FERMILAB Conf-84/59-A
June 1984

COSMOLOGY AND GUTS*

Keith A. Olive

Astrophysics Theory Group
Fermi National Accelerator Laboratory
Batavia, Illinois 60510 U.S.A.

* To be published in Grand Unification With and Without Supersymmetry and Cosmological Implications, by C. Kounnas, A. Masiero, D. V. Nanopoulos and K. A. Olive; International School for Advanced Studies Lecture Series No. 2. Trieste, Italy, Summer 1983. World Scientific Publishing Co., Singapore.

COSMOLOGY AND GUTS

Keith A. Olive

Astrophysics Theory Group
Fermi National Accelerator Laboratory
Batavia, Illinois 60510 U.S.A.

Introduction

Up until the last fifteen to twenty years, cosmology had been considered as a link between astrophysics and general relativity. Although nuclear physics began to play a role as early as the 1940s when big bang nucleosynthesis[1] was first discussed as a possible cosmological origin of the elements, it was not until the discovery of the microwave background radiation[2] that the big bang model was placed as the front-running cosmological model. More recently, with the advent of grand unified theories (GUTs) cosmology has played an integral role in particle physics. Among the most important results due to the interplay between cosmology and particle physics are, for example, the cosmological limits that one sets on particle abundances, masses, and lifetimes which set guides for building unified models. On the other hand, the incorporation of GUTs into the big bang model led to a theory of big bang baryosynthesis[3] to explain the slight excess of matter over antimatter.

In the first three parts of these lectures, we have presented the current status GUTs and supersymmetric GUTs (global and local). In this final part we will try to cover the status of the big bang model and in particular its role in unification models. To do this, we will begin by reviewing the essential ingredients from general relativity needed to describe fully the Friedmann-Robertson-Walker and De Sitter models. These models which include the standard big bang model will be discussed in some detail. In section 3 we will review the current status of big bang nucleosynthesis and the origin of the light elements. This discussion will include the cosmological bounds on the baryon to photon ratio and limits on the number of light neutrino flavors. In section 4, we will derive the limits on neutrino masses and lifetimes. In section 5, we arrive at the junction between GUTs and cosmology and will review the present status of baryon generation in the very early Universe. We will examine both standard and supersymmetric GUTs. The remaining two sections are those which link cosmology and particle physics most closely. In those sections, we will discuss the major problems with the standard big bang model and describe in detail their solution in the inflationary Universe scenario.[4] Once again, our discussion will include both supersymmetric and non-supersymmetric GUTs.

Section 1. Essentials From General Relativity

Let us begin these lectures on cosmology by stating our main guiding assumptions. These assumptions are in fact so basic that they are really at the foundations of all modern physics. They are:

1) The Copernican Principle: we are not privileged observers. On the average we do not expect the Universe to look any different from any other spatial position.

2) The Relativity Principle: physical laws do not depend on space-time. Without such an assumption, it would, of course, be impossible to consider any cosmological model or any description of the Universe as a whole.

These two principles taken together are commonly referred to as:

The Cosmological Principle: the Universe is isotropic in all measurable properties at all times over all space. That is, the Universe is spatially homogeneous and isotropic. This is of course an extremely strong assumption which may or may not have been true throughout the history of the Universe. In section 6, we will look more closely as to how good an assumption this really is.

There are two immediate consequences of the Cosmological Principle. The first is that the only true velocity fields allowed are either overall expansion or contraction. Other possibilities, such as, rotation, shear, combined expansion and contraction are all contained in the anisotropic Bianchi models.[5)] Though these may have been important initially, we will not consider them here. Furthermore, any expansion or contraction present must have no apparent center. That is to say that the relative velocity between **any** two observers most depend only on their separation

$$v_{12} = Hr_{12}, \tag{1.1}$$

where H is a Universal spatial constant. This (1.1) is also known as Hubble's Law. The second consequence of the Cosmological Principle is that there must exist a measure of distance which is independent of direction. Such a measure might be, for example

$$d = zc/H, \tag{1.2}$$

where z is the redshift (blueshift) due to the expansion (contraction) of an emitted signal.

More generally, the second consequence implies that there exists a metric which does not depend on direction. Formally, a metric g is a symmetric tensor of the form

$$g = g_{\mu\nu}\, dx^{\mu}\, dx^{\nu} \tag{1.3}$$

and defines the line element ds^2. To each vector X, the metric assigns a magnitude $(|g(X,X)|)^{1/2}$. The vector X will be time-like, null, or space-like depending on whether

$$\begin{array}{ll} g(X,X) < 0 & \text{time-like} \\ g(X,X) = 0 & \text{null} \\ g(X,X) > 0 & \text{space-like.} \end{array} \tag{1.4}$$

Furthermore, the metric must be non-singular so that it has an inverse defined by

$$g^{-1} = g^{\mu\nu} \frac{\partial}{\partial x^{\mu}} \frac{\partial}{\partial x^{\nu}} \tag{1.5}$$

and

$$g^{\mu\nu} g_{\nu\lambda} = \delta^{\mu}{}_{\lambda}. \tag{1.6}$$

If we now apply the Cosmological Principle to the metric g we see that we must have

$$\begin{array}{ll} g_{oi} = 0 & \\ g_{ij} = 0 & i \neq j \end{array} \tag{1.7}$$

or

$$ds^2 = g_{oo}\, dt^2 + g_{ii} dx^i dx^i. \tag{1.8}$$

One can further define a set of coordinates so that without loss of generality the homogeneous and isotropic metric will take the form

$$ds^2 = -dt^2 + R^2(t)\, d\sigma^2, \tag{1.9}$$

where $d\sigma^2$ is the three-space metric of constant curvature and is time-independent. The different three-space geometries will then be those of positive, negative, and zero curvature. In general, we can write

$$d\sigma^2 = dr^2 + f^2(r)\, (d\theta^2 + \sin^2\theta\, d\phi^2) \tag{1.10a}$$

and

$$f(r) = \begin{array}{lll} \sin r & & k = +1 \\ \sinh r & \text{for} & k = -1, \\ r & & k = 0 \end{array} \tag{1 .10b}$$

where k is the curvature constant representing the sign of the intrinsic curvature of the space time (see below for a more formal definition of k). Homogeneity and isotropy guarantee that the form of f will be independent of θ and ϕ. This metric known as the Freidmann-Robertson-Walker metric can be written in a more compact form,

$$ds^2 = -dt^2 + R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\,(d\theta^2 + \sin^2\theta d\phi^2)\right]. \tag{1.11}$$

In order to derive the equations describing the dynamical evolution of a cosmological model with metric (1.11), we will have to briefly review the necessary ingredients from general relativity. To begin with, the covariant derivative of a vector X^ν is defined by

$$\begin{aligned} D_\mu X^\nu &= X^\nu_{,\mu} + \Gamma^\nu_{\mu\sigma} X^\sigma = X^\nu_{;\mu} \\ D_\mu X_\nu &= X_{\nu,\mu} - \Gamma^\sigma_{\mu\nu} X_\sigma = X_{\nu;\mu}\,, \end{aligned} \tag{1.12}$$

where

$$X^\nu_{,\mu} = \partial_\mu X^\nu \text{ and } X_{\nu,\mu} = \partial_\mu X_\nu \tag{1.13}$$

is the ordinary derivative and the connection (Christoffel symbol) is given in terms of the metric by

$$\Gamma^\mu_{\nu\sigma} = \frac{1}{2}\left\{g_{\rho\sigma,\nu} + g_{\nu\rho,\sigma} - g_{\nu\sigma,\rho}\right\} g^{\rho\mu}. \tag{1.14}$$

(Note that $\Gamma^\mu_{\nu\sigma} = \Gamma^\mu_{\sigma\nu}$ is symmetric in its lower two indices.) Using the connection (1.14), the Riemann curvature tensor is defined by

$$R^\sigma_{\mu\rho\nu} = \Gamma^\sigma_{\nu\mu,\rho} - \Gamma^\sigma_{\rho\mu,\nu} + \Gamma^\lambda_{\nu\mu}\,\Gamma^\sigma_{\rho\lambda} - \Gamma^\lambda_{\rho\mu}\,\Gamma^\sigma_{\lambda\nu}. \tag{1.15}$$

A space-time will be defined to be flat if $R^\sigma_{\mu\rho\nu} = 0$. Contracting on σ and ρ we have the Ricci tensor $R_{\mu\nu}$ and on further contraction we have the curvature scalar R_c which are defined by

$$R_{\mu\nu} = R^\sigma_{\mu\sigma\nu} = \Gamma^\sigma_{\mu\nu,\sigma} - \Gamma^\sigma_{\sigma\mu,\nu} + \Gamma^\sigma_{\mu\nu}\,\Gamma^\rho_{\sigma\rho} - \Gamma^\sigma_{\rho\mu}\,\Gamma^\rho_{\sigma\nu} \tag{1.16}$$

$$R_c = R^\mu_\mu = R_{\mu\nu} g^{\mu\nu}. \tag{1.17}$$

The curvature constant used in (1.10b) is related to the three-space Ricci tensor (i,j running only from 1 to 3) which is given by (for maximally symmetric three spaces)

$${}^3R_{ij} = 2kg_{ij} \tag{1.18}$$

with a three-space curvature scalar

$${}^3R_c = 6k. \tag{1.19}$$

k is then normalized to ±1,0 by adjusting the scale factor R(t) in (1.9) and (1.11). Finally, we must define our energy-momentum tensor

$$T_{\mu\nu} = pg_{\mu\nu} + (p + \rho)\,u_\mu\,u_\nu, \tag{1.20}$$

where p is the isotropic pressure, ρ is the total mass-energy density and $u_\mu = (1,0,0,0)$ is the velocity vector for an isotropic fluid.

