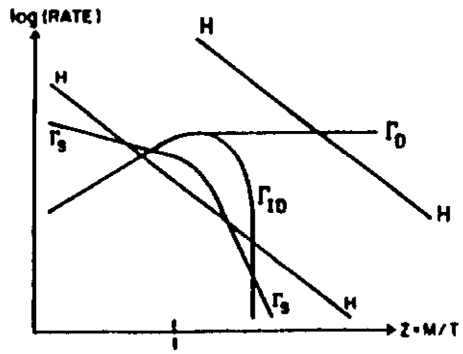

Figure 20: Rates in the early Universe

what sets the scale of H relative to the other rates. As shown in Fig. 20, H can be either large or small compared to the rates at $T = m_X$.

The relevant quantity to determine if X will be over abundant is the ratio of the decay rate to the expansion rate at $T = m_X$, given by

$$K \equiv (\Gamma_D/H)_{T=m_X} = \frac{\alpha m_{Pl}}{g_*^{1/2} m_X}. \tag{4.7}$$

If $K \ll 1$, equilibrium will not be maintained and the X will become over abundant. In this limit the X just drifts along and then decays. In the limit of pure drift and decay (the British limit) $B = g_*^{-1}\epsilon$. If $K \gg 1$, the X will be in good causal contact, there will be no departure from equilibrium, and baryogenesis will be thwarted. Exactly what is meant by "$\gg$" and "$\ll$" can be found by a simple model. If X is a gauge boson, then $\alpha = g^2_{\rm GUT}/4\pi \simeq 1/45$, and $K = 7\times 10^{15}{\rm GeV}/M$. If X is not a gauge boson, the effective α may be smaller, and the corresponding K may be smaller.

4.3 A Simple Model

The simple model of this section is constructed to study the dynamics of baryogenesis. The simple model consists of a massive boson $X \equiv \bar{X}$, light particles b and $\bar{b}$ with baryon numbers $+1/2$ and $-1/2$, and light particles γ with baryon number zero. The amplitudes for X decay are taken to be [13]

$$\begin{aligned} |M(X \to bb)|^2 = |M(\bar{b}\bar{b} \to X)|^2 &= \frac{1}{2}(1+\epsilon)|M_0|^2 \\ |M(X \to \bar{b}\bar{b})|^2 = |M(bb \to X)|^2 &= \frac{1}{2}(1-\epsilon)|M_0|^2. \end{aligned} \tag{4.8}$$

If the interactions with the "photons" γ are rapid enough, kinetic equilibrium will be maintained, and the phase space densities will be given as

$$\begin{aligned} f_b(E) &= \exp\left[-(E-\mu)/T\right] \\ f_{\bar{b}}(E) &= \exp\left[-(E+\mu)/T\right] \\ f_X(E) &= \exp\left[-E/T\right], \end{aligned} \tag{4.9}$$

where μ is the baryon chemical potential. The assumption has been made that the reactions establishing *kinetic* equilibrium (baryon number conserving reactions like $b\bar{b} \leftrightarrow \gamma$'s) are rapid compared to H. Kinetic equilibrium implies $\mu \equiv \mu_b = -\mu_{\bar{b}}$. *Chemical* equilibrium depends upon baryon number violating reactions, and would imply $\mu = 0$. The baryon to photon ratio is given by

[13] CPT invariance relates $M(i \to j)$ and $M(\bar{j} \to \bar{i})$.

$$\frac{n_B}{n_\gamma} = \left(\frac{n_b - n_{\bar b}}{n_\gamma}\right) = 2\sinh\left(\frac{\mu}{T}\right) = 2\frac{\mu}{T} + \cdots, \tag{4.10}$$

where Maxwell-Boltzmann statistics have been used for all particles.

A Boltzmann equation for the decay and inverse decay of X can be developed along the same lines as for annihilation and freeze out. The equation for the evolution of the X number density is

$$\begin{aligned} \dot n_X + 3Hn_X &= \Lambda^X_{12}\Big[-f_X(p_X) + \frac{1}{2}(1+\epsilon)|M_0|^2 f_{\bar b}(p_1)f_{\bar b}(p_2) \\ &\quad + \frac{1}{2}(1-\epsilon)|M_0|^2 f_b(p_1)f_b(p_2)\Big] \\ &= \Lambda^X_{12}[-f_X(p_X) + f_X^{eq}(p_X)]\,|M_0|^2 + O(\epsilon) + O(\mu/T) \\ &= -\Gamma_D(n_X - n_X^{eq}). \end{aligned} \tag{4.11}$$

The equation for the evolution of n_b also includes terms involving exchange of an intermediate X in $bb \leftrightarrow \bar b\bar b$. [14] The equation for $\dot n_b$ is

$$\begin{aligned} \dot n_b + 3Hn_b = \;& \Lambda^X_{12}\;\left[-(1-\epsilon)f_b(p_1)f_b(p_2) + (1-\epsilon)f_X(p_X)\right]|M_0|^2 \\ & + 2\Lambda^{34}_{12}\Big[-f_b(p_1)f_b(p_2)|M'(bb \to \bar b\bar b)|^2 \\ & + f_{\bar b}(p_3)f_{\bar b}(p_4)|M'(\bar b\bar b \to bb)|^2\Big]. \end{aligned} \tag{4.12}$$

The corresponding equation for $n_{\bar b}$ is obtained from Eq. 4.12 by the interchange $b \leftrightarrow \bar b$; $\epsilon \leftrightarrow -\epsilon$.

The baryon number density is defined as $n_B = (n_b - n_{\bar b})/2$, and is given by

$$\dot n_B + 3Hn_B = \epsilon\Gamma_D(n_X - n_X^{eq}) - (n_B/n_\gamma)n_X^{eq}\Gamma_D - n_B 2n_b(|v|\sigma)' + O(\epsilon^2). \tag{4.13}$$

Eqs. 4.11 and 4.13 form a coupled system of equations that can be solved to yield $n_B(t)$. Eq. 4.13 illustrates the three ingredients for baryogenesis. The first term on the right-hand side is the production term. It vanishes unless CP is violated ($\epsilon \neq 0$), B is violated ($\Gamma_D \neq 0$), and non-equilibrium is obtained ($n_X \neq n_X^{eq}$). The last two terms describe damping of the baryon asymmetry due to inverse decays and two-body scattering processes.

The two-body scattering term in $\dot n_B$ is only important for $K \geq K_C$, where $K_C \simeq 10^3/A\alpha$, if the low-energy cross section is $\sigma = A\alpha^2T^2/m_X^4$.

The numerical solution of the system of equations is shown if Fig. 21. As expected, for $K \ll 1$, $B = \epsilon/g_*$, while for $K \geq 1$, the baryon number is damped. The baryon number as a function of K is shown in Fig. 22. For $K \leq K_C$ the

[14] Of course the contribution due to an on-shell s-channel X exchange is already accounted for by sequential X inverse decay and decay. The $\prime$ superscript on the amplitudes and cross section indicates that this contribution is to be removed.

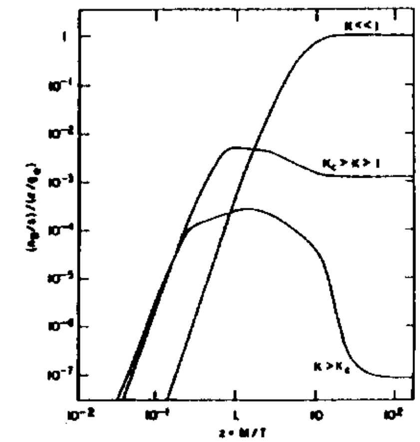

Figure 21: The development of the baryon number $B = n_B/n_\gamma$

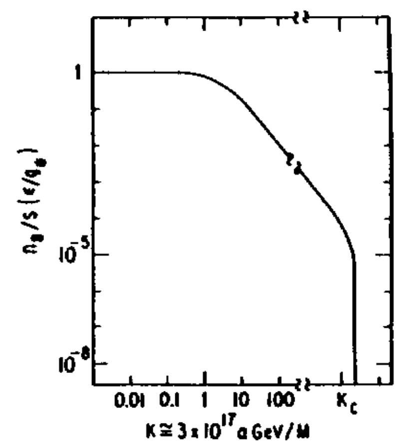

Figure 22: The baryon number as a function of $K = \alpha m_{Pl}/g_*^{1/2}m_X$

the different vacua are accompanied by a change in the baryon number. This process is usually associated with instantons that describe the tunneling between different θ-vacua. Since the process is inherently non-perturbative, the rate for baryon number non-conservation is proportional to $\exp(-4\pi/\alpha_W)$. This quantum tunneling is unimportant today, and was certainly unimportant in the early Universe.

At finite temperature the transitions between different vacua can be driven by finite temperature effects. Kuzmin, Rubakov, and Shaposhnikov (KRS) have recently shown that the tunneling at finite temperature may be appreciable.[8] In their analysis they considered an SU_2 gauge theory, which is a good approximation to the Weinberg-Salam model in the limit $\sin^2\theta_W \to 0$, α_W fixed. The Lagrangian for the SU_2 model with gauge coupling g_W is given by

$$\mathcal{L} = \frac{1}{2}D_\mu\phi^a D^\mu\phi^a - \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} - \frac{1}{4}\lambda(\phi^a\phi^a - \sigma^2/2)^2 + \mathcal{L}_{\text{fermions}}. \tag{4.14}$$

The basic point is that fluctuations in A_μ and ϕ caused by finite temperature effects cause transitions between different θ-vacua, with associated baryon number violation. The calculation of the transition rate is related to the calculation of the fate of the false vacuum at finite temperature. However, in this case the calculation is much more complicated. The first approximation made is to ignore the effect of the fermion fields, and to only consider A_μ and ϕ as the dynamical fields. The validity of this approximation will be discussed later. The second approximation is to replace the finite temperature "bounce" action (S_3) in Eq. 3.14 by the free energy at the maximum of the barrier between adjacent θ-vacua.

The maximum free energy in the transition is dominated by a static configuration, given in the $A_0 = 0$ gauge by ($r^2 = \vec{x}^2$)

$$\begin{aligned} A^i_{cl} &= \frac{i}{g}\frac{\epsilon_{ijk}x_j\tau_k}{r^2}f(\xi) \\ \phi &= \frac{\sigma}{\sqrt{2}}\frac{i\vec{\tau}\cdot\vec{x}}{r}\begin{bmatrix}0\\1\end{bmatrix}h(\xi), \end{aligned} \tag{4.15}$$

where $\vec{\tau}$ are the Pauli spin matrices, and

$$f(0) = h(0) = 0, \qquad f(\infty) = h(\infty) = 1 \tag{4.16}$$

are functions of the dimensionless parameter $\xi = r/r_0 = rg_W\sigma$. The free energy of this configuration is given by

$$F \simeq \frac{2M_W}{\alpha_W}, \tag{4.17}$$

where $M_W = g_W\sigma/2$.

With the KRS assumption that the action for the tunneling rate at finite temperature is given by F, the change in the baryon number is

$$\dot{B} = -CBT\exp(-F/T), \tag{4.18}$$

where C is a dimensionless constant of order unity, and the overall factor of T appears on dimensional grounds. Since CP is conserved, B cannot be generated by the transitions. In the early Universe g, λ, and σ are a function of temperature. The temperature dependence of g and λ is only proportional to $\ln T$ and will be ignored. The temperature dependence of σ around the critical temperature is much more important. At $T \geq T_c$, $\sigma = 0$, and as the temperature passes through the critical temperature, $\sigma \to \sigma_0 = 250$ GeV.

The rate for baryon number non-conservation, as seen from Eq. 4.18, is given by $\Gamma_{\Delta B} = CT\exp(-F/T)$, which is greater than the expansion rate for $T \geq T^* \simeq 200$ GeV. This would be a long enough period to eradicate any baryon number. Since there is no anomaly in $B - L$,the transitions conserve $B - L$, so more precisely, it would eradicate any baryon number with zero projection on $B - L$. The two possible solutions would be either to generate an asymmetry in $B - L$ in the same manner an asymmetry in B is generated, [16] or to generate the baryon asymmetry after the electroweak transition.