We are now able to begin to derive the equations describing the evolution of the Friedmann-Robertson-Walker models. We begin with Einstein's equation

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R_c = 8\pi G_N T_{\mu\nu} - \Lambda g_{\mu\nu}, \tag{1.21}$$

where G_N is Newton's gravitational constant and Λ is what is known as the cosmological constant. To derive the field equations we must first work out the set of Christoffel symbols (i,j run from 1 to 3 only)

$$\begin{aligned}
\Gamma^o_{ij} &= \frac{\dot{R}}{R} g_{ij} \\
\Gamma^i_{oj} &= \Gamma^i_{jo} = (\dot{R}/R)\delta^i_j \\
\Gamma^1_{11} &= kr\ (1-kr^2)^{-1} \\
\Gamma^1_{22} &= -(1-kr^2)r\ ;\ \Gamma^1_{33} = -(1-kr^2)\ r\sin^2\theta \\
\Gamma^2_{12} &= \Gamma^2_{21} = \Gamma^3_{13} = \Gamma^3_{31} = 1/r \\
\Gamma^2_{33} &= -\sin\theta\cos\theta\ ;\ \Gamma^3_{23} = \Gamma^3_{32} = \cot\theta.
\end{aligned} \tag{1.22}$$

Using this, we find that the only non-vanishing Ricci coefficients are

$$\begin{aligned}
R^0_0 &= 3\ddot{R}/R \\
R^1_1 = R^2_2 = R^3_3 &= \lfloor 2\ (\dot{R}/R)^2 + (\ddot{R}/R) + (2k/R^2)\rfloor.
\end{aligned} \tag{1.23}$$

Hence the curvature scalar is

$$R_c = R^\mu_\mu = 6\lfloor(\dot{R}/R)^2 + (\ddot{R}/R) + (k/R^2)\rfloor. \tag{1.24}$$

Let us now examine Einstein's equation (1.21) using (1.23) and (1.24). If we concentrate first on the 0-0 term in the field equations we find the standard Friedmann equation

$$H^2 \equiv (\dot{R}/R)^2 = \frac{8\pi G_N \rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3} \tag{1.25}$$

and defines the Hubble parameter H. This equation may be thought of as describing the total energy content of the Universe. If we just rewrite (1.25) as

$$-k = \dot{R}^2 - \frac{8\pi G_N \rho}{3} R^2 - \frac{\Lambda R^2}{3}, \tag{1.25'}$$

we can interpret $-k$ as the total energy of the system, the kinetic term represented by $\dot{R}^2$, the gravitational potential energy by the term containing ρ, with an additional energy source available in Λ. A helpful analogy is the total energy of a rocket at lift-off. If the total energy is positive (in this case $k < 0$) then the initial kinetic energy is great enough (the initial velocity is greater than escape velocity) and the rocket will escape the gravitational pull of the earth, or, in our case, the Universe will continue to expand forever, i.e., the Universe is open. If on the other hand, the total energy is negative ($k > 0$), the rocket will fall back to earth, and the Universe will recollapse, i.e., the Universe is closed. In the third possibility ($k = 0$), the Universe corresponds to the rocket just at escape velocity and the Universe will expand indefinitely. This is known as the critical or spatially flat Universe.

There is one additional equation which comes from the spatial components in Einstein's equation (1.21).

$$2(\ddot{R}/R) + (\dot{R}/R)^2 + (k/R^2) = \Lambda - 8\pi G_N p, \tag{1.26}$$

or substituting for $\dot{R}/R$ we have an equation for the acceleration

$$(\ddot{R}/R) = \frac{\Lambda}{3} - \frac{4\pi G_N}{3} (\rho + 3p). \tag{1.27}$$

The final equation that we need in order to set up the class of homogeneous and isotropic cosmological models comes from energy conservation

$$T^{\mu\nu}{}_{;\nu} = T^{\mu\nu}{}_{,\nu} + \Gamma^{\mu}_{\nu\rho} T^{\nu\rho} + \Gamma^{\nu}_{\nu\rho} T^{\mu\rho} = 0 \tag{1.28}$$

or

$$\dot{\rho} = -3 (\dot{R}/R) (\rho + p). \tag{1.29}$$

In the remaining six sections of these lectures, we will concentrate on interpreting the consequences of equations (1.25), (1.27), and (1.29). In particular, we will search through the class of solutions for the set which can most closely resemble our observed Universe. Given general relativity the rest should be easy. All we need to know is the full equation of state ($p[\rho]$) at all temperatures and whether or not $k = +1$, -1, or 0.

Section 2. Standard Cosmological Models

The Friedmann-Robertson-Walker metric (1.11) covers the full range of isotropic and homogeneous cosmological models. In these models, there are basically four independent quantities which need to be specified. They are 1) the sign of the curvature constant k; 2) the value of the cosmological constant Λ; 3) the equation of state $p(\rho)$. The fourth quantity, as we will see below, essentially corresponds to a measure of total entropy. This makes, however, no qualitative difference between the models.

The simplest type of space-time described by (1.11) is that of empty space, i.e., $p = \rho = \Lambda = 0$. As can be seen from Eq. (1.25), such a space time can be either critical or open. The critical case ($k = 0$) is just that of a non-expanding Minkowski space which is used in special relativity,

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2. \tag{2.1}$$

It is also possible to give the space-time some intrinsic curvature with $k = -1$. In this case we find [from Eq. (1.25)] that the scale factor grows linearly with time

$$R \propto t. \tag{2.2}$$

Although $R_c = R_{\mu\nu} = 0$, there are components of $R^{\sigma}{}_{\mu\rho\nu} \neq 0$. This space is open (by definition $k = -1$) and continues to expand indefinitely.

The more interesting class of models have either or both $\rho \neq 0$, $\Lambda \neq 0$. Before attempting to classify the full range of these models, let us look at two more simple examples. In each case we will take $\Lambda = k = 0$. In the first case we specify the equation of state as

$$p = \rho/3, \tag{2.3}$$

which is the equation of state of a free gas of radiation. If we use this equation of state in the equation for energy conservation (1.29), we have

$$\dot{\rho} = -4(\dot{R}/R)\rho \tag{2.4}$$

or $\rho \propto R^{-4}$. From the Friedmann equation (1.25) we then find that

$$R \propto t^{1/2}. \tag{2.5}$$

Finally using both (1.25) and (1.29) we can solve for the "age" of the Universe by

$$(\dot{\rho}/\rho) = -4(8\pi G_N\rho/3)^{1/2} \tag{2.6}$$

leading to

$$t = (3/32\pi G_N\rho)^{1/2} + \text{constant}. \qquad (2.7)$$

This is typically referred to as a radiation-dominated Universe.

The second useful example is described by choosing the following equation of state

$$p = 0 \quad \rho \neq 0, \qquad (2.8)$$

i.e., we have a dust-filled or matter-dominated Universe. Once again, energy conservation (1.29) tells us that

$$\dot{\rho} = -3(\dot{R}/R)\,\rho \qquad (2.9)$$

or $\rho \propto R^{-3}$. From (1.25), we find the time dependence of the scale factor

$$R \propto t^{2/3}. \qquad (2.10)$$

In the standard big bang model, the Universe has spent nearly all of its lifetime in one of these two cases.