The potential impact of the KRS calculation certainly warrants more detailed calculations of the finite-temperature induced tunneling between θ-vacua. In particular, one approximation made by KRS that might be questioned is the neglect of the plasma effects on the transition. The classical field configuration of Eq. 4.15 are spatially large. The characteristic size of the configurations are several r_0. Within this size are $10^3 - 10^4$ particles with weak charge. Since the configuration of A_μ is pure magnetic and the plasma is a magnetic conductor, there are no plasma screening effects in the *static* configuration. However in the evolution to and from the static configuration there are time-dependent gauge fields and A_μ must have an electric component which *will* be affected by screening. The problem of screening can be addressed in a U_1 theory by calculating how long it takes to establish a U_1 field configuration with a characteristic size of several r_0. By detailed balance the time it takes to establish this unstable field configuration is the same as the time it takes for the field configuration to decay. In the limit that the conductivity of the Universe is infinite, the time is infinite, i.e., in a perfect conductor currents in the plasma are induced to oppose the establishment of the field. Of course, the Universe is not a perfect conductor, and the time for the establishment of the coherent field, Γ_A, will be proportional to the conductivity.

It is possible to map the problem at hand into one treated before. Consider the field configuration in Eq. 4.15 as a particle (called the Sphaleron) with mass $m = F$ and decay width $\Gamma = \Gamma_A$. Since baryon number is violated in the decay (and CP is conserved), the change in the baryon number can be found from Eq. 4.13:

$$\dot{n}_B + 3Hn_B = -n_B n_S^{eq}\Gamma_S, \tag{4.19}$$

[16] In many GUTs (in particular SO_{10}) this is easy to accomplish.

where as usual $n_S^{eq} \simeq (mT)^{3/2}\exp(-m/T)$. Determination of Γ_S will allow a determination of the efficiency of the process in damping the baryon asymmetry.

4.5 References

1. The evidence for the asymmetry is reviewed by G. Steigman, *Ann. Rev. Ast. Astrophys.* **14**, 339 (1976). The view that the Universe is symmetric is reviewed by F. Stecker, *Ann. NY Acad. Sci.* **375**, 69 (1981).
2. A. D. Sakharov, *JETP Lett.* **5**, 24 (1967).
3. E. W. Kolb and S. Wolfram, *Nucl. Phys.* **B172**, 224 (1980).
4. E. W. Kolb and M. S. Turner, *Ann. Rev. Nucl. Part. Sci.* **33**, 645 (1983).
5. J. N. Fry, K. A. Olive, and M. S. Turner, *Phys. Rev. D.* **22**, 2953 (1980).
6. M. Yoshimura, *Phys. Rev. Lett.* **41**, 281 (1978); D. Toussaint, S. B. Trieman, F. Wilczek, and A. Zee, *Phys. Rev. D* **19**, 1036 (1979); S. Dimopoulos and L. Susskind, *Phys. Rev. D* **18**, 4500 (1978); J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, *Phys. Lett.* **B80**, 360 (1979); S. Weinberg, *Phys. Rev. Lett.* **42**, 850 (1979).
7. G. 't Hooft, *Phys. Rev. Lett.* **37**, 8 (1976).
8. V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, *Phys. Lett.* **155B**, 36 (1985).

5 INFLATION

The cosmology developed in the first four sections (augmented with a model for the growth of structure as discussed elsewhere in this book) provides a remarkably simple and beautiful model to describe the Universe. Nevertheless, there are some aspects of the standard picture that strongly suggests that the model is not a complete one. After discussing the problems of the cosmology developed so far, a possible solution to the problems will be presented. This solution goes by the name of "inflation."[1]

5.1 Loose Ends of the Standard Cosmology

• *Large-Scale Smoothness*: The Robertson-Walker metric describes a space that is homogeneous and isotropic. Why is space homogeneous and isotropic? There are other possibilities, including homogeneous but anisotropic spaces, and inhomogeneous spaces. The most precise indication of the smoothness of the Universe is provided by the microwave background radiation. If the entire observable Universe was in causal contact when the radiation last scattered, it might be imagined that microphysical processes would have damped any fluctuations and a single temperature would have obtained. However in the standard cosmology the distance to the horizon increases with time. The size of the horizon is conveniently expressed in terms of the entropy within the horizon

$$S_H = s\frac{4\pi}{3}d_H^3 \simeq T^3t^3. \tag{5.1}$$

The entropy within the horizon today is $S_H(0) \simeq 10^{88}$. In a matter-dominated Universe, $S_H = S_H(0)(1+z)^{-3/2}$, while in a radiation-dominated Universe, $S_H = S_H(0)(1+z)^{-3}$. The entropy in the horizon at recombination when the radiation last scattered was $S_H(t = t_{rec}) \simeq 10^{83}$. The Universe as presently observed consisted of about 10^5 causally disconnected regions at recombination, so causal processes could not have led to smoothness. At the time of primordial nucleosynthesis, the entropy within the horizon volume was $S_H(t_{nucleo}) \simeq 10^{63}$, or about 10^{-30} of the present Universe.

The first untidy fact about the standard cosmology is that there is no physical explanation for why the Universe is smooth.

• *Density Perturbations*: Although the Universe is smooth on large scales, there is a rich structure on small scales. It is usually assumed that the structures observed today were once small perturbations on a smooth background, and have grown as the result of the gravitational instabilities in an expanding Universe. The relic photons did not take part in the gravitational collapse, and remain as fossil evidence of the once-smooth Universe.

Density inhomogeneities are usually expressed in a Fourier expansion

$$\left(\frac{\delta\rho}{\rho}\right) = (2\pi)^{-3}\int \delta_k \exp(-i\vec{k}\cdot\vec{x})d^3k. \qquad (5.2)$$

Here k is a co-moving label. The *physical* wavenumber and wavelength are related to k by $k_{ph} = k/R(t)$, $\lambda_{ph} = (2\pi/k)R(t)$. It is also convenient to express the scale of the perturbation in terms of the mass in baryons contained within the perturbation. For constant B, the baryon mass on scale λ is proportional to λ^3. The baryon mass within the horizon at time t is $M_H(t) \simeq m_p B s d_H^3 \propto S_H$. The quantity usually referred to as $(\delta\rho/\rho)$ on a given scale is the r.m.s. mass fluctuations on that scale

$$\left(\frac{\delta\rho}{\rho}\right)_k^2 = (2\pi)^{-3}k^3|\delta_k|^2. \qquad (5.3)$$

The exact nature of the perturbations required for galaxy formation is unknown. A promising choice for density perturbations is that as every distance scale comes within the horizon, the r.m.s. fluctuations in the density are $10^{-4} - 10^{-5}$ *independent of the scale.* This is usually expressed as

$$\left(\frac{\delta\rho}{\rho}\right)_H \simeq 10^{-4}. \qquad (5.4)$$

Here $(\delta\rho/\rho)_H$ is $(\delta\rho/\rho)$ on the scale $\lambda = d_H = t$ at time $t = d_H$.

The evolution of the perturbations within the horizon is determined by local physics, e.g., the Jeans criteria. The behavior of the perturbations outside of the horizon is complicated by the fact that there is a "gauge dependence" that reflects the freedom of the choice for a reference spacetime. Nevertheless, the growth of metric perturbations on scales larger than the horizon can be studied by using the uniform Hubble flow gauge (time slices chosen to give constant H). From the Friedmann equation with H constant, fluctuations in ρ are equivalent to fluctuations in the spatial curvature k/R^2

$$\delta\left(\frac{k}{R^2}\right) \Longleftrightarrow \delta\left(\frac{8\pi G}{3}\rho\right). \qquad (5.5)$$

In a radiation-dominated (matter-dominated) Universe, $\rho \propto R^{-4}$ (R^{-3}), so

$$(\delta\rho/\rho) \propto \begin{cases} R^{-2}/R^{-4} \sim (1+z)^{-2} & \text{(RD)} \\ R^{-2}/R^{-3} \sim (1+z)^{-1} & \text{(MD)}. \end{cases} \qquad (5.6)$$

Since $S_H \propto (1+z)^{-3}$ for (RD) and $S_H \propto (1+z)^{3/2}$ for (MD), $(\delta\rho/\rho) \propto S_H^{2/3} \propto M_H^{2/3}$ for both (RD) and (MD). So as any scale comes within the horizon, the growth that scale has experienced while outside the horizon depends upon the mass contained in the scale as it enters the horizon

$$\left(\frac{\delta\rho}{\rho}\right)_H \sim \left(\frac{\delta\rho}{\rho}\right)_0 (M_H(t))^{2/3}, \qquad (5.7)$$

where t_0 is some arbitrary initial time. If $(\delta\rho/\rho)_0$ is proportional to $M^{-2/3}$, as each scale comes within the horizon, $(\delta\rho/\rho)$ will be a constant. Larger scales have smaller initial amplitudes, but they have a longer time to grow outside the horizon. If $(\delta\rho/\rho)_0$ is characterized by a steeper spectrum, the first scales that come within the horizon would have been non-linear. If $(\delta\rho/\rho)_0$ is characterized by a flatter spectrum, larger scales would have larger $(\delta\rho/\rho)$ at horizon crossing.

The standard model can shed no light on the origin of the density perturbations. It must simply assume that at $t = 0$ there are perturbations of the appropriate magnitude and spectrum impressed upon the metric.

• *Spatial Flatness - Age*: In the standard Friedmann cosmology, $\Omega - 1 = k/R^2H^2$. In the past, $H^2 \propto \rho$, which for a matter-dominated Universe gives $H^2 \propto R^{-3}$, and for a radiation-dominated Universe gives $H^2 \propto R^{-4}$. Since today $|\Omega - 1|$ is of order unity, at previous epochs

$$|\Omega - 1| \simeq \begin{cases} R/R_0 = (1+z)^{-1} & \text{(MD)} \\ (R/R_0)^2 = (1+z)^{-2} & \text{(RD)}. \end{cases} \qquad (5.8)$$

At the time of primordial nucleosynthesis, $|\Omega - 1| \leq 10^{-16}$, and at the planck time $|\Omega - 1| \leq 10^{-60}$. Obviously Ω was *very* close to one at early times, i.e., the curvature term was small compared to H^2 and $8\pi G\rho/3$.

The smallness of the curvature term is necessary for the Universe to survive as long as it has without either re-collapsing (for $k = +1$) or becoming curvature dominated (for $k = -1$). The natural time scale in the Friedmann equation is the planck time $t_{Pl} = 2\times 10^{-43}$ sec. The difference between the kinetic term (H^2) and the potential term $(8\pi G\rho/3)$ is the curvature term. This must be small in order for the Universe to expand for 10^{17} sec. $\sim 10^{60}t_{Pl}$.

The standard Friedmann model has no explanation for the present spatial flatness of the Universe.

• *Cosmological Constant*: The most general form of Einstein's equations includes a cosmological constant

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu} + \Lambda g_{\mu\nu}. \qquad (5.9)$$

With the stress-tensor in the perfect-fluid form (U_μ is the fluid velocity vector, $U_\mu = (1,0,0,0)$ in the fluid rest frame)

$$T_{\mu\nu} = -p g_{\mu\nu} + (\rho + p)U_\mu U_\nu, \qquad (5.10)$$

the effect of the cosmological constant is to add to the fluid contributions to ρ and p, terms $\rho_\Lambda = -p_\Lambda = \Lambda/8\pi G$. The generalized energy and pressure are given by $\rho^* = \rho + \rho_\Lambda$, $p^* = p + p_\Lambda$, and the Einstein equations can be written in terms of $T^*_{\mu\nu}$, which is Eq. 5.10 with $\rho \to \rho^*$, $p \to p^*$,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T^*_{\mu\nu}. \tag{5.11}$$

If ρ^* and p^* are dominated by ρ_Λ and p_Λ, the conservation and Friedmann equations become

$$\begin{aligned} \rho^* &\propto R^0 = \text{constant} \\ H^2 &= \frac{8\pi G\rho^*}{3} = \frac{\Lambda}{3}, \end{aligned} \tag{5.12}$$

which has solution $R \propto \exp(Ht)$.

Today the contribution of a cosmological constant to the energy density of the Universe must be less than ρ_C. In useful units, $\rho_C = 8.07 \times 10^{-47}h^2$ GeV^4. Among the contributions to Λ are contributions from the condensates of Higgs particles due to SSB. During cosmological phase transitions, the vacuum energy density changes by σ^4, where σ is the zero-temperature vacuum expectation value of the Higgs field. This change in the vacuum energy is 10^9GeV^4 for the electroweak transition, and 10^{60}GeV^4 for the GUT transition. A cosmological constant of this order must be present before the transition to ensure that after all transitions are complete the energy density of the vacuum is less than about 10^{-47}GeV^4.

The standard cosmology cannot explain why the present vacuum energy density is so small.

• *Unwanted Relics*: There are a variety of particles that are expected to survive annihilation and contribute to the present energy density. Particles with very large masses typically have very small annihilation cross sections and should be abundant. This is rather unfortunate, as their contribution to the mass density typically is many orders of magnitude larger than ρ_C. The magnetic monopoles produced in the GUT phase transition are an example of such an unwanted relic.

The standard cosmology has no mechanism of ridding the Universe of unwanted particles.

The problems mentioned here do not invalidate the standard cosmology. They are accommodated by the standard cosmology, but they are not explained. The goal of cosmology is to explain the present structure of the Universe on the basis of physical law, and one hopes that physical law will one day explain the above points. Inflation is a model for such an explanation.

5.2 New Inflation - The Basic Picture

Consider as a model for new inflation[2,3] a phase transition associated with SSB with a scalar potential given by

$$V(\phi) = \frac{1}{4}\lambda(\phi^2 - \sigma^2)^2. \tag{5.13}$$

At temperatures $T \gg T_c = 2\sigma$, $\langle\phi\rangle = 0$, and $V(\langle\phi\rangle) = \lambda\sigma^4/4 \equiv \rho_V$. At temperatures $T \ll T_c$, $\langle\phi\rangle = \sigma$, and $V(\langle\phi\rangle) = \rho_V = 0$. New inflation will occur as ϕ makes the transition from the high temperature minimum of the potential to the low temperature minimum of the potential.

At some temperature $T \leq T_c$, in some region of the Universe, the Higgs field will make the transition from $\phi = 0$ to $\phi \neq 0$. Assume that in this region of the Universe ϕ is spatially uniform. The evolution of ϕ to the low-temperature ground state is not instantaneous, but requires a time determined by the dynamics of the theory. The equation of motion for ϕ can be found from $T^{\mu\nu}_{;\nu} = 0$, where $T_{\mu\nu} = -\partial_\mu\phi\partial_\nu\phi - \mathcal{L}g_{\mu\nu}$. With the assumption that ϕ is spatially homogeneous

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0, \tag{5.14}$$

where $V' = \partial V/\partial\phi$, and $H^2 = 8\pi G\rho/3$. The contributions to ρ include a radiation term ρ_R, a kinetic term for ϕ, and a potential term for ϕ:

$$\rho = \rho_R + \frac{1}{2}\dot{\phi}^2 + V(\phi). \tag{5.15}$$

If there is a "flat" region in $V(\phi)$, the evolution of ϕ will be "slow" and the $\ddot{\phi}$ term can be neglected in the equation of motion. In this flat region ϕ will change very slowly and $V(\phi)$ will be roughly constant. Therefore the contribution to ρ from $V(\phi)$ will be roughly constant and will rapidly come to dominate ρ_R which decreases in proportion to R^4. When ρ is dominated by potential energy the scale factor increases exponentially. If this flat region in the potential extends from ϕ_i to ϕ_e, R will increase by an amount

$$R(\phi_e) = R(\phi_i)\exp(H\Delta t), \tag{5.16}$$

where Δt is the time it takes to make the transition from ϕ_i to ϕ_e, and $H^2 \sim V(\phi)/m_{Pl}^2 \simeq \sigma^4/m_{Pl}^2$. For a concrete example, assume for the moment that $\Delta t = 100H^{-1}$.

Now assume that after traversing the flat region in the potential, at $\phi \geq \phi_e \simeq \sigma$ there is a "steep" region in the potential. In this steep region the oscillations in the zero momentum mode of ϕ will rapidly convert the potential energy to radiation. If this conversion is efficient, the Universe will be reheated to a temperature T_{RH}

found by equating the potential energy density to the radiation energy density: $V(\phi) \simeq T^4$, or $T_{\rm RH} \simeq \sigma$.

This is the basis scenario for new inflation. To illustrate the scenario, take $\sigma = 10^{14}$GeV, and the initial size of the region to be the size of the horizon at T_c, $R_i = H^{-1} \simeq m_{Pl}/\sigma^2 = 10^{-23}$cm. [16] The initial entropy in this region is $S_i \simeq (R_i T_i)^3 \simeq 10^{14}$. The final size of the region in the example where $\Delta t = 100 H^{-1}$ is $R_f = \exp(100) R_i \simeq 3 \times 10^{20}$cm. With efficient reheating $T_{\rm RH} = \sigma$, and the final entropy contained in the region is $S_f = (R_f T_{\rm RH})^3 \simeq 10^{144}$.

This large creation of entropy has helped with three out of four problems. *Large-Scale Smoothness*: At $T = 10^{14}$GeV, the presently observable Universe ($S = 10^{88}$) was contained in a size of 10cm, and easily fit within the smooth region after inflation. *Density Perturbations*: To see how inflation generates density perturbations it is necessary to treat the dynamics of the scalar field in greater detail than done so far. This will be done shortly. *Spatial Flatness - Age*: After inflation R has increased by exp(100) but the final temperature is close to the initial temperature. Thus, immediately after inflation the spatial curvature term k/R^2 is a factor of exp(−200) smaller, while the energy density term is unchanged. *Cosmological Constant*: Inflation does not help the cosmological constant problem. *Unwanted Relics*: The number density of particles present before inflation is decreased by a factor of $R_i^3/R_f^3 \simeq \exp(-300)$. This is true also for the original photons. It is crucial to create entropy in the termination of inflation.

In this example it was assumed that the slow-roll period lasted for 100 e-folds. The minimum number of e-folds is the number required to fit the observed entropy of 10^{88} into a single inflation region. The final entropy in the inflation region is $S_f \simeq T_{\rm RH}^3 R_f^3$. The size of the final region is related to the number of e-folds by $R_f^3 = \exp(3N) R_i^3$, assuming little or no growth during the oscillation phase. The largest possible smooth initial region is the size of the horizon at the phase transition, $R_i = H^{-1}(T_c) \simeq m_{Pl}/\sigma^2$, assuming $T_c = \sigma$. The maximum reheat temperature is $T_{\rm RH} \simeq \sigma$, so the final entropy is $S_f \simeq \sigma^3 \exp(3N) m_{Pl}^3/\sigma^6 \simeq \exp(3N) m_{Pl}^3/\sigma^3$. The requirement $S_f \geq 10^{88}$ gives

$$N \geq 58 + \ln(\sigma/10^{15}{\rm GeV}). \tag{5.17}$$

5.3 Dynamics of Inflation

The evolution of the spatially homogeneous scalar field (zero momentum mode of the scalar field) is crucial for inflation. If the coupling of the scalar field to other fields are included, the equation of motion for the zero-momentum mode of ϕ is (ϕ will denote the zero-momentum mode unless otherwise indicated)

$$\ddot{\phi} + 3H\dot{\phi} + \Gamma_\phi \dot{\phi} + V'(\phi) = 0, \tag{5.18}$$

[16] It is reasonable to expect ϕ to be uniform on scales that are in causal contact.

where Γ_ϕ is the ϕ decay width. The decay width is typically $\Gamma_\phi = h^2 m_\phi$, where h is a coupling constant, and m_ϕ is the mass of ϕ. [17] The energy density and pressure of ϕ are given by

$$\begin{aligned} \rho_\phi &= \frac{1}{2}\dot{\phi}^2 + V(\phi) \\ p_\phi &= \frac{1}{2}\dot{\phi}^2 - V(\phi). \end{aligned} \tag{5.19}$$

The "slow roll" regime is when the $\ddot{\phi}$ and Γ_ϕ terms in Eq. 5.18 can be neglected, and $V(\phi)$ is the dominant term in Eq. 5.15. The equation of motion during slow roll is

$$3H\dot{\phi} = -V'(\phi). \tag{5.20}$$

Neglecting $\ddot{\phi}$ is consistent if

$$\begin{aligned} |V''(\phi)| &\leq 9H^2 \\ |V'(\phi) m_{Pl}/V(\phi)| &\leq (48\pi)^{1/2}. \end{aligned} \tag{5.21}$$

These conditions will determine the duration of slow roll.

The number of e-folds of inflation while ϕ rolls from ϕ_1 to ϕ_2 during slow roll is given by

$$N(\phi_1 \to \phi_2) = \int_{t_1}^{t_2} H\,dt = -3\int_{\phi_1}^{\phi_2} \frac{H^2(\phi)}{V'(\phi)} d\phi, \tag{5.22}$$

where $dt = \dot{\phi}^{-1} d\phi = -3H/V' d\phi$.

With ρ_ϕ given by Eq. 5.19, $\dot{\rho}_\phi = \dot{\phi}\ddot{\phi} + \dot{\phi}V'(\phi)$, and using Eq. 5.18, $\dot{\rho}_\phi = -3H\dot{\phi}^2 - \Gamma_\phi\dot{\phi}^2$. The two terms in the equation for $\dot{\rho}_\phi$ represent the change due to the redshift of the kinetic energy in the ϕ field (proportional to H) and the change due to decay of the ϕ field (proportional to Γ_ϕ). When ϕ starts oscillating about the minimum of the potential, the energy transfers between kinetic and potential energy until ϕ decays. Over an oscillation cycle $\langle\dot{\phi}^2\rangle = \rho_\phi$, and $\dot{\phi}^2$ can be replaced by ρ_ϕ in the equation for $\dot{\rho}_\phi$. The energy from ϕ decay is transferred into radiation, and the equation for the evolution of ρ_R becomes $\dot{\rho}_R = -4H\rho_R + \Gamma_\phi\rho_\phi$.

The equations for $\dot{\rho}_R$ and $\dot{\rho}_\phi$ can be integrated to study reheating. If oscillation about the minimum begins at $t = t_3$ and $R = R_3$ with $\rho_\phi = \sigma^4$, the ϕ energy density will decrease as

$$\rho_\phi = \sigma^4 \left(\frac{R}{R_3}\right)^{-3} \exp[-\Gamma(t - t_3)]. \tag{5.23}$$

[17] It is crucial to remember that $m_\phi \equiv \partial^2 V(\phi)/\partial\phi^2$ is a function of ϕ, and will be small in the flat region of the potential.

Until decay, the ϕ energy density decreases in expansion as the energy density for massive particles. When ϕ decays, the remaining energy is converted to radiation ($\rho_\phi \to (\pi^2/30)g_* T_{RH}^4$). Obviously, the longer ϕ oscillates before decay, the less energy will be available for conversion into radiation, and the lower will be the reheat temperature. If the decay width is large compared to the expansion rate at the start of oscillation, $H_3 \simeq \sigma^2/m_{Pl}$, reheating will occur before damping of ρ_ϕ, and $T_{RH} \simeq g_*^{-1/4}\sigma$. If the decay width is small compared to H_3, ϕ will oscillate until the age of the Universe is equal to the ϕ lifetime, i.e., until $\Gamma_\phi = H = \rho_\phi^{1/2}/m_{Pl}$. Then when $\rho_\phi \to g_* T_{RH}^4$, the reheat temperature will be $T_{RH} = g_*^{-1/4}\rho_\phi^{1/4} = g_*^{-1/4}(\Gamma_\phi m_{Pl})^{1/2}$.