Let us now examine more completely the full class of the Friedmann-Robertson-Walker models. First let us define a quantity Q

$$Q = \frac{3k}{R^2} - 8\pi G_N\rho \qquad (2.11)$$

we can then rewrite (1.25) as

$$(\dot{R}/R) = \pm\,[(\Lambda - Q)/3]^{1/2} \qquad (2.12)$$

which immediately tells us that $Q < \Lambda$ (see below for explanation of $\Lambda > Q$). Furthermore, we will specify the equation of state by

$$p = (\gamma - 1)\rho \qquad 1 < \gamma < 2 \qquad (2.13)$$

and from (1.29) we know that $\rho \propto R^{-3\gamma}$. We can see the qualitative behavior of Q by looking at its derivative with respect to the scale factor R,

$$\frac{dQ}{dR} = \frac{-6k}{R^3} + \frac{24\pi G_N\gamma\rho}{R}. \qquad (2.14)$$

Hence we see that for $k = -1,0$ Q has no extrema, begins at $-\infty$ (as $R \to 0$) and monotonically increases to $Q = 0$ as $R \to \infty$. When $k = +1$, Q again begins at $-\infty$. The Universe becomes curvature dominated (i.e., when the curvature term dominates over the energy density and constant terms), when $Q = 0$ (and Λ is sufficiently small). Q then has an extrema when $dQ/dR = 0$ which occurs at R_0 when

$$Q_{max} = 4\pi G_N(3\gamma-2)\rho. \qquad (2.15)$$

At larger values of R, Q monotonically decreases to $Q = 0$ as $R \to \infty$. The behavior of Q (R) is schematically shown in Fig. 1

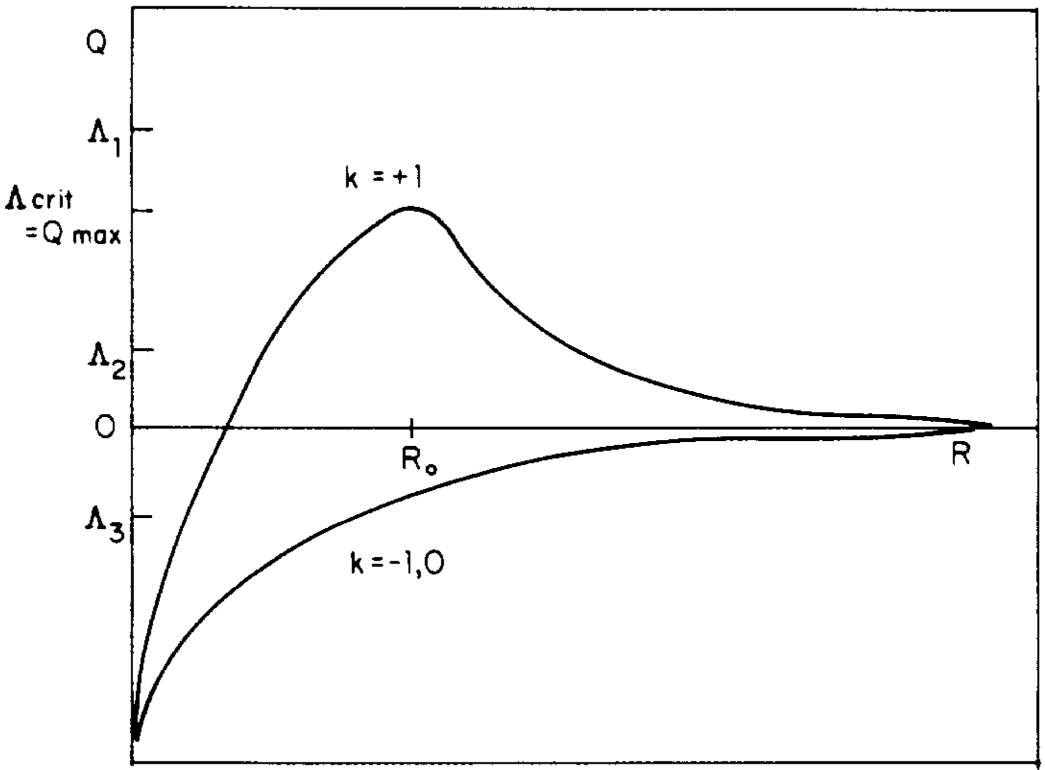

Fig. 1. Schematic plot of Q(R) for k = +1, 0, -1.

Let us begin the classification of cosmological models with those containing matter ($\rho \neq 0$) and with a variety of choices for Λ. Depending on the choice of Λ, we arrive at a wide class of models. The most interesting cases are those of closed Universes. ($k = +1$).

A. $k = +1$, $\Lambda = \Lambda_1 > Q_{max}$.

There are two solutions in this case corresponding to the + and - signs of Eq. (2.12). The + solution corresponds to an expanding Universe. The model begins with a singularity at $R = 0$ and expands to infinity. The expansion rate ($\dot{R}/R$) will have a minimum when $Q = Q_{max}$. As $R \to \infty$ the model approaches a De Sitter-type solution (see below) with a typical expansion rate given by

$$\dot{R}/R = (\Lambda/3)^{1/2} \quad (\Lambda >> Q) \tag{2.16}$$

or

$$R \sim \exp[(\Lambda/3)^{1/2}\, t]. \tag{2.17}$$

It is interesting to note that although, we have a closed Universe here, it does continue to expand forever without a recollapse typical of most closed models. This is of course due to the presence of the cosmological constant suppling an additional "force" for expansion. The - solution to (2.12) corresponds to a contracting Universe which begins ($t = 0$) at $R = \infty$ and contracts to a singularity. Both of these solutions would in a sense track the full curve for $k = +1$ in Fig. 1.

At this point, it is worthwhile to make two comments which apply to this solution and to those which follow. Nearly all of the solutions contain singularities in the space-time. This has been shown[6] to be a necessity for a wide class of models which satisfy the energy condition

$$R_{\mu\nu}X^{\mu}X^{\nu} \geqslant 0 \tag{2.18}$$

for every non-space-like vector X^{μ}. We will point out those solutions which do not satisfy this condition and hence do not have singularities. The second point is that the vast majority of models discussed do not correspond to our physical Universe. As we will discuss below, we know the value of Λ is very close to zero and hence $Q < 0$ today. We also know that the Universe is expanding, hence the contracting models are also not possibilities. We will point out those solutions which might be candidates for describing our present state.

B, $k = +1$, $\Lambda = \Lambda_{crit} = Q_{max}$.

This case actually has five independent solutions. The first of these is known as the Einstein Static Universe. It corresponds to the non-expanding, non-contracting solutions with $Q = \Lambda$ always and $R = R_0$ is constant. We see now, for the first time, that in order to completely describe this model, it is necessary to specify the value of R_0 or equivalently the total energy density. Recall in this case

$$\frac{k}{R_0^2} = \frac{1}{R_0^2} = 4\pi G_N \gamma \rho. \tag{2.19}$$

As we will see below, this is also equivalent to specifying the total entropy. This solution was the original motivation for Einstein to introduce the cosmological constant in order to cancel-out the expansion of the Universe. In this model, all quantities remain constant with time. This model also has no singularities in either the past or future.

The remaining four solutions all asymptotically approach the Einstein static model in either the past or the future. There are two solutions which begin with (very nearly) static solutions and depending on the sign in (2.12) either expand out towards $R = \infty$ (non singular) or contract in towards a singularity at $R = 0$. The other two solutions expand from a singularity or contract from $R = \infty$ (non singular) and asymptotically approach the static Universe in the future.

C. $k = +1, \ \Lambda_2 \geqslant 0$

To trace the evolution for this case, one can imagine cutting Fig. 1 horizontally at $Q = \Lambda_2$, there are then two separate possibilities. The first is such that the Universe begins at a singularity $R(0) = 0$ and expands until $Q = \Lambda_2$ at which time $\dot{R} = 0$, the expansions halt and the Universe begins its collapse back to a singularity. This is more typical of what we imagine for a closed Universe. Once again the total entropy must be specified in order to determine the maximum radius the Universe takes when $\dot{R} = 0$. For values of Λ_2 which are very close (or equal) to zero, this model is a possible candidate for our physical Universe.

The second possibility for this case is on the right-hand side of Fig. 1 for $Q < \Lambda_2$ and $R > R_0$. This solution is also non-singular but does not include $\Lambda_2 = 0$ as a possibility. In this solution the Universe initially contracts from $R = \infty$ until once again $Q = \Lambda_2 > 0$ when $\dot{R} = 0$ and the Universe "bounces" and begins to re-expand. Because this solution is surely not (nor ever was) dominated by matter or radiation, it does not represent a possible candidate.

D. $k = +1, \ \Lambda = \Lambda_3 < 0$

This solution is essentially identical to the first one of case C. It begins at a singularity and expands until $Q = \Lambda_3$ and then collapses. Unless Λ_3 is again very close to zero, this is not a realistic choice.

E. $k = -1, \ 0, \ \Lambda \geqslant 0$

Because the cosmological solutions to $k = -1$ and $k = 0$ are qualitatively similar, we will not distinguish them here. For $\Lambda \geqslant 0$, there are again two solutions corresponding to the two signs in (2.12). One solution is an expansion from an initial singularity to $R = \infty$. For $\Lambda = 0$, this is perhaps the most likely candidate. The second solution is a contraction from $R = \infty$ to a singularity at $R = 0$.

F. $k = -1, \ 0 \quad \Lambda < 0$

This case is qualitatively identical to case D.

G. $\rho = 0, \ k = +1, \ \Lambda > 0$

For completeness, we also present the possibilities in which there is no matter present. For $\rho = 0$, we have

$$Q = \frac{3k}{R^2} \tag{2.20}$$

The behavior of Q is shown in Fig. 2.