Now consider the generation of fluctuations in ρ. In the FRW radiation-dominated Universe $H^{-1} \propto t$, while during the slow-roll epoch, $H^{-1} \simeq m_{Pl}/V(\phi)^2 \simeq$ constant. If H is constant, the Universe in approximately in a de Sitter phase. H^{-1} sets the scale over which microphysical processes can act. H^{-1} will be called the "physics horizon." [18] During the slow roll phase the physics horizon is constant and physical scales increase exponentially. Eventually, physical scales once smaller than the horizon, will become larger than the horizon. After termination of the slow-roll phase the Universe reheats, behaves like a FRW radiation-dominated Universe, and scales outside the horizon will eventually come (back) within the horizon. This double-cross of the physics horizon is illustrated in Fig. 24.

Notice that the *last* scales to go outside H^{-1} during the de Sitter phase are the *first* scales to come back inside H^{-1} during the FRW phase. Ignoring the σ dependence in Eq. 5.17, the Hubble radius today ($\simeq$ 3000 Mpc) crossed the horizon 58 e-folds before the end of inflation. Any scale smaller than the Hubble radius today crossed the horizon $58 + \ln(\sigma/10^{15}\text{GeV}) + \ln(\lambda/3000\text{Mpc})$ e-folds before the end of inflation. Using $B = 10^{-10}$, the mass in baryons inside the horizon today is $10^{21} M_\odot$. Since $B \propto \lambda^3$, any baryon mass scale crossed the horizon $58 + \ln(\sigma/10^{15}\text{GeV}) + (1/3)\ln(M/10^{21}M_\odot)$ e-folds before the end of inflation. Scales that will eventually contain a galaxy mass ($M = 10^{11} M_\odot$) crossed the horizon 50 e-folds before the end of inflation, while scales that will eventually contain a galaxy cluster mass ($M = 10^{14}M_\odot$) crossed the horizon 53 e-folds before the end of inflation.

So far it has been assumed that the ϕ field is constant. However there are quantum fluctuations in ϕ due to the fact that during the slow-roll epoch the Universe is approximately in a de Sitter phase. If the fluctuations $\delta\phi$ are expressed as a Fourier expansion

$$\delta\phi = (2\pi)^{-3}\int d^3k\,\delta\phi_k \exp(-i\vec{k}\cdot\vec{x}), \tag{5.24}$$

[18] Note the difference between the "physics horizon" (H^{-1}) and the particle horizon, which is the distance a massless particle propagates from the time of the bang. The physics horizon is the relevant quantity in calculation of perturbations.

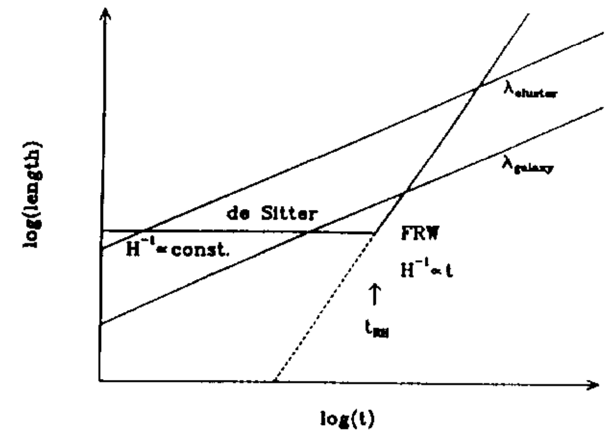

Figure 24: Physical scales cross the physics horizon twice

then the de Sitter fluctuations result in (note: $\Delta\phi \equiv \Delta\phi_k$)

$$(\Delta\phi)^2 \equiv (2\pi)^{-3}k^3|\delta\phi_k|^2 = \left(\frac{H}{2\pi}\right)^2. \tag{5.25}$$

These fluctuations obtain on scales less than the physics horizon during the de Sitter phase. As each scale goes outside the horizon during slow roll, it has fluctuations $(\Delta\phi)^2 = (H/2\pi)^2$. Since the energy density depends upon ϕ, the fluctuations in ϕ lead to fluctuations in ρ of $\delta\rho = (\partial V/\partial\phi)\Delta\phi$. Using $\rho \simeq V \simeq \sigma^4$ and $\partial V/\partial\phi = -3H\dot\phi$, fluctuations in ϕ lead to

$$\left(\frac{\delta\rho}{\rho}\right)_k \simeq \left(\frac{\dot\phi H^2}{\sigma^4}\right). \tag{5.26}$$

Once the scale is larger than H^{-1}, it can no longer be affected by microphysics. The behavior of the perturbation outside the horizon is gauge-dependent. However the behavior outside the horizon can be characterized by a parameter ς, given by

$$\varsigma \equiv \frac{\delta\rho}{\rho + p} \simeq \begin{cases} \delta\rho/\rho & \text{FRW} \\ -3H\dot\phi\Delta\phi/\dot\phi^2 & \text{de Sitter.} \end{cases} \tag{5.27}$$

When a scale comes back within the horizon during the FRW phase, ς is the same as when it first went outside the horizon during inflation. Therefore, $(\delta\rho/\rho)$

relevant for galaxy formation is given by [19]

$$\left(\frac{\delta\rho}{\rho}\right)_H \simeq \left(\frac{-3H\dot{\phi}\Delta\phi}{\dot{\phi}^2}\right)_H \simeq \left(\frac{H^2}{\dot{\phi}}\right)_H . \tag{5.28}$$

With the approximation that H and $\dot{\phi}$ are constant during the slow-roll phase, $(\delta\rho/\rho)$ as it re-enters the horizon will be scale free . In the slow-roll period, $\dot{\phi} = -V'(\phi)/3H$, and the equation for $(\delta\rho/\rho)$ becomes

$$\left(\frac{\delta\rho}{\rho}\right)_H \simeq \left(\frac{-3H^3}{V'(\phi)}\right). \tag{5.29}$$

5.4 Specific Models

The first example considered is the original attempt to implement new inflation. The model is based upon a SU_5 GUT with symmetry breaking via the Coleman-Weinberg mechanism.[2,3] The scalar field responsible for inflation (hereafter referred to as the *inflaton*) is in the 24-dimensional representation of SU_5 and is responsible for the symmetry breaking $SU_5 \to SU_3 \times SU_2 \times U_1$. Let ϕ denote the magnitude of the Higgs field in the $SU_3 \times SU_2 \times U_1$ direction. The one-loop, zero-temperature Coleman-Weinberg potential is

$$V(\phi) = B\sigma^4/2 + B\phi^4\left[\ln(\phi^2/\sigma^2) - 1/2\right], \tag{5.30}$$

where $B = 25\alpha_{GUT}^2/16 \simeq 10^{-3}$, and $\sigma = 2 \times 10^{15}$GeV. Because of the absence of a mass term, the potential is very flat near the origin (SSB arises due to one-loop radiative corrections). For $\phi \ll \sigma$, the potential may be approximated in the slow-roll regime by

$$\begin{aligned} V(\phi) &\simeq B\sigma^4/2 - \lambda\phi^4/4 \\ \lambda &\simeq |4B\ln(\phi^2/\sigma^2)| \simeq 0.1. \end{aligned} \tag{5.31}$$

For $\phi \ll \sigma$

$$\begin{aligned} V(\phi) &\simeq B\sigma^4/2 \\ H^2 &= \frac{8\pi G\rho}{3} \simeq \frac{4\pi}{3}\frac{B\sigma^4}{m_{Pl}^2}. \end{aligned} \tag{5.32}$$

The critical temperature for this potential is about 10^{14}GeV. The finite temperature potential has a small temperature-dependent barrier near the origin, and it is not until $T = 10^9$GeV or so that this barrier is low enough that the action for bubble nucleation drops to order unity. At this time the Universe will undergo "spinodal decomposition" and break up into irregularly shaped fluctuation regions within which ϕ is approximately constant.

Consider the evolution of ϕ in the slow-roll regime. Slow roll commences at ϕ_s and ends at ϕ_e. The end of slow roll is determined by the condition $|V''(\phi_e)| = 9H^2$, or $\phi_e^2 = 3H^2/\lambda$. For any ϕ in the region $\phi_s \le \phi \le \phi_e$, the number of e-folds from ϕ to ϕ_e (time t to time t_e) is given by

$$N(\phi \to \phi_e) = \int_t^{t_e} H\,dt = \int_\phi^{\phi_e} H\dot{\phi}^{-1}d\phi. \tag{5.33}$$

Using $3H\dot{\phi} = -dV/d\phi$ during slow roll,

$$N(\phi \to \phi_e) = \frac{3H^2}{2\lambda}\left(\frac{1}{\phi^2} - \frac{1}{\phi_e^2}\right). \tag{5.34}$$

The total number of e-folds in slow roll depends upon ϕ_s. To have enough inflation, $N(\phi_s \to \phi_e)$ must be greater than 58. Since λ is 10^{-1}, ϕ_s must be smaller than H in order to have sufficient e-folds. However de Sitter space fluctuations introduce uncertainties in ϕ of this order. The quantum fluctuations may prematurely terminate inflation. At the very least they suggest that the semiclassical equations of motion may be invalid.

A more serious problem is the magnitude of the density fluctuations.[4,5,6,7] During slow roll for the Coleman-Weinberg potential $V'(\phi) \simeq \lambda\phi^3$, and Eq. 5.29 gives

$$\left(\frac{\delta\rho}{\rho}\right)_H \simeq \frac{3H^3}{\lambda\phi^3} \simeq \left(\frac{\lambda}{3}\right)^{1/2}[2N(\phi \to \phi_e)]^{3/2}, \tag{5.35}$$

where Eq. 5.34 has been used to express ϕ in terms of the number of e-folds before the end of inflation. Although $(\delta\rho/\rho)$ depends upon N to a power, N depends upon the *logarithm* of the length or mass scale, so the scale dependence of $(\delta\rho/\rho)$ is only logarithmic. The problem with the Coleman-Weinberg potential is not the spectrum, but the magnitude of the perturbations. Using $\lambda = 0.1$ and $N(\phi \to \phi_e) = 58 + (1/3)\ln(M/10^{21}M_\odot)$, $(\delta\rho/\rho)_H$ on the scale of galaxies is 182, and on the scale of clusters is 199. The spectrum is very flat, but about 10^6 too large. Notice that a smaller λ cures both problems.

Although the original model for new inflation was a failure, it pointed the way for the construction of somewhat more successful models. The potential of the original Coleman-Weinberg model was not flat enough, i.e., λ was too large. If ϕ couples to gauge fields, λ will be of order α_{GUT}^2, which is too large. If ϕ is a weakly-coupled gauge singlet, the effective λ can be small, and will remain small

[19] There should be no confusion between the sub-H which indicates the quantity is to be evaluated at the time of horizon crossing, and the expansion rate H.

after radiative corrections. If $\lambda \leq 10^{-12}$, the density perturbations from Eq. 5.35 will be small enough. However a weakly-coupled inflaton will have a small decay width, and the reheat temperature will be low. If λ is also the magnitude for the coupling of the inflaton to other fields, the decay width at the minimum will be $\Gamma_\phi \simeq \lambda^2 m_\phi \simeq \lambda^2\sigma$, and the reheat temperature will be $T_{\rm RH} \simeq (\Gamma_\phi m_{Pl})^{1/2} \simeq 10^8$GeV for $\lambda = 10^{-12}$ and $\sigma = 10^{15}$GeV. A more careful calculation may give one or two orders of magnitude larger value of $T_{\rm RH}$, but it is clear that a weakly-coupled field will have a low reheat temperature. This presents a problem for baryogenesis. Any baryon asymmetry present before inflation will be diluted due to the creation of the large amount of entropy, so it is necessary to create the baryon asymmetry either during or after the reheating epoch. Many inflation models are squeezed between the requirement of a weakly coupled inflaton for a flat potential and an inflaton that has a large enough decay width to give $T_{\rm RH}$ large enough for baryogenesis.