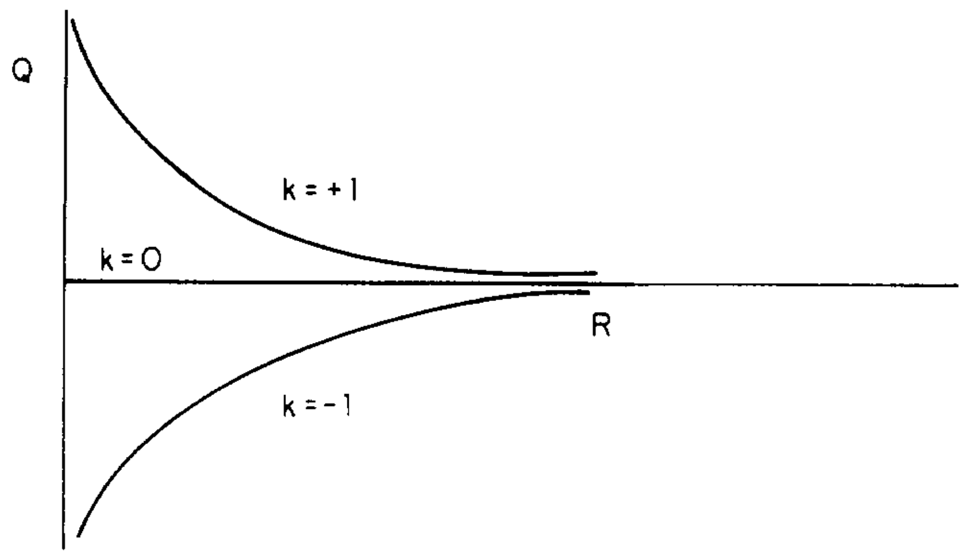

Fig. 2. Schematic plot of Q(R) for k = +1, 0, -1, and $\rho = 0$.

There is only one solution in this case which has the Universe initially contracting in from $R = \infty$ until $Q = \Lambda$ when the Universe bounces and begins to expand back out to $R = \infty$. For causal Universes, $k = +1$ forbids the possibility that $\Lambda < 0$.

H. $\rho = 0$, $k = 0$, $\Lambda > 0$

This solution most clearly has the De Sitter exponential expansion. It is indeed a space with constant curvature and constant expansion rate

$$H^2 = \dot{R}^2/R^2 = \Lambda/3. \tag{2.21}$$

It may either exponentially expand to $R = \infty$ or exponentially contract to a singularity from $R = \infty$. In this case, $\Lambda = 0$ corresponds to Minkowski space and $\Lambda < 0$ is again forbidden.

I. $\rho = 0$, $k = -1$, any value of Λ

This case is qualitatively identical to that of $k = -1,0$ with matter, cases E and F.

J. $\Lambda < Q$

As we have indicated, $Q < \Lambda$ is forbidden in a causal Universe. It does represent a class of solutions known as anti-De Sitter spaces and contains closed time-like paths. It will not be in the scope of these lectures to pursue these solutions any further.

The above set of cosmological models covers the full range of homogeneous and isotropic Universes. We will now briefly review some properties and observables of the Universe today. It is important to note that because of the scaling with R there is a sequential relevance of terms in the expansion rate. As we have seen in a matter-dominated Universe $\rho \sim R^{-3}$, while for a radiation-dominated Universe $\rho \sim R^{-4}$ which means that at early times (small R) the expansion rate will be dominated by either matter or radiation. Eventually, at large enough R if $k \neq 0$, the curvature term will begin to dominate until finally the Universe is controlled by the cosmological constant if it exists. Each of these periods has a different time-dependent expansion. In particular, we know that the expansion today is not governed by the cosmological term and hence we will assume $\Lambda = 0$ until section 7. With that, the Hubble parameter can be expressed as

$$H^2 = \frac{-k}{R^2} + \frac{8\pi G_N}{3}\rho. \tag{2.22}$$

We can define a critical energy density ρ_c such that $\rho = \rho_c$ for $k = 0$

$$\rho_c = \frac{3H^2}{8\pi G_N}. \tag{2.23}$$

In terms of the present value of the hubble parameter

$$\rho_c = 1.88\times10^{-29}\ h_0^2\ \mathrm{g\ cm}^{-3}, \tag{2.24}$$

where

$$h_0 = H_0/(100\ \mathrm{km\ Mpc^{-1}\ s^{-1}}) \tag{2.25}$$

is the present value of the Hubble parameter in units of 100 km Mpc^{-1} s^{-1}. The cosmological density parameter is then defined as the ratio of the present energy density to the critical density

$$\Omega \equiv \rho/\rho_c. \tag{2.26}$$

Furthermore, the value of Ω will determine the sign of k. For $\Omega > 0$ we have $k = +1$, $\Omega = 1$ corresponds to $k = 0$ and $\Omega < 0$ to $k = -1$. In terms of Ω the Friedmann equation can be rewritten as

$$(\Omega - 1)\,H_0^2 = \frac{k}{R^2}. \tag{2.27}$$

It is also useful to define an additional quantity known as the deacceleration parameter

$$q_0 \equiv -R\ddot{R}/\dot{R}^2 \qquad (2.28)$$

and from Eq. (1.27) we have

$$q_0 H_0^2 = 4\pi G_N (\rho + 3p)/3 \qquad (2.29)$$

or

$$2q_0 = (3\gamma-2)\Omega. \qquad (2.30)$$

The measurement of q_0 is extremely difficult and at best there is only a limit $q_0 < 2$ which for a matter-dominated present Universe corresponds to $\Omega < 4$. To be sure a low value q_0 corresponding to $\Omega < 1$ is also allowed indicating that as we can not even determine the sign of k, we are not as yet curvature dominated.

As we have said repeatedly, the Universe is expanding. This is perhaps the oldest discovery of modern cosmology and involves the measurement of the cosmological redshift. (In a contracting Universe there would be a cosmological blue shift.) If we take again our example of section 1 of two observers in which a light signal is emitted by 1 and received by 2, the redshift is then defined by

$$z \equiv \frac{\nu_1 - \nu_2}{\nu_2} = v_{12}/c \qquad (2.31)$$

for nearby observers, where ν_1, ν_2 are the emitted and observed frequencies corresponding to a relative velocity v_{12} between the observers. For large separations, care must be taken, and distances, and timescales, and hence velocities must be derived by integrating over the metric. However, v_{12} is determined by the expansion by

$$v_{12} = \dot{R}\ \delta r, \qquad (2.32)$$

where δr is the coordinate separation of the two observers. For light signals ($ds^2 = 0$) Eq. (2.32) can be rewritten as

$$v_{12} = \left(\frac{\dot{R}}{R}\right) R\delta r = H\delta t = Hd, \qquad (2.33)$$

where d is the physical separation of the two observers. Finally, we see that a determination of z and d will lead to a value of the Hubble parameter. Present limits are

$$50\ \mathrm{km\ Mpc^{-1}\ s^{-1}} \lesssim H_0 \lesssim 100\ \mathrm{km\ Mpc^{-1}\ s^{-1}} \qquad (2.34)$$

or

$$1/2 \lesssim h_0 \lesssim 1. \qquad (2.35)$$

In addition to determining the present density of the Universe or the density parameter by measuring q_0, it is also possible by means of measuring mass-to-light ratios. The mass of a galaxy or gravitational system if in gravitational equilibrium can be computed via the virial theorem from measured rotational velocities. The total mass of the system is then compared with its absolute luminosity which is derived from the measured apparent luminosity. The total density ρ is then

$$\rho = \left(\frac{M}{L} \right) \mathcal{L}, \tag{2.36}$$

where (M/L) is the above described mass-to-light ratio and $\mathcal{L}$ is the total luminosity density of the night sky[7])

$$\mathcal{L} \approx 2\times10^8 \, h_0 \, L_\odot \, \mathrm{Mpc}^{-3}, \tag{2.37}$$

where $L_\odot$ is the solar luminosity $L_\odot = 3.9\times10^{33}$ erg s^{-1}. We can now define a critical mass-to-light ratio

$$(M/L)_c = \rho_c/\mathcal{L} \approx 1200 \, h_0 \tag{2.38}$$

and the cosmological density parameter is given by

$$\Omega = (M/L)/(M/L)_c. \tag{2.39}$$

In principle this could give us an accurate determination of Ω. The problem is that the derived value of Ω seems to depend on what scale we measure (M/L). For example, the following four systems all give different values[8]) of Ω

1) solar neighborhood

$$(M/L) \sim 2 \pm 1 \Rightarrow \Omega \sim (0.0016 \pm 0.0008)/h_0$$

2) central parts of galaxies

$$(M/L) \sim (10\text{-}20) \, h_0 \Rightarrow \Omega \sim (0.008 - 0.017)$$

3) binaries and small groups of galaxies

$$(M/L) \sim (60\text{-}180) \, h_0 \Rightarrow \Omega \sim (0.05 - 0.15)$$

4) clusters of galaxies

$$(M/L) \sim (300\text{-}1000) \, h_0 \Rightarrow \Omega \sim (0.25 - 0.8).$$

The dependence on h_0 of the last three mass-to-light ratios is due to the uncertainties in estimating the mass and absolute luminosities of distant objects. It is evident that as we look on larger and larger scales the value of Ω seems to be increasing. This is known as the missing mass problem. In particular, it seems to indicate that there is dark matter present in the Universe on large scales. Neutrinos (and/or perhaps stable supersymmetric particles) are a popular candidate. Although there is no evidence for $\Omega > 1$ this alone is not sufficient to prove that we live in an open Universe.