Supersymmetric models have been proposed as a mechanism to stabilize small couplings in the inflaton potential against radiative corrections. Supersymmetric models introduce several additional potential problems. The high-temperature minimum of the potential is generally not at $\phi = 0$, and $\langle\phi\rangle$ may smoothly evolve to the zero-temperature minimum. There are two possible solutions. If the high-temperature minimum is at $\phi \leq 0$, there will always be a barrier between the high-temperature and the low-temperature minimum. The other solution is to ignore the problem. Since the inflaton must be weakly coupled, it may never be in LTE, and the initial value of ϕ may be random. Another problem with supersymmetric models is the gravitino problem. Gravitinos are weakly-interacting, long-lived particles present in supersymmetric models. They will be produced in reheating in embarrassingly large numbers unless the reheat temperature is less than about 10^9GeV. Finally, in supersymmetric models where supersymmetry breaking is done with a Polonyi field, the Polonyi field can be set into oscillations that will not decay because the Polonyi field is "hidden." Since the energy density in the oscillating field behaves like non-relativistic matter, it will eventually come to dominate the Universe.

For successful new inflation, several requirements must be fulfilled. The requirements occur during different periods of inflation.[8]

• *Start Inflation*: The scalar field must be smooth in a region such that the energy density and pressure associated with spatial gradients in ϕ are smaller than the potential energy. If the average value of ϕ is ϕ_0 and the region has typical spatial dimension L, this requirement implies

$$(\nabla\phi)^2 = O(\phi_0/L) \ll V(\phi_0) = O(\sigma^4). \tag{5.36}$$

If this requirement is not met and the $(\nabla\phi)^2$ term dominates, $R(t)$ will expand as t to a power and inflation will not occur. However once $V(\phi)$ does dominate, the gradient terms rapidly become small in the exponential expansion and can be ignored.

In supersymmetric models where LTE is obtained, the high-temperature minimum of $V(\phi)$ should be at $\phi \leq 0$ to prevent ϕ from smoothly evolving to the zero-temperature minimum without inflating.

• *Start Slow Roll*: If ϕ is not a gauge singlet it may roll in the "wrong" direction. For instance for the Coleman-Weinberg SU_5 model, the steepest direction for ϕ near the origin is toward a minimum where $SU_4 \times U_1$ is the unbroken symmetry. If ϕ is a gauge singlet there is no problem with ending up in the wrong phase.

In order to have slow roll, the potential must have a flat region in which $|V''(\phi)| \leq 9H^2$ and $|V'(\phi)m_{Pl}/V(\phi)| \leq (48\pi)^{1/2}$.

• *Roll Far Enough*: The interval of slow roll, $[\phi_s, \phi_e]$, must be large enough that quantum fluctuations do not terminate slow roll. This condition will be met if $\phi_e - \phi_s \gg H$.

The number of e-folds, $N = \int H dt$ from ϕ_s to ϕ_e, must be greater than $58 + \ln(\sigma/10^{16}\text{GeV})$.

• *Small Perturbations*: The magnitude of the perturbations must be less than of order 10^{-4} on the scale of galaxies to clusters in order to avoid large fluctuations in the MBR. If the fluctuations produced in inflation are to lead to structure formation, they should be greater than of order 10^{-6}. Therefore during slow roll $H^2/\dot\phi \leq 10^{-4}$.

In addition to the scalar perturbations discussed so far, inflation will produce tensor perturbations. These tensor perturbations can be thought of as gravity waves. As each scale leaves the horizon during inflation the energy density of gravity waves on that scale is $\rho_{GW} \simeq H^4$. In terms of a dimensionless amplitude $h = H/m_{Pl}$ and wavelength λ, $\rho_{GW} \simeq (m_{Pl}^2 h^2/\lambda^2)_{\lambda=H^{-1}}$. These gravity waves will re-enter the horizon during the FRW phase with the same dimensionless amplitude h, and induce an anisotropy in the MBR of order h. For $\delta T/T \leq 10^{-4}$, $h = H/m_{Pl} \leq 10^{-4}$. Since $H \simeq \sigma^2/m_{Pl}$, σ must be less than about 10^{17}GeV.

• *Exit Properly*: The reheat temperature must be high enough so the Universe is radiation dominated during primordial nucleosynthesis. Using $T_{\rm RH} = (\Gamma_\phi m_{Pl})^{1/2}$, $T_{\rm RH} \geq 1$ MeV requires $\Gamma_\phi \geq 10^{-25}$GeV. If baryogenesis proceeds in the standard way, then $T_{\rm RH} \geq 10^9$GeV, which implies $\Gamma_\phi \geq 10^{-1}$GeV. In order to ameliorate the problem of low reheat temperature and baryogenesis, it has been proposed that a baryon asymmetry is created by the decay of the inflaton. The energy density in the coherent oscillations can be thought of as due to a collection of zero momentum inflatons with number density $n_\phi = \rho_\phi/m_\phi$. In reheating, $\rho_\phi \rightarrow g_* T_{\rm RH}^4$, so $n_\phi = g_* T_{\rm RH}^4/m_\phi$ at reheating. Suppose the inflaton decays into a particle, S, which, in turn, decays out of equilibrium with baryon number violation. The number density of S's that decay is the same as the number density of parent inflatons. If the CP parameter in the decay of the S is ϵ, then the asymmetry in baryons produced by the S is $n_B = \epsilon n_\phi = \epsilon g_* T_{\rm RH}^4/m_\phi$. The entropy density produced after thermalization of the inflaton decay products is $s = g_* T_{\rm RH}^3$. Therefore $B \equiv n_B/s = \epsilon T_{\rm RH}/m_\phi$. If $B \geq 10^{-10}$, then $T_{\rm RH} \geq 10^{-10} m_\phi/\epsilon$.

EPOCH	PROBLEM	POSSIBLE SOLUTION
Start	ϕ Smooth	$(\nabla\phi)^2 \ll V(\phi)$
Inflation	Thermal Constraint	$(\phi) \leq 0$
Execute	Roll in Right Direction	ϕ is gauge singlet
Slow Roll	Flat Region in $V(\phi)$	$\lvert V''(\phi)\rvert \leq 9H^2$, and
		$\lvert V'(\phi)m_{Pl}/V(\phi)\rvert \leq (48\pi)^{1/2}$
Roll Far	Quantum Fluctuations	$\phi_e - \phi_s \gg H$
Enough	Sufficient e-folds	$N = \int H dt \geq 58$
Small	Scalar Perturbations	$(H^2/\dot{\phi}) \leq 10^{-4}$
Perturbations	Tensor Perturbations	$H/m_{Pl} \leq 10^{-4}$
Exit Properly	Nucleosynthesis	$T_{RH} \geq 1$ MeV
	Baryogenesis	$T_{RH} \geq 10^{-10}\epsilon^{-1}m_\phi$
	Gravitinos	$T_{RH} \leq 10^9$GeV

Table 3: Possible problems and solutions in new inflation

There is a model-dependent upper limit on T_{RH} to avoid making unwanted relics. For example, in supersymmetric models, $T_{RH} \leq 10^9$GeV to avoid overproducing gravitinos.

The above problems and some possible solutions are given in Table 3. Although there are models that satisfy all the above requirements, none of them seem so compelling that they must be the final answer. In fact, in the past few years there has been increasing effort in the generalization of inflation as a phenomena that is decoupled from a cosmological phase transition.

5.5 Present Status and Future Directions

Although the general scenario of inflation presents a very attractive means to ameliorate at least some of the untidiness of the standard model, it is by no means clear that all (or even any) problems are solved or understood. It is now clear that there are models, both supersymmetric and non-supersymmetric, which can successfully implement the program of new inflation as outlined above. It is useful to normalize the more non-standard models of inflation by comparing them to these two "standard" models of inflation.

The non-supersymmetric model is a GUT model based upon SU_5. The model was first proposed by Shafi and Vilenkin,[9] and refined by Pi.[10] In the latest version of the model the inflaton is the real part of a complex gauge-singlet field with a Coleman-Weinberg potential of the form in Eq. 5.30, with ϕ representing the magnitude of the complex field, and $B = O(10^{-14})$. It must be assumed that the couplings of the ϕ to all other fields in the theory are less than about 10^{-7} to prevent quantum corrections from spoiling the smallness of B. The real part of ϕ will be the inflaton, and the imaginary part of ϕ will be the axion. ϕ couples to the adjoint Higgs, and induces SU_5 breaking when it receives a VEV. This requires $\sigma = 10^{18}$GeV. Since B is so small (and will remain small after radiative cprrections), the problems with the original Coleman-Weinberg SU_5 model vanish. The reheat temperature is barely high enough to produce a baryon asymmetry through the decay of the inflaton as discussed above. At the expense of introducing a small number, the model is simple and it works.

An example of a supersymmetric model that works was proposed by Holman, Ramond, and Ross.[11] The superpotential in their model has a "inflation sector" with superpotential $I = (\Delta^2/M)(\phi - M)^2$, where $M = m_{Pl}/(8\pi)^{1/2}$. The scalar potential in supersymmetric models is typically an expansion in ϕ/M, given in this case by

$$V(\phi) = \Delta^4(1 - 4\phi^3/M^3 + 6.5\phi^4/M^4 - 8\phi^5/M^5 + \cdots). \quad (5.37)$$

For $\Delta/M \simeq 10^{-4}$, ($\Delta \simeq 2 \times 10^{14}$GeV), density fluctuations are small enough and sufficient e-folds obtain. The decay width of the ϕ (which has only gravitational coupling to other fields) is $\Gamma_\phi \simeq \Delta^6/M^5$, which for Δ small enough to satisfy the perturbations constraint, leads to $T_{RH} \simeq (\Gamma_\phi m_{Pl})^{1/2} \simeq 10^6$GeV. With the baryon asymmetry produced via inflaton decay, this is large enough. At the expense of the introduction of a sector whose sole purpose is inflation, the model is simple and it works.

Both the above models have two potential problems. The first problem is that to this point the calculations of the evolution of the scalar field have been semi-classical. It may be that a true quantum calculation of the evolution of ϕ, including production of density perturbations, will give a result much different than the semi-classical result. Preliminary work on this problem suggests that the semi-classical approximations are reasonable. The second potential problem has to do with the initial value of ϕ. Both fields are extremely weakly coupled and are unlikely ever to be in LTE. There is no reason to assume $\phi \simeq 0$ for an initial condition (in fact, it may not even be the high-temperature minimum for the supersymmetric example). It is tempting to say that this is not a problem, and that it is only necessary for $\phi \simeq 0$ in some region of the Universe where the kinetic contributions to ρ are small enough to start inflation.

The above two models are existence proofs that it is possible to implement new inflation. Whether new inflation is the final answer will be discussed briefly after mentioning some other approaches for inflation that do not involve SSB.

For weakly coupled scalar fields there is no reason to believe the inflaton will be in LTE at high temperature, and the value of ϕ at high temperature might be random (hence the name "chaotic inflation"). Imagine a simple scalar potential of the form $V(\phi) = \lambda\phi^4$, with minimum at $\langle\phi\rangle = 0$. Assume as initial conditions that $\phi = \phi_0 \neq 0$, and that ϕ is sufficiently smooth in a large enough region to inflate. The number of e-folds of inflation is

$$N(\phi \to 0) = \int_\phi^0 H\,dt \simeq \pi \left(\frac{\phi}{m_{Pl}}\right)^2. \tag{5.38}$$

In order to obtain 58 e-folds of inflation, $\phi_0 \gtrsim 4.3 m_{Pl}$. The density perturbations are

$$\left(\frac{\delta\rho}{\rho}\right)_H \simeq \left(\frac{3H^3}{V'(\phi)}\right) \simeq \lambda^{1/2}\left(\frac{\phi}{m_{Pl}}\right)^3 \simeq \lambda^{1/2} N(\phi \to 0)^{3/2}. \tag{5.39}$$

Again, using $N = 50$, λ must be smaller than about 10^{-14} for sufficiently small density perturbations. Since Linde originally proposed this model[12] several refinements have been made. First, it has been shown that it is possible to use a $m^2\phi^2$ potential rather than a $\lambda\phi^4$ potential. Some work has been done in examining and formalizing what exactly is meant by "chaotic" initial conditions, and which regions of phase space will inflate. Linde's model is an example of how general inflation is, and that it is possible, perhaps even desirable, to separate inflation from SSB phase transitions. Chaotic inflation (at least for the $\lambda\phi^4$ case) has the possible problem of using classical gravity in the regime $\phi \geq m_{Pl}$. At present it also has the undesirable feature of involving the dynamics of a scalar field introduced for the sole purpose of inflation.

A model even further from the original idea of an SSB phase transition is a pure gravity model based upon including an ϵR^2 term in the gravity Lagrangian. Such higher-derivative terms are expected to be present in theories with extra dimensions. Mijić, Morris, and Suen[13] have examined this possibility in detail, including questions of density perturbations and reheating and find that all constraints can be met for $10^{11} \leq \epsilon^{-1/2} \leq 10^{13}$GeV.

Yet further from the original idea of inflation is the possibility that the inflaton is related to the size of extra dimensions. This will be discussed in the next section. A possibility not discussed here is the role of quantum gravity and the program of the "wave function of the Universe."

In a Universe without inflation, the space of initial conditions that give the Universe we observe is a set of measure zero. The inflationary Universe enlarges the space of initial data that will lead to the observable Universe. However, it does not imply that every imaginable set of initial data will lead to inflation. A trivial example is a closed Universe that becomes curvature dominated, and collapses before the vacuum energy dominates and causes inflation. The question "is inflation inevitable" has not yet been completely answered. Inflation may be the final answer, part of the final answer, or none of the final answer.

If inflation did occur there are two general predictions. The first prediction is that Ω is very close to 1. It would be hard to imagine that *exactly* 58 e-folds of inflation occurred. With all models that give small density perturbations, the number of e-folds of inflation is enormous, and the intrinsic curvature will only appear on scales far larger than our present horizon. Of course, scale-free density perturbations would appear on the horizon today, so a $(\delta\rho/\rho) \simeq 10^{-4}$ would lead to $\Omega = 1 \pm 10^{-4}$. The second prediction is that of scale-free density perturbations. At present there is no convincing data to support either prediction. Dynamical measurements of Ω seem to give $\Omega = 0.1 \to 0.3$. This has (at least) three possible explanations. Either there are systematic uncertainties in all the measurements, there is unclustered matter (like massless particles) that give the unseen part of Ω, or there is a present vacuum energy that can account for spatial flatness (the actual prediction of inflation) and $\Omega \neq 1$. None of these explanations are compelling. If the recent determination of the velocity field on large-scales are correct, it is evidence against a scale-free spectrum. Possible ways out are the measurements are wrong, cosmic strings, and double inflation.

The last point is that some explanation must be found for the present smallness of the cosmological constant.

5.6 References

For a general review and a complete list of original literature, see M. S. Turner in Ref. [14].

1. A. Guth, Phys. Rev. D **23**, 347 (1981).
2. A. Albrecht and P. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
3. A. Linde, Phys. Lett. **108B**, 389 (1982).
4. A. Starobinsky, Phys. Lett. **117B**, 175 (1982).
5. J. M. Bardeen, P. Steinhardt, and M. S. Turner, Phys. Rev. D **28**, 679 (1983).
6. A. Guth and S.-Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982).
7. S. Hawking, Phys. Lett. **115B**, 295 (1982).
8. P. Steinhardt and M. S. Turner, Phys. Rev. D **29**, 2162 (1984).
9. Q. Shafi and A. Vilenkin, Phys. Rev. Lett. **52**, 691 (1984).
10. S.-Y. Pi, Phys. Rev. Lett. **52**, 1725 (1984).
11. R. Holman, P. Ramond, and G. Ross, Phys. Lett. **137B**, 343 (1984).

12. A. D. Linde, Phys. Lett. **129B**, 177 (1983).

13. M. B. Mijić, M. S. Morris, and W.-M. Suen, "The R^2 Cosmology - Inflation Without a Phase Transition," Caltech Report CATT-68-1320, (Feb. 1986).

14. M. S. Turner, in *Proceedings of the 1984 Cargese School on Fundamental Physics and Cosmology*, ed. J. Audouze and J. Tran Thanh Van (Editions Frontieres, Gif-Sur-Yvette, 1985).

6 COSMOLOGY AND EXTRA DIMENSIONS

In the past few years the search for a consistent quantum theory of gravity and the quest for a unification of gravity with other forces have led to a great deal of interest in theories with extra spatial dimensions (extra time dimensions seem to lead to ghosts). These extra spatial dimensions are unseen because they are compact and small, presumably with typical dimensions of the Planck length, $l_{Pl} = 1.616 \times 10^{-33}$cm. If the "internal" dimensions are static and small compared to the large "external" dimensions the only role they would play in the dynamics of the expansion of the Universe is in determining the structure of the physical laws. However, if the big bang is extrapolated back to the Planck time, then the characteristic size of *both* internal and external dimensions were the same, and the internal dimensions may have had a more direct role in the dynamics of the evolution of the Universe. This chapter presents some speculations about the role of extra dimensions in cosmology.

6.1 Microphysics in Extra Dimensions

Theories that have been formulated in extra dimensions include Kaluza-Klein theories,[1] supergravity theories, [1] and superstring theories.[2] The exact motivation and goals of these approaches are quite different, but for many applications to cosmology they have several common features and they will be referred to simply as theories in extra dimensions. Among the common features of theories in extra dimensions are:

• *There are large spatial dimensions and small spatial dimensions*: If some of the dimensions are compact and smaller than the three large dimensions, it is possible to dimensionally reduce the system (integrate over the extra dimensions) and obtain an "effective" 3+1-dimensional theory. Present accelerators have probed matter at distances as small as 10^{-16}cm without finding evidence of extra dimensions. This is not surprising, as the extra dimensions are expected to have a size characteristic of the Planck length. The large dimensions may also be compact. If so, their characteristic size is greater than the Hubble distance, 10^{28}cm. This disparity of about 61 orders of magnitude is somewhat striking. This disparity is usually posed by the question "what makes the extra dimensions so small?" However, if gravity has anything to do with the size of dimensions, the only reasonable size *is* the Planck length, and a more appropriate question to ask is "what makes the observed dimensions so large?" One possible answer to the the last question is inflation. The possible connection between inflation and extra dimensions will be explored.

• *The effective low-energy theory depends upon the internal space*: In Kaluza-Klein theories the low-energy gauge group is determined by the continuous isometries of the internal manifold. In superstring theories, the structure of the internal space determines the number of generations of chiral fermions, whether there is

THEORY	α/α^0	G/G^0	G_F/G_F^0
Kaluza-Klein	$(b/b_0)^{-2}$	$(b/b_0)^{-D}$	$(b/b_0)^{-2}$
(D internal dimensions)			
Superstrings	$(b/b_0)^{-6}$	$(b/b_0)^{-6}$	$(b/b_0)^{-6}$
(6 internal dimensions)			

Table 4: Variation of fundamental constants with the size of the internal manifold

low-energy supersymmetry, etc. If the internal space is distorted in any way the effective low-energy physics could be very different.

• *The fundamental constants we observe are not truly fundamental*: In theories with extra dimensions the truly fundamental constants are constants in the higher dimensional theory. The constants that appear in the dimensionally reduced theory are the result of integration over the extra dimensions. If the volume of the extra dimensions would change, the value of the constants we observe in the dimensionally-reduced theory would change. Exactly how they would change depends upon the theory. In Kaluza-Klein theories, gauge symmetries arise from continuous isometries in the internal manifold, while in superstring theories the gauge symmetries are part of the fundamental theory. In all theories the gravitational constant is inversely proportional to the volume of the internal manifold. In the most general case there is not a single radius in the internal manifold. However, for the sake of simplicity it will be assumed that there is a single radius, b, which characterizes the internal manifold. The b dependence of some fundamental constants are given in Table 4. In Table 4, α^0 is the present value of the fine structure constant, G^0 is the present value of the gravitational constant, G_F^0 is the present value of Fermi's constant, and b_0 is the present value of b.

• *The internal dimensions are static*: If the internal dimensions change, fundamental constants change. Limits on the time variability of the fundamental constants can be converted to limits on the time variability of the extra dimensions. Limits on time rate of change of the fine structure constant (assuming that the change is a power law in cosmological time) are given in Table 5. The look-back time, Δr, is the maximum time over which the limit may be applied. For the look-back time, an $\Omega = 1$ cosmology was assumed, i.e., a present age of $(2/3)H_0^{-1} = 6.6 \times 10^9 h^{-1}$yr. Long look-back times are relevant if the change is not a power law in cosmological time. It is interesting to know how soon after the bang the internal space had essentially the size it has today. The limit with the longest look-back time is the limit from primordial nucleosynthesis.

Primordial nucleosynthesis is a sensitive probe of changes in α, since the neutron-proton mass difference $Q = m_n - m_p = 1.293$ MeV has an electromagnetic component. Although the details of the neutron-proton mass difference are not known, it is reasonable to assume that the electromagnetic contribution is the same size (but the opposite sign) as the entire difference. With this assumption $\alpha/\alpha^0 = Q/Q^0$, where Q^0 is the value today.

$\lvert\dot{\alpha}/\alpha\rvert$	METHOD	Δr
$5 \times 10^{-15}\text{yr}^{-1}$	^{187}Re/^{187}Os	5×10^9yr
$1 \times 10^{-17}\text{yr}^{-1}$	Oklo reactor	1.8×10^9yr
$13 \times 10^{-13} h\ \text{yr}^{-1}$	Radio galaxies	$2 \times 10^9 h^{-1}$ yr
$2 \times 10^{-14} h\ \text{yr}^{-1}$	QSO	$5 \times 10^9 h^{-1}$ yr
$15 \times 10^{-15} h\ \text{yr}^{-1}$	Primordial nucleosynthesis	$6.6 \times 10^9 h^{-1}$ yr

Table 5: Constraints on the time variation of the fine structure constant

The neutron-proton ratio at freeze out given by Eq. 1.78 is $\exp(-Q/T_f)$, so n/p is very sensitive to small changes in Q. The primordial ^{4}He mass fraction as a function of b/b_0 is given in Fig. 25, assuming that α, G, and G_F depend on b/b_0 as in Table 4. The curve labeled "SS" is the superstring model ($D = 6$), and the curves marked "KK$_2$" and "KK$_7$" are Kaluza-Klein models with $D = 2$ and $D = 7$ internal dimensions. The allowed range of the primordial ^{4}He, $Y_P = X_4 = 0.24 \pm 0.01$. For the superstring model, the primordial helium is within acceptable limits only if at the time of primordial nucleosynthesis $1.005 \geq b/b_0 \geq 0.995$. The Kaluza-Klein models give the slightly less stringent result $1.01 \geq b/b_0 \geq 0.99$. In either case, by the time of primordial nucleosynthesis the internal dimensions had obtained a size very close to the size they have today.[3]

• *The ground state geometry does not have all the symmetries of the theory*: It is generally assumed that the ground state geometry is of the form $M^4 \times B^D$, where M^4 is four-dimensional Minkowski space, [19] and B^D is some compact D−dimensional space. The symmetries of the ground state are generally not as large as the symmetries of the theory, i.e., there is spontaneous symmetry breaking. One of the results of SSB is the existence of a massless (at least at the classical level) Nambu-Goldstone boson, which is sometimes called the dilaton.