There is one additional quantity which is a measureable relic of the big bang, that is, the temperature of microwave background radiation. If we take the premise that at very early times as $R \to 0$, the density of radiation becomes very high corresponding to a very high temperature, we should see a relic of this temperature today. In particular, as we will see shortly, the temperature of the radiation falls off as

$$T \sim 1/R \tag{2.40}$$

in an adiabatically expanding Universe. The radiation would have remained in thermal contact with the matter just until the recombination of free electrons and protons to make neutral hydrogen at about $T \sim 4000°K$. Subsequently, the radiation would have redshifted down to a very low temperature today. In fact, this radiation is exactly what was observed in 1965 by Penzias and Wilson[2] when they measured an isotropic blackbody with a temperature* of 2.7°K.

Today, the content of the microwave background consists of photons. We can calculate the energy density of photons by

$$\rho_\gamma = \int E_\gamma dn_\gamma, \tag{2.41}$$

where the density of states is given by

$$dn_\gamma = \frac{g_\gamma}{2\pi^2} \left[\exp(E_\gamma/T)-1\right]^{-1} q^2 dq \tag{2.42}$$

and $g_\gamma = 2$ simply counts the number of degrees of freedom for photons, $E_\gamma = q$ is just the photon energy (momentum). (We are using units such that $\hbar = c = k_B = 1$ and will do so throughout the remainder of these lectures.) On performing the integral in (2.41) we have that

$$\rho_\gamma = \frac{\pi^2}{15} T^4 \tag{2.43}$$

which is the familiar blackbody result.

In general, at very early times, at very high temperatures, other particle degrees of freedom join the radiation background when $T \sim m_i$ for each particle type i if that type is brought into thermal equilibrium through interactions. In equilibrium (we will define this

*The present range for the temperature is between 2.7 and 3°K. The original measurement by Penzias and Wilson was not as exact, they found $T_0 \simeq 3.5 \pm 1°K$.

notion more precisely shortly) the energy density of a particle type i is given by

$$\rho_i = \int E_i dn_{q_i} \tag{2.44}$$

and

$$dn_{q_i} = \frac{g_i}{2\pi^2} \left[\exp[(E_{q_i} - \mu_i)/T] \pm 1\right]^{-1} q^2 dq, \tag{2.45}$$

where again g_i counts the total number of degrees of freedom for type i,

$$E_{q_i} = (m_i^2 + q_i^2)^{1/2}. \tag{2.46}$$

μ_i is the chemical potential if present and $\pm$ corresponds to either fermi or bose statistics.

We can also define the other thermodynamic quantities such as the entropy density.

$$s_i = \frac{1}{T} \left[\int E_{q_i} dn_{q_i} \mp \frac{Tg_i}{(2\pi)^3} \int \ln\,(1 \mp n_{q_i})\, d^3q_i, \tag{2.47}$$

where

$$n_{q_i} = \left[\exp\left[(E_{q_i} - \mu_i)/T\right] \pm 1\right]^{-1}. \tag{2.48}$$

The free energy is just

$$F_i = \rho_i - Ts_i = \mu_i n_i - p_i \tag{2.49}$$

and

$$n_i = \int dn_{q_i}. \tag{2.50}$$

The chemical potential is generally taken for net baryon number. However, as we will discuss in section 3, the net baryon number

$$\eta = \frac{n_B - n_{\bar{B}}}{n_\gamma} \approx \frac{n_B}{n_\gamma} \sim 0(10^{-10}) \tag{2.51}$$

is very small, and one usually neglects the chemical potential.

At this point, it will be useful to note that the conservation of energy Eq. (1.29) also implies conservation of entropy. Having set $\mu = 0$, we can rewrite (2.49) as

$$s = \frac{1}{T}(\rho + p). \tag{2.52}$$

In addition, we have the thermodynamic identity

$$-\frac{dF}{dT} = \frac{dP}{dT} = s. \tag{2.53}$$

Equation (1.29) can also be rewritten as

$$R^3\frac{dp}{dt} = \frac{d}{dt}\left[R^3\ (\rho + p)\right]. \tag{2.54}$$

Combining these two equations, we have

$$\frac{d}{dt}\left[R^3\ s\right] = 0 \tag{2.55}$$

or conservation of total entropy.

We can now see how the specification of total entropy determines, for example, the maximum radius of a closed Universe. For a simple gas of photons $s = (1/T)\ (\rho + p) = (4/3T)\ \rho \propto T^3$ hence we have that

$$RT = \text{constant}. \tag{2.56}$$

Thus, in addition to the equation of state, we must specify the temperature (or total entropy) at the maximum radius, i.e., the constant in (2.56).

Returning to our discussion of a free gas at high temperatures, it will be useful to look at the limit at which $T \gg m_i$, i.e., a relativistic gas. In general, the total energy density is given by

$$\rho = \left(\sum_B g_B + \frac{7}{8}\sum_F g_F\right)\frac{\pi^2}{30}\,T^4 \equiv \frac{\pi^2}{30}\,N\ (T)\ T^4, \tag{2.57}$$

where $g_{B(F)}$ are the total number of boson (fermion) degrees of freedom and the sum runs over all boson (fermion) states with $m \ll T$. The factor of 7/8 is due to the difference between the fermi and bose integrals. Equation (2.57) defines N(T) by taking into account new particle degrees of freedom as the temperature is raised.

Once again, we can compute the pressure and entropy density in the high temperature limit

$$p = \rho/3 = \frac{\pi^2}{90}\,N\ (T)\ T^4 \tag{2.58}$$

$$s = 4/3\left(\frac{\rho}{T}\right) = \frac{2\pi^2}{45}\,N\ (T)\ T^3. \tag{2.59}$$

We can also rewrite Eq. (2.7) giving us a relationship between the age of the Universe and its temperature

$$t = (90/32\pi^3\ G_N N(T))^{1/2} T^{-2}. \tag{2.60}$$

Put into a more convenient form

$$tT^2_{MeV} = 2.4\ [N(T)]^{-1/2}, \tag{2.61}$$

where t is measured in seconds and T_{MeV} in units of MeV.

The value of N(T) at any given temperature depends on the particle physics model. In the standard SU(3)×SU(2)×U(1) model, we can specify N(T) up to temperatures of O(100)GeV. This is done in the following table.

Table 1

Temperature	New particles	4N(T)
$T < m_e$	γ's + ν's	29
$m_e < T < m_\mu$	$e^\pm$	43
$m_\mu < T < m_\pi$	$\mu^\pm$	57
$m_\pi < T < T_c^*$	π's	69
$T_c < T < m_{strange}$	π's + u, $\bar{u}$, d, $\bar{d}$ + gluons	205
$m_s < T < m_{charm}$	$s\bar{s}$	247
$m_c < T < m_\tau$	$c\bar{c}$	289
$m_\tau < T < m_{bottom}$	$\tau^\pm$	303
$m_b < T < m_{top}$	$b\bar{b}$	345
$m_t < T < m_W$	$t\bar{t}$	387

*T_c corresponds to the confinement-deconfinement transition between quarks and hadrons.

At higher temperatures, N(T) will be model dependent. For example, in the minimal SU(5) model, one needs to add to N(T), 6 states coming from $W^\pm$, Z, 24 for the X and Y gauge bosons, another 24 from the adjoint Higgs, and another 10 from the $\underset{\sim}{5}$. Hence for $T > M_X$ in minimal SU(5) N(T) = 160.75. In a supersymmetric model this would at least double, with some changes possibly necessary in the table if the selectron (scalar partner of the electron) has a mass below M_W.

Much of the preceding discussion has involved the notion of a temperature and all of the thermodynamic quantities (2.57-2.59) depend on the assumption that the particle states which are counted in N(T) must be in thermal equilibrium. Therefore, we will define this notion in the context of an expanding universe. Particle states will be said to be in thermal equilibrium if there is a reaction rate involving that state which is fast on an expansion time-scale.

If, for example, the Universe were not expanding, then given enough time, every particle state would come into equilibrium with each other. Because of the expansion of the Universe, certain rates might be too slow indicating, for example, in a scattering process that the two incoming states might never find each other to bring about an interaction. Depending on their rates, certain interactions may pass in and out of thermal equilibrium during the course of the Universal expansion. Quantitatively, for each particle i, we will require that some rate Γ_i involving that type be larger than the expansion rate of the Universe or

$$\Gamma_i > H \qquad (2.62)$$

in order to be in thermal equilibrium.