• *The spectrum contains an infinite number of massive states*: If the radius of the internal space is b, then b^{-1} sets the scale for the massive states. The spectrum of the massive states depends upon the type of theory and the structure of the internal manifold. Since b is expected to be close to l_{Pl}, the massive states should have masses close to m_{Pl}.

[19] The assumption of M^4 is not quite correct in a cosmological context, and should be replaced by $R^1 \times S^3$ for the closed model, $R^1 \times Q^3$ for the open model.

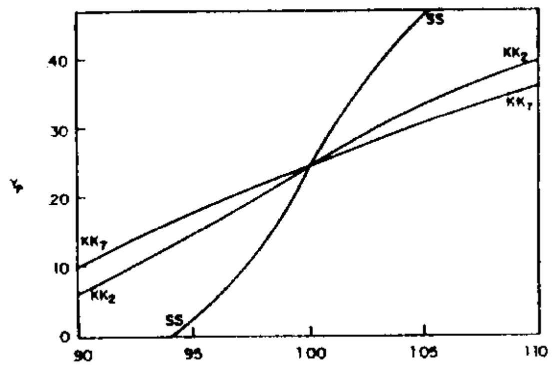

Figure 25: The primordial mass fraction as a function of b/b_0

6.2 Stability of the Internal Space

All theories formulated in extra dimensions must contain some mechanism to keep the internal dimensions static. In the absence of such a mechanism, the extra dimensions would either contract or expand. The origin of the vacuum stress responsible for this is unknown. Here, some toy models are given, along with some possible cosmological effects.

In theories with extra dimensions new types of interactions may arise. For a starting point, consider the Chapline-Manton action, [4] which is an $N = 1$ supergravity and an $N = 1$ super-Yang-Mills theory in 10 space-time dimensions. This theory is thought to be the field theory limit of a 10-dimensional superstring theory. It is not at all clear that the 10-dimensional field theory limit of the superstring ever makes sense. The 10-dimensional field theory description obtains only in the region between two similar energy scales. The first scale is determined by the string tension. It is the scale above which it is necessary to include the massive excitations of the string. Above this scale physics is "stringy" and any point-like field theory description is inadequate. The second scale is the compactification scale, which is determined by the radius of the internal space. At distances smaller than the compactification scale dimensional reduction no longer makes sense, the 3+1-dimensional description is inadequate, and the 10-dimensional theory must be used. The 10-dimensional field theory description makes sense at distance scales larger than the string tension scale, but smaller than the compactification scale. Since these two scales are expected to be the same order of magnitude, it is not clear if the 10-dimensional field theory description ever obtains. Nevertheless, it offers a convenient starting point for an exploration of cosmology in extra dimensions.

The Chapline-Manton Lagrangian contains the $N = 1$ supergravity multiplet $\{e^A_M;\ \psi_M;\ B_{MN};\ \lambda;\ \sigma\}$, where e^A_M is the vielbein, ψ_M is the Rarita-Schwinger field, B_{MN} is the Kalb-Ramond field, λ is the sub-gravitino, and σ is the dilaton, and the super Yang-Mills multiplet $\{G_{MN};\ \chi\}$, where G_{MN} is the Yang-Mills field strength and χ is the gluino field. The Lagrangian is [20]

$$
\begin{aligned}
e^{-1}\ \mathcal{L} \ =\ & -\frac{1}{2}R - \frac{1}{2}\bar{\psi}_M\Gamma^{MPS}D_P\psi_S - \frac{3}{4}\exp(-\sigma)H_{MNP}H^{MNP} \\
& -\frac{1}{4}\partial_M\sigma\partial^M\sigma - \frac{3\sqrt{2}}{8}\bar{\psi}_M\ \partial\!\!\!/\sigma\Gamma^M\lambda - \frac{1}{2}\bar{\lambda}\ D\!\!\!\!/\lambda \\
& +\frac{\sqrt{2}}{16}\exp(-\sigma/2)H_{MNP}\left(\bar{\psi}_Q\Gamma^{QMNPR}\psi_R + 6\bar{\psi}^M\Gamma^N\psi^P\right. \\
& \left.-\sqrt{2}\bar{\psi}_R\Gamma^{MNP}\Gamma^R\lambda\right) - \frac{1}{2}\mathrm{Tr}\bar{\chi}\ D\!\!\!\!/\chi - \frac{1}{4}\exp(-\sigma/2)\mathrm{Tr}G_{MN}G^{MN} \\
& -\frac{3}{4}\left(\mathrm{Tr}\bar{\chi}\Gamma_{MNP}\chi\right)^2 + \exp(-\sigma/2)H_{MNP}\mathrm{Tr}\bar{\chi}\Gamma^{MNP}\chi + \cdots
\end{aligned}
\tag{6.1}
$$

where $\Gamma^{MNP} = \Gamma^{[M}\Gamma^N\Gamma^{P]}$, and $H_{MNP} = \partial_{[M}B_{NP]}$. Four fermion couplings and other terms have been omitted.

The "Einstein equations" are straightforward to obtain:

$$
\begin{aligned}
R_{MN} \ =\ & \frac{9}{2}\exp(-\sigma)\left(H_{MPQ}H_N^{\ PQ} - \frac{1}{12}g_{MN}H_{PQR}H^{PQR}\right) \\
& -\exp(-\sigma/2)\left(\mathrm{Tr}G_{MP}G_N^{\ P} - \frac{1}{16}g_{MN}\mathrm{Tr}G_{PQ}G^{PQ}\right) \\
& -\frac{1}{2}\partial_M\sigma\partial^M\sigma - \frac{1}{8}\left(\mathrm{Tr}\bar{\chi}\Gamma_{PQR}\chi\right)\left(\bar{\lambda}\Gamma^{PQR}\lambda\right)g_{MN} \\
& -\frac{3}{16}\left(\mathrm{Tr}\bar{\chi}\Gamma_{PQR}\chi\right)^2 g_{MN} + \frac{9}{2}\exp(-\sigma/2)H_M^{\ PQ}\mathrm{Tr}\bar{\chi}\Gamma_{NPQ}\chi \\
& -\frac{3}{16}\exp(-\sigma/2)g_{MN}H_{PQR}\mathrm{Tr}\bar{\chi}\Gamma^{PQR}\chi + \cdots
\end{aligned}
\tag{6.2}
$$

The task at hand is to solve Eq. 6.2 to find the equations of evolution of the scale factor(s) in the expansion of the Universe toward the quasi-static ground state of the system where there are D static dimensions and 3 dynamic dimensions expanding as in a standard FRW cosmology.

[20]The following notation will be used: D =number of extra dimensions; M, N, P, Q, ... run from 0 to $D+3$; μ, ν, ρ, ... are indices in the extra dimensions; and m, n, p, q, ... are indices in the large spatial dimensions.

In general it is necessary to choose background field configurations. For example consider the "bosonic" parts of the equations. What are the symmetries of the metric? What are the vacuum (background) values of H_{MNP}, of G_{MN}, of $\bar{\chi}\Gamma\chi$, of $\bar{\lambda}\Gamma\lambda$, of σ? In general, many (possibly infinitely many) solutions of the field equations are expected, even if there is but one ground state that describes the microphysics of our Universe. The immediate question to ask is what picks out the ground state and what is the evolution of the Universe to this ground state? Perhaps when the true string nature of the equations are taken into consideration there will be but one possible solution to the string equations even if there are many solutions to the field theory. Perhaps something in the evolution of the Universe prefers a unique or small number of possibilities. Such questions are reminiscent of the questions considered in inflation. If the conditions in some region of the Universe are such as to enter an inflationary phase, that region of the Universe will grow relative to a region that does not undergo inflation. It is possible to imagine that the Universe starts in a state with no particular background field configuration, but in a quantum state described by a wave function Ψ that describes the probability of a given configuration, Ψ(field configurations). If in some region of the Universe the wave function is peaked about a particular configuration that will inflate some spatial dimensions, that region will grow. All that is required to produce the Universe we observe is that there is some region that will lead to three spatial dimensions inflating (and some mechanism to keep D dimensions static). It may be that the theory is unique, but the ground state is not. It may be that somewhere outside of our horizon the Universe is quite different. There may be a different number of small versus large dimensions, or the internal space may have different topological properties leading to drastically different microphysics. Before this speculation is considered, it is necessary to understand the mechanism that leads to the stabilization of the internal space. This problem will be studied by considering individual contributions to the right-hand side of Eq. 6.2.

For simplicity, the metric will be taken to have the symmetry $R^1 \times S^3 \times S^D$

$$g_{MN} = \begin{pmatrix} 1 & & \\ & -a^2(t)\tilde{g}_{mn} & \\ & & -b^2(t)\tilde{g}_{\mu\nu} \end{pmatrix} \tag{6.3}$$

where $\tilde{g}_{mn}$ is the metric for S^3 of unit radius and $a(t)$ is the actual radius, and $\tilde{g}_{\mu\nu}$ is the metric for S^D of unit radius and $b(t)$ is the actual radius. The components of the Ricci tensor are

$$\begin{aligned} -R_{00} &= 3\frac{\ddot{a}}{a} + D\frac{\ddot{b}}{b} \\ -R_{mn} &= \left[\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + D\frac{\dot{a}}{a}\frac{\dot{b}}{b} + \frac{2}{a^2}\right] g_{mn} \\ -R_{\mu\nu} &= \left[\frac{\ddot{b}}{b} + (D-1)\frac{\dot{b}^2}{b^2} + 3\frac{\dot{a}}{a}\frac{\dot{b}}{b} + \frac{D-1}{b^2}\right] g_{\mu\nu}. \end{aligned} \tag{6.4}$$

With the Einstein equations in the form

$$R_{MN} = 8\pi\bar{G}\left[T_{MN} - \frac{1}{D+2}g_{MN}T^P_P - \frac{1}{D+2}\frac{\Lambda}{8\pi\bar{G}}g_{MN}\right] \tag{6.5}$$

where $\bar{G}$ is the gravitational constant in $D+4$ dimensions,[21] and Λ is a possible cosmological constant in $D+4$ dimensions. All the terms on the right-hand side of Eq. 6.2 contribute to T_{MN} and Λ.

Symmetries of the stress tensor are usually chosen such that the only non-vanishing components of the stress tensor are

$$\begin{aligned} T_{00} &\equiv \rho \\ T_{mn} &\equiv -p_3 g_{mn} \\ T_{\mu\nu} &\equiv -p_D g_{\mu\nu} \end{aligned} \tag{6.6}$$

with $T^M_M = \rho - 3p_3 - Dp_D$. In terms of ρ, p_3, p_D, and $\rho_\Lambda = \Lambda/8\pi\bar{G}$ the Einstein equations are

$$\begin{aligned} 3\frac{\ddot{a}}{a} + D\frac{\ddot{b}}{b} &= -\frac{8\pi\bar{G}}{D+2}\left[(D+1)\rho + 3p_3 + Dp_D - \rho_\Lambda\right] \\ \frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + D\frac{\dot{a}}{a}\frac{\dot{b}}{b} + \frac{2}{a^2} &= \frac{8\pi\bar{G}}{D+2}\left[\rho + (D-1)p_3 - Dp_D + \rho_\Lambda\right] \\ \frac{\ddot{b}}{b} + (D-1)\frac{\dot{b}^2}{b^2} + 3\frac{\dot{a}}{a}\frac{\dot{b}}{b} + \frac{D-1}{b^2} &= \frac{8\pi\bar{G}}{D+2}\left[\rho - 3p_3 + 2p_D + \rho_\Lambda\right]. \end{aligned} \tag{6.7}$$

Some possible contributions to the right hand side will be considered in turn.

• R_{MN} =NOTHING: The simplest possible form for the right hand side is zero. For the moment abandon the choice of $R^1 \times S^3 \times S^D$, and consider a $D+3$ torus for the ground state geometry. The spatial coordinates can be chosen to take the values $0 \le x^i \le L$, where L is a parameter with dimension of length. The general cosmological solutions of the vacuum Einstein equations are the Kasner solutions. The Kasner metric is

$$ds^2 = dt^2 - \sum_{i=1}^{D+3}\left(\frac{t}{t_0}\right)^{2p_i}(dx^i)^2. \tag{6.8}$$

The Kasner metric is a solution to the vacuum Einstein equations provided the Kasner conditions are satisfied

[21] $\bar{G}$ is related to Newton's constant G by $\bar{G} = GV_D^0$, where V_D^0 is the volume of the internal space today.

$$\sum_{i=1}^{D+3} p_i = \sum_{i=1}^{D+3} p_i^2 = 1. \tag{6.9}$$

In order to satisfy the Kasner conditions at least one of the p_i must be negative. It is possible to have 3 spatial dimensions expanding in an isotropic manner and D dimensions contracting in an isotropic manner by the choice[5]

$$\begin{aligned} p_1 = p_2 = p_3 \quad &\equiv \quad p = \frac{3 + (3D^2 + 6D)^{1/2}}{3(D+3)} \\ p_4 = \ldots = p_{3+D} \quad &\equiv \quad q = \frac{D - (3D^2 + 6D)^{1/2}}{D(D+3)}. \end{aligned} \tag{6.10}$$

Note that $p > 0$ and $q < 0$. With this choice the metric may be written

$$ds^2 = dt^2 - a^2(t)d\vec{x}^2 - b^2(t)d\vec{y}^2, \tag{6.11}$$

where x^i are coordinates of the 3 expanding dimensions, and y^i are coordinates of the D contracting dimensions. The two scale factors are given by $a(t) = (t/t_0)^p$, $b(t) = (t/t_0)^q$.

Somewhat more complicated classical cosmologies have been considered. The Kasner model can be regarded as an anisotropic generalization of the flat FRW cosmology, i.e., a Bianchi I cosmology. A generalization of the closed FRW model is the Bianchi IX model. The Bianchi IX vacuum solutions have the feature that the general approach to the singularity is "chaotic."[6] On approach to the initial singularity the scale factors in different spatial directions undergo a series of oscillations, contractions, and expansions. This feature is quite general, and independent of the state of the Universe after the singularity. The oscillation of the scale factors is well described by a sequence of Kasner models in which expanding and contracting dimensions are interchanged in "bounces." Such anisotropic behavior is predicted to be the general approach to the initial singularity. The question of whether such a chaotic approach to the initial singularity is present in more than three spatial dimensions has been considered. It has been shown that chaotic behavior obtains only for models with between 3 and 9 spatial dimensions.[7] The importance of this observation is clouded by the fact that at the approach to the singularity curvature may not dominate the right hand side of the Einstein equations, and near the singularity classical gravity may be a poor description.

The solutions above do not have solutions with a static internal space and if they are ever relevant, it is only for a limited time. The right-hand side must be more complicated than nothing. The next simplest thing to consider on the right-hand side is free scalar fields. Before discussing their effect on the evolution of the Universe it is necessary to discuss regularization in the background geometry.

The free energy of a non-interacting spinless boson of mass μ is given by[8]

$$F = T\frac{1}{2}\ln \mathrm{Det}\left(-\Box_{4+D} + \mu^2\right). \tag{6.12}$$

since finite temperature effects are of interest, the time is periodic with period of $1/2\pi T$, the relevant geometry is $S^1 \times S^3 \times S^D$, and the radii of the spheres are $1/2\pi T$, a, and b. The eigenvalues of $\Box$ on the compact space are discrete, and are given by the triple sum (hereafter μ will be set to zero)

$$\begin{aligned} 2T^{-1}F = \sum_{r=-\infty}^{\infty} \sum_{m,n=0}^{\infty} D_{mn} \ln \Big[r^2(2\pi T)^2 \quad &+ \quad m(m+2)a^{-2} \\ &+ \quad n(n+D-1)b^{-2}\Big], \end{aligned} \tag{6.13}$$

where D_{mn} is a factor that counts the degeneracy

$$D_{mn} = (m+1)^2(2n+D-1)(n+D-2)!/(D-1)!n!. \tag{6.14}$$

The free energy given by Eq. 6.13 is, of course, infinite. To deal with the infinities, a regularization scheme will be found to extract the relevant finite part. For the purpose of regularization, each term in the sum can be expressed as an integral using the formula[8] 22

$$\ln X = \frac{d}{ds}X^s \,|_{s=0} = \frac{d}{ds}\left(\frac{1}{\Gamma(-s)}\int_0^\infty dt\, t^{s-1}\exp(-tX)\right)_{s=0}. \tag{6.15}$$

The finite part of the free energy is given by

$$2T^{-1}F = \frac{d}{ds}\left[\frac{1}{\Gamma(-s)}\int_0^\infty dt\, t^{-s-1}\sigma_1(4\pi^2T^2t)\sigma_3(a^{-2}t)\sigma_D(b^{-2}t)\right]_{s=0}, \tag{6.16}$$

where the functions σ_i are given by

$$\sigma_i(x) = \sum_{n=0}^{\infty} \frac{(2n+i-1)(n+i-1)!}{(i-1)!n!} \exp[-n(n+i-1)x]. \tag{6.17}$$

The full expression for the free energy is quite difficult to evaluate, but the free energy is simple in several limits. In the "flat-space" limit the radius of S^3 is much larger than the radius of S^D ($a \gg b$) and $\sigma_3 \to (\sqrt{\pi}/4)a^3t^{-3/2}$. In the limit $a \gg b$ the free energy can be approximated by

$$F = \frac{\Omega_3 a^3}{b^4}\left[c_1 - c_2(bT)^4 - c_3(bT)^{D+4}\right], \tag{6.18}$$

where Ω_i is found from the volume of the i-sphere, $V_i = R^i\Omega_i$ with R the radius and $\Omega_i = (2\pi)^{(i+1)/2}/\Gamma[(i+1)/2]$. For S^3, the volume is $V_3 = R^3 2\pi^2$, and Ω_3 has the familiar form $\Omega_3 = 2\pi^2$. The term proportional to c_1 is the Casimir term (c_1

---

22 This regularization is only valid for D =odd. The D =even case will be discussed below.

	Casimir $T=0$	Low Temperature $0 \leq T \leq b^{-1}$	High Temperature $T \gg b^{-1}$	Monopole T=0
ρ	$c_1/\Omega_D b^{4+D}$	$(\pi^2/30)T^4/\Omega_D b^{4+D}$	$(D+3)c_3 T^{D+4}/\Omega_D$	$f_0^2/2b^{2D}$
p_3	$-c_1/\Omega_D b^{4+D}$	$(\pi^2/90)T^4/\Omega_D b^{4+D}$	$c_3 T^{D+4}/\Omega_D$	$-f_0^2/2b^{2D}$
p_D	$4c_1/D\Omega_D b^{4+D}$	0	$c_3 T^{D+4}/\Omega_D$	$f_0^2/2b^{2D}$
T^M_M	0	0	0	$(4-D)f_0^2/2b^{2D}$

Table 6: Contributions to thermodynamic quantities

is c_N of Candelas and Weinberg[9]). The term proportional to $c_2 = \pi^2/90$ is the leading temperature-dependent term when $T \ll b^{-1}$. When $T \gg b^{-1}$, the term proportional to $c_3 = (2\zeta(D+4)/\pi^{3/2})\Gamma[(D+4)/2]/\Gamma[(D+1)/2]$ dominates. In the "low-temperature" limit the radius of the S^1 becomes large and $\sigma_1 \rightarrow (4\pi t T^2)^{-1/2}$. In the flat-space, zero-temperature limit only the term proportional to c_1 survives.

The internal energy is given in terms of the free energy, the temperature, and the entropy

$$S = -\left[\frac{\partial F}{\partial T}\right]_{a,b}, \tag{6.19}$$

by $U = F + TS$. The thermodynamic quantities ρ, p_3, and p_D are defined in terms of the internal energy:

$$\begin{aligned} \rho &= \frac{U}{\Omega_3 \Omega_D a^3 b^D} \\ p_3 &= -\frac{a}{3\Omega_3\Omega_D a^3 b^D}\left[\frac{\partial U}{\partial a}\right]_{b,S} \\ p_D &= -\frac{b}{D\Omega_3\Omega_D a^3 b^D}\left[\frac{\partial U}{\partial b}\right]_{a,S}. \end{aligned} \tag{6.20}$$

The thermodynamic quantities in zero temperature, low temperature, and high temperature limits are given in Table 6. There are several obvious limits of Table 6. In the zero-temperature or in the low-temperature limits, dimensional reduction is possible. Upon integration over the internal dimensions the effective three-dimensional energy density and pressure is obtained by multiplication by $V_D = \Omega_D b^D$. After dimensional reduction the Casimir terms are proportional to $c_1 b^{-4}$. The low-temperature limit after dimensional reduction is $\rho = 3p_3 \rightarrow (\pi^2/30)T^4$ and $p_D = 0$, which is the expected contribution for a spinless boson in 3+1 dimensions. In the high-temperature limit dimensional reduction does not make sense.

It is possible to perform a similar analysis for particles of higher spin. The technical details are more difficult, but the physics is quite similar.

•R_{MN} =RADIATION:[10] Consider the "high-temperature" ($T \geq b^{-1}$) "flat-space" ($a \gg b$) limit with $\Lambda = 0$. In this limit T_{MN} is isotropic in the sense that $p_3 = p_D \equiv p$ (see Table 6). The Einstein equations are

$$\begin{aligned} 3\frac{\ddot{a}}{a} + D\frac{\ddot{b}}{b} &= -8\pi\bar{G}\rho \\ \frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + D\frac{\dot{a}}{a}\frac{\dot{b}}{b} &= 8\pi\bar{G}p \\ \frac{\ddot{b}}{b} + (D-1)\frac{\dot{b}^2}{b^2} + 3\frac{\dot{a}}{a}\frac{\dot{b}}{b} + \frac{D-1}{b^2} &= 8\pi\bar{G}p. \end{aligned} \tag{6.21}$$

In keeping with the flat space assumption the $2/a^2$ term has been dropped in R_{mn}. The equation of state is $\rho = Np$, where $N \equiv D + 3$. The conservation law $T^{MP}{}_{;P} = 0$ implies

$$\rho\bar{a}^{N+1} = \text{constant}, \tag{6.22}$$

where $\bar{a} \propto (a^3 b^D)^{1/N}$ is the mean scale factor. Since $\rho \propto T^{N+1}$, there is a conserved quantity $S_N = (\bar{a}T)^N$ that is constant. This is simply the total N-dimensional entropy.

The Einstein equations (or a subset of the Einstein equations and the $T^{MP}{}_{;P} = 0$ equation) can be integrated to give $a(t)$ and $b(t)$. A typical solution is shown in Fig. 26. Both scale factors emerge from a initial singularity. The scale factor for the internal space reaches a maximum and recollapses to a second singularity. As b approaches the second singularity a is driven to infinity. The parameter x/x_* in Fig. 26 is a measure of the time in units of the time necessary to reach the second singularity.

The evolution of the temperature is shown in Fig. 27. The figure demonstrates the rather striking feature that as the second singularity is approached, the temperature increases. The expansion of a together with an increase of T seems unusual. However it is simply due to the conservation of entropy. In the region of growing T the mean volume of the Universe is actually *decreasing*, and the temperature must increase to keep S_N constant.

The assumption of the flat-space limit for S^3 can be easily justified. Imagine that the spatial geometry is $S^3 \times S^D$. If $a \geq b$ in the high-temperature region, once the maximum of b is reached, the S^3 will be inflated. The only requirement is that the curvature term, $1/a^2$, is small compared to the thermal term, $8\pi\bar{G}\rho$, at $b = b_{MAX}$.

In the approach to the second singularity the combination of expanding and contracting dimensions behaves like a Kasner model. A recurring feature in the