A good example of processes in equilibrium at some stage and out of equilibrium at others is that of neutrinos. If we consider the standard neutral or charged-current interactions such as $e^+ + e^- \to \nu + \bar{\nu}$ or $e + \nu \to e + \nu$, etc., very roughly the rates for these processes will be

$$\Gamma = \sigma\, n\, v, \qquad (2.63)$$

where σ will be taken as the weak interaction cross section

$$\sigma \sim O\,(10^{-2})\; T^2/M_W^4. \qquad (2.64)$$

n is the number density of leptons

$$n \sim T^3 \qquad (2.65)$$

and v their relative velocity ($v \sim 1$). Hence the rate for these interactions is

$$\Gamma_{wk} \sim O(10^{-2})\; T^5/M_W^4. \qquad (2.66)$$

The expansion rate, on the other hand, is just

$$\begin{aligned} H = \left(\frac{8\pi G_N \rho}{3}\right)^{1/2} &= \left(\frac{8\pi^3}{90}\, N(T)\right)^{1/2} T^2/M_p \\ &\sim 1.66\; N(T)^{1/2}\; T^2/M_p, \end{aligned} \qquad (2.67)$$

where the planck mass is defined by

$$M_p = G_N^{-1/2} \simeq 1.22\times10^{19}\ \text{GeV}. \qquad (2.68)$$

Neutrinos will be in equilibrium when $\Gamma_{wk} > H$ or

$$T > (500\; M_W^4/M_p)^{1/3} \sim 1\ \text{MeV}. \qquad (2.69)$$

The temperature at which these rates are equal is commonly referred to as the decoupling or freeze-out temperature and is defined by

$$\Gamma\,(T_d) = H(T_d). \qquad (2.70)$$

For temperatures $T > T_d$, neutrinos will be in equilibrium, while for $T < T_d$ they will not. Basically, in terms of their interactions, the expansion rate is just too fast and they never "see" the rest of the matter in the Universe (nor themselves). Their momenta will simply redshift and their effective temperature (the shape of their momenta distribution is not changed from that of a blackbody) will simply fall with $T \propto 1/R$. It is interesting to note that at very high temperatures neutrinos were again out of equilibrium. The interaction rate for $T \gg M_W$ behaves as

$$\Gamma \sim \frac{g^4}{T^2} \cdot T^3, \tag{2.71}$$

where g is some gauge or Yukawa coupling then $\Gamma \sim H$ when

$$g^4 T \sim N^{1/2}\ T^2/M_p \tag{2.72}$$

or

$$T \sim N^{-1/2}\ g^4\ M_p \sim 10^{15}\ \text{GeV}. \tag{2.73}$$

Thus at temperature scales much higher than the GUT scale, equilibrium might be a very bad approximation.[9]

At the epoch of neutrino decoupling, $T \sim 1$ MeV, the neutrino "temperature" is still equal to the thermal background temperature which includes only photons, electrons, and positrons. Soon after decoupling the $e^\pm$ pairs begin to annihiliate (when $T \lesssim m_e$). The energy released, as we will see, is served to heat up the photon background relative to the neutrinos. Because the neutrinos are decoupled, their entropy must be conserved separately from the entropy of interacting particles. If we call $T_>$, the temperature of photons, $e^\pm$ before annihilation, we also have $T_\nu = T_>$ as well. The entropy density at $T = T_>$ is just

$$s_> = \frac{4}{3}\ \frac{\rho_>}{T_>} = \left(\frac{4}{3}\right)\left(\frac{\pi^2}{30}\right)\left(\frac{11}{2}\right) T_>^3 \tag{2.74}$$

while at $T = T_<$, the temperature of the photons just after annihilation, the entropy density is

$$s_< = \frac{4}{3}\ \frac{\rho_<}{T_<} = \left(\frac{4}{3}\right)\left(\frac{\pi^2}{30}\right)(2)\ T_<^3 \tag{2.75}$$

and by conservation of entropy (2.55) $s_< = s_>$ and

$$(T_</T_>)^3 = 11/4. \tag{2.76}$$

Thus, the photon background is at higher temperature than the neutrinos because the annihilation energy could not be shared among the neutrinos, and

$$T_\nu = (4/11)^{1/3}\ T_\gamma \approx 1.9°\text{K}. \tag{2.77}$$

The same type of phenomena would also occur if there are other neutral weakly interacting particles which decoupled at higher temperatures.[10] A possible example of such a particle might be the graviton. If we assume that the decoupling temperature in this case is $T_d = M_p$ then the photons would have received the energy due to the annihilation of every other particle species relative to the graviton. In this case $s_> = (4/3)\ (\pi^2/30)\ N\ (M_p)\ T^3$ and $s_< = 4/3\ (\pi^2/30)\ (2)\ T^3$ so that $T_G = (2/N(M_p))^{1/3} T\gamma$. As we have said for minimal SU(5), we must have $N(M_p) \gtrsim 160$ or $T_G < 0.7°K$. In a minimal supersymmetric model $N(M_p) \gtrsim 350$ and $T_G < 0.5°K$.

This concludes the review of the standard cosmological models. We will now build on this by tackling the questions on the origins of elements, the baryon asymmetry and even the homogeneity and isotropy.

Section 3. Big Bang Nucleosynthesis

The two most important pieces of evidence in support of the standard big bang model are the observation[2)] of the 3°K microwave background radiation and the explanation[1)] of the origin of the light elements and their abundances. Because of the initially high temperatures and densities and the large abundance of neutrons relative to protons, the chains of nuclear reactions similar to those occurring in stars might have occurred. Indeed in the simplest model of nucleosynthesis, one can compute the produced abundances of deuterium, ^{3}He, ^{4}He and ^{7}Li and one finds an amazing degree of agreement with the observed abundances (The observations which must be compared with the big bang abundances must be from sources where little or no subsequent nucleosynthesis has taken place.) In this section we will look closely at the predictions of big bang nucleosynthesis and its cosmological consequences in terms of limits on particle physics.

The temperature region of interest is one typical of nuclear energies, i.e., $T \sim 1$ MeV. The initial conditions for the problem will therefore be set at $T >> 1$ MeV. Once again, because the asymmetry between baryons and antibaryons is so small and since we do not expect very different asymmetries among the leptons (standard GUT models even predict their similarity) we will take all chemical potentials to be zero. One of the chief quantities of interest will be the neutron-to-proton ratio (n/p). At very high temperatures ($T >> 1$ MeV), the weak interaction rates for the processes

$$\begin{aligned} n + \nu_e &\leftrightarrow p + e^- \\ n + e^+ &\leftrightarrow p + \bar{\nu}_e \\ n &\leftrightarrow p + e^- + \bar{\nu}_e \end{aligned} \tag{3.1}$$

were all in equilibrium, i.e., $\Gamma_W > H$. Thus we would expect that initially $(n/p) \approx 1$. Actually in equilibrium, the ratio is essentially controlled by the boltzmann factor so that

$$(n/p) \simeq \exp(-\Delta m/T), \tag{3.2}$$

where $\Delta m = m_n - m_p$ is the neutron-proton mass difference. For $T >> \Delta m$, $(n/p) \simeq 1$.

At tempertures $T >> 1$ MeV, nucleosynthesis can not begin to occur even though the rate for forming the first isotope, deuterium, through

$$n + p \leftrightarrow D + \gamma \tag{3.3}$$

is sufficiently rapid. To begin with, at $T \gtrsim 1$ MeV deuterium is photodissociated because $E_\gamma > 2.2$ MeV (the binding energy of deuterium; $E_\gamma \approx 2.7T$ for a blackbody). Furthermore, the density of photons is very high $n_\gamma/n_B \sim 10^{10}$. Thus the onset of nucleosynthesis will depend on the quantity

$$\eta^{-1} \exp[-2.2 \text{ MeV}/T] \tag{3.4}$$

where

$$\eta \equiv n_B/n_\gamma \tag{3.5}$$

is the baryon to photon ratio. When this quantity (3.4) becomes $\lesssim O(1)$, the rate for $p + n \to D + \gamma$ finally becomes greater than the rate for dissociation $D + \gamma \to p + n$. This occurs when $T \sim 0.1$ MeV or when the Universe is a little over 2 min. old.

Because nucleosynthesis begins when $T < 1$ MeV, the rates for processes which control (n/p) (3.1) as well as those which keep neutrinos in equilibrium are frozen out. As we have seen, neutrinos are effectively at a lower temperature at $T \lesssim 1/2$ MeV, this must be taken account in the expansion rate (2.67) which now has*

$$\begin{aligned} N_\rho &= N_\gamma + \left(\frac{4}{11}\right)^{4/3} N_\nu \\ &= 2 + \left(\frac{4}{11}\right)^{4/3} \cdot 3 \cdot \frac{7}{4} = 3.36. \end{aligned} \tag{3.6}$$

Furthermore, because the rates for processes (3.1) also freeze out (at $T \lesssim 1$ MeV), the neutron to proton ratio must be adjusted from its equilibrium value. When freeze out occurs, the ratio (n/p) is relatively fixed at

$$(n/p) \sim 1/6. \tag{3.7}$$

This equilibrium value is adjusted by taking into account the free neutron decays up until the time at which nucleosynthesis begins. This reduces the ratio to

$$(n/p) \sim 1/7. \tag{3.8}$$

Since virtually all the neutrons available end up in deuterium which gets quickly converted to ^{4}He, we can estimate the ratio of the ^{4}He nuclei formed compared with the number of protons left over

$$X_4 \equiv (N_{^4He}/N_H) = \frac{1}{2}(n/p)/\left(1 - (n/p)\right) \tag{3.9}$$

or more importantly the ^{4}He mass fraction

*We distinguish between N_ρ and $N_s = N_\gamma + (4/11) N_\nu = 3.91$ because of the difference in the temperature dependence of $\rho \propto T^4$ and $s \propto T^3$.

$$Y_4 \equiv 4X_4/(1 + 4X_4) = 2(n/p)/\left(1 + (n/p)\right). \tag{3.10}$$

For $(n/p) \approx 1/7$, we estimate that $Y_4 \approx 0.25$ which is very close to the observed value.

The actual calculated value of Y_4 will depend on a numerical calculation which runs through the complete sequence of nuclear reactions.[11] Once deuterium is produced by Eq. (3.3), tritium can be produced by

$$D + D \rightarrow T + p \tag{3.11}$$

which then gets converted to make ^4He by

$$D + T \rightarrow {}^4He + n. \tag{3.12}$$

^4He has in addition several other processes which go towards its production

$$\begin{aligned} D + D &\rightarrow {}^4He + \gamma \\ {}^3He + n &\rightarrow {}^4He + \gamma \\ {}^3He + D &\rightarrow {}^4He + p \\ {}^3He + {}^3He &\rightarrow {}^4He + 2p \\ T + p &\rightarrow {}^4He + \gamma. \end{aligned} \tag{3.13}$$

Additional processes for producing T and ^3He include

$$\begin{aligned} n + D &\rightarrow T + \gamma \\ p + D &\rightarrow {}^3He + \gamma \\ D + D &\rightarrow {}^3He + n. \end{aligned} \tag{3.14}$$

The nuclear chain is temporarily halted at this point because there are gaps at masses A = 5 and A = 8, i.e., there are no stable nuclei with those masses. There is some further production, however, which accounts for the abundances of ^6Li and ^7Li through

$$\begin{aligned} &{}^3He + {}^4He \rightarrow {}^7Be\,|\, + \gamma \\ &\qquad\qquad\qquad\quad \rightarrow {}^7Li + e^+ + \nu_e \\ &T + {}^4He \rightarrow {}^7Li + \gamma \\ &p + {}^7Li \rightarrow {}^4He + {}^4He, \text{ etc.} \end{aligned} \tag{3.15}$$

Once again because of the gap at A = 8 there is very little subsequent nucleosynthesis in the big bang. A second chief factor in the ending of nucleosynthesis is that during this whole process the Universe continues to expand and cool. At lower temperatures it becomes exponentially difficult to overcome the Coulomb barriers in nuclear collisions. In spite of these effects, numerical calculations of the elemental abundance continue the chain up until Al.

Before reviewing the results of the big bang nucleosynthesis[11-14)] calculations, it is important to realize that there are three additional parameters which have a very strong effect on the results.* They are 1) the baryon-to-photon ratio η (3.5); 2) the neutron half-life $\tau_{1/2}$; 3) the number of light particles or, in particular, the number of neutrino flavors N_ν.

As we have seen above, the value of η controls the onset of nucleosynthesis (3.4). Basically what happens is that for a larger baryon-to-photon ratio η the quantity (3.4) becomes smaller thus allowing nucleosynthesis to begin earlier at a higher temperature. Remember also that a key ingredient in determining the final mass fraction of ^{4}He, Y_4, was (n/p) [see Eq. (3.10)] and that the final value of (n/p) was determined by the time at which nucleosynthesis begins thus controlling the time available for free decays after freeze out. If nucleosynthesis begins earlier, this leaves less time for neutrons to decay and the value of (n/p) and hence Y_4 is increased.

The value of η can not be determined directly from observations. If we break it up and try to look individually at the number density of baryons and photons present in the Universe today** we find that

$$n_B = \rho_B/m_B = \Omega_B\rho_c/m_B \tag{3.16}$$
$$= 1.13 \times 10^{-5}\Omega_B h_0^2 \text{ cm}^{-3},$$

where ρ_B is the energy density in baryons, m_B is the nucleon mass, Ω_B is that part of Ω (2.26) which is in the form of baryons and ρ_c is the critical energy density (2.23). The number density of photons is just given by

$$n_\gamma = \int dn_\gamma = \left(2\zeta(3)/\pi^2\right) T_\gamma^3 \tag{3.17}$$

*Once again we are not considering the effects of a chemical potential, which can also greatly vary the results.[14a)]

**The baryon-to-photon ratio should not have changed since nucleosynthesis. At these energy scales, baryon number is conserved and there are no major processes which would have produced entropy to change the photon number.

or in more useful units

$$n_\gamma = 400\ (T_0/2.7)^3\ \mathrm{cm}^{-3}, \tag{3.18}$$

where T_0 is the present temperature of the microwave background radiation. Putting η back together we find

$$\eta = 2.81 \times 10^{-8}\Omega_B h_0^2 (2.7/T_0)^3. \tag{3.19}$$

Thus we could determine η if we knew Ω_B, h_0, and T_0. It is important to keep in mind however that physically it is η which is determined at the level of microphysics through baryon generation (see section 6) and not the other way around.

If we now use the limits on h_0 and T_0 we can get a feeling of where η lies in terms of Ω_B. If we use $2.7°K \leq T_0 \leq 3°K$ and $1/2 \leq h_0 \leq 1$ we find that

$$5.1 \times 10^{-9}\Omega_B \leq \eta \leq 2.8 \times 10^{-8}\Omega_B. \tag{3.20}$$

Furthermore, as we saw in section 2, depending on what scale we consider as typically representing the overall density we might have Ω_B in the range 0.0008 - 0.8. Thus we see that the observations leave us with more than three orders of magnitude uncertainty. As we will see shortly, however, consistency of the standard big bang nucleosynthesis model allows only for a factor of about 20 in Ω and only about 2-3 in η.

The second parameter, $\tau_{1/2}$, is important in that it also plays a role in determining the value of Y_4. Although we don't usually consider $\tau_{1/2}$ a parameter, the uncertainties in its measured value are significant from the point of view of nucleosynthesis. After all, it is this quantity which will control the weak interaction rates and hence determine the freeze-out temperature. The common value of $\tau_{1/2} \simeq 10.6$ min. is actually uncertain by about two percent and this is enough to affect the production of ^4He. The range we will consider is

$$10.4 \text{ min.} < \tau_{1/2} \leq 10.8 \text{ min.}$$

As in the case of η, increasing $\tau_{1/2}$ leads to a larger value of Y_4. We can see this by looking again at a comparison between the weak interaction rates and the expansion rate. If we parametrize the weak interaction rate by $\Gamma_{wk} = AT^5$ and the expansion rate by $H = BT^2$ then the freeze-out temperature is given by Eq. (2.70)

$$T_d^3 = B/A. \tag{3.22}$$

If we now increase $\tau_{1/2}$, this corresponds to decreasing $\Gamma_{wk} \sim \tau_{1/2}^{-1}$ or decreasing the value of A. This in turn gives a higher value for T_d. Now if T_d is larger, this will give a larger value of (n/p) at freeze-out via Eq. (3.2) and hence more ^4He via Eq. (3.10).

The final input parameter, we said was the number of light particles. Specifically, what we mean is the number of degrees of freedom corresponding to particles which are still relativistic ($m \ll T$) when $T < O(1)$ MeV. In addition, we must require that these particles be relatively stable so that they will be present when freeze-out occurs, thus $\tau >$ few seconds. As we hinted to above, likely candidates for these particles are neutrinos and thus the number of neutrino flavors N_ν becomes important. Of course any other types of light particles such as photinos or axions, etc., may also be important.*

The number of neutrino flavors N_ν will also affect the primordial abundance of ^{4}He and like η and $\tau_{1/2}$, increasing N_ν increases Y_4. The expansion rate (2.67) is proportional to $[N(T)]^{1/2}$. At $T \gtrsim 1$ MeV, $N(T)$ is given by

$$N(T) = 2 + \frac{7}{2} + \frac{7}{4} N_\nu \tag{3.23}$$

which takes into account the contribution of γ's, $e^\pm$'s, and N_ν flavors of neutrinos. Thus increasing N_ν, increases B in the notation of Eq. (3.22) and again leads to higher value of T_d, with the same effect of producing more ^{4}He.

Let us now look at the observations[15] which tells us the abundances of the light elements. In particular, we will be interested in the abundances of D, ^{3}He, ^{4}He, and ^{7}Li. Deuterium is the most easily destroyed of the light elements. It is also very difficult to produce in astrophysical systems where it is not further processed to form ^{3}He. Therefore, any of the observed D is generally assumed to be primordial. Furthermore because deuterium is so easily destroyed (or burned) we must assume that the abundance of D produced in the big bang is greater than the observed value or

$$(D/H)_{BB} > (D/H)_{OBS}, \tag{3.24}$$

where (D/H) is the ratio (by number) of deuterium to hydrogen.

The abundance of deuterium is found by a number of methods which include the analysis of meteoritic and solar wind data,[16] line spectra from the atmospheres of the giant planets,[17] and ultraviolet absorption studies on interstellar gas[18] which indicate that D/H lies in the range $(1-4) \times 10^{-5}$ and a good lower limit to the deuterium abundance would be

$$(D/H) > (1 - 2) \times 10^{-5}. \tag{3.25}$$

In Fig. 3, we have plotted[14] the produced D/H ratio in big bang

*For particles which interact more weakly than neutrinos, care must be taken in that they may have a lower temperature if they have decoupled earlier.[10]

nucleosynthesis as a function of η for $N_\nu = 3$ and $\tau_{1/2} = 10.6$ min. (The ^{4}He abundance is the most sensitive to variations in these qualities, see Fig. 5.) If we now require that $(D/H)_{BB} > (1 - 2) \times 10^{-5}$ from Eqs. (3.24) and (3.25), we find an upper limit[14] to the baryon-to-photon ratio η

$$\eta < (7 - 10) \times 10^{-10}. \qquad (3.26)$$

Any larger value of η would have led to increased burning rates for D in the big bang with very little D left over.

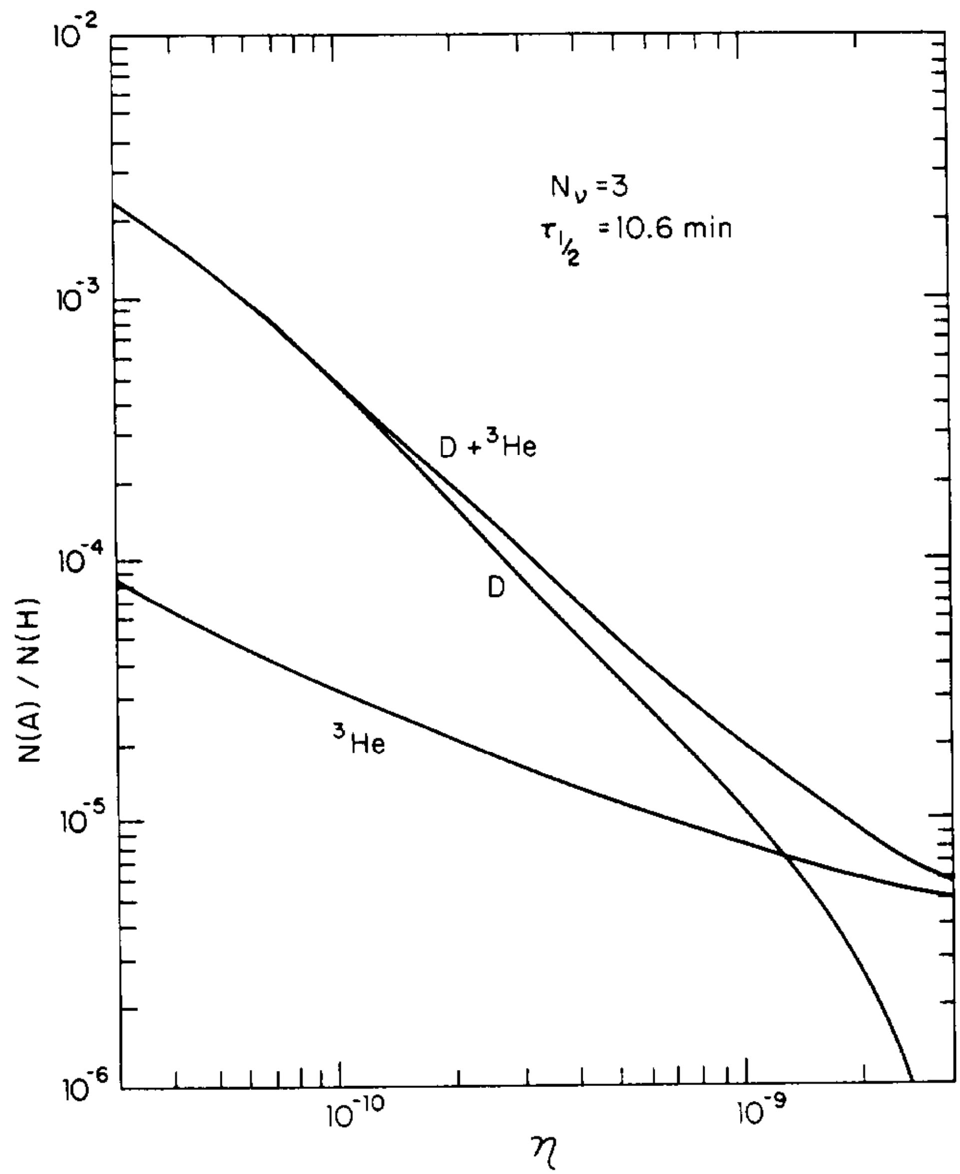

Fig. 3. The abundances (by number relative to hydrogen) of D, ^{3}He and their sum as a function of η for $N_\nu = 3$ and $\tau_{1/2} = 10.6$ min.

In general, direct measurements of ^{3}He are very difficult. Once again, most of the available data is derived from meteoritic and solar wind data.[16] However, as we have said, deuterium is burned to ^{3}He in the sun, hence the solar wind data actually reflects the presolar abundance of the sum (D + ^{3}He)/H. The observations indicate that

$$(D + {}^3He)/H|_{pre\odot} \lesssim 4 \times 10^{-5}. \tag{3.27}$$

Meteoritic data[19] can give the pre-solar abundance of ^{3}He/H where it is supposed that no processing of D takes place. These measurements indicate that

$${}^3He/H|_{pre\odot} \lesssim 2 \times 10^{-5}. \tag{3.28}$$

Unlike deuterium, ^{3}He is very difficult to destroy in its entirety in stellar systems. Pre-main-sequence stars are very efficient in burning deuterium to ^{3}He via $D + p \rightarrow {}^3He + \gamma$. ^{3}He is only destroyed at high temperatures ($T > 7 \times 10^6$ °K) through ${}^3He + {}^3He \rightarrow {}^4He + 2p$ and ${}^3He + {}^4He \rightarrow {}^7Be + \gamma$. At higher temperatures ($T > 10^8$ °K), ^{4}He is burned to carbon and oxygen. The point is that, in general, some fraction g of the initial ^{3}He abundance will survive stellar processing. If one takes into account the fact some of this ^{3}He is redeposited in the interstellar medium (pre-solar) then in terms of g we have

$$(D + {}^3He)/H|_{BB} \leq (D/H)|_{pre\odot} + \frac{1}{g}\,({}^3He/H)|_{pre\odot} \tag{3.29}$$

which can be rewritten as

$$(D + {}^3He/H)|_{BB} \leq (D + {}^3He)/H|_{pre\odot} + \left(\frac{1}{g} - 1\right)\,{}^3He/H|_{pre\odot} \tag{3.30}$$

The value of g, however, can only be determined[20] by models of stellar evolution and in fact may differ depending on the mass of the star. In low mass stars ($M < 8M_\odot$), $g > 0.7$ is not unreasonable while for high mass stars ($8M_\odot < M < 100\ M_\odot$), g may be as low as 1/4. Since an initial spectrum of stellar masses would cover all ranges, perhaps a lower limit to g of 1/2 - 1/4 would be safe.

If we put together the presolar limits on (D + ^{3}He)/H and ^{3}He/H with the above limit on g we have an upper limit on the sum of primordial ^{3}He + D,

$$(D + {}^3He)/H|_{BB} \leq (6 - 10) \times 10^{-5}. \tag{3.31}$$

In Fig. 3, we have also displayed the behavior of the ^{3}He/H and (D + ^{3}He)/H ratios as a function of η. In this case an upper limit on (D + ^{3}He)/H$|_{BB}$ corresponds to a lower limit[14] on η,

$$\eta \geq (3 - 4) \times 10^{-10} \tag{3.32}$$

from Eq. (3.31). Putting this together with the upper limit on η Eq. (3.26) we have

$$(3\text{-}4) \times 10^{-10} \leq \eta \leq (7 - 10) \times 10^{-10} \quad (3.33)$$

as the range of η consistent with the abundances of D and ^3He.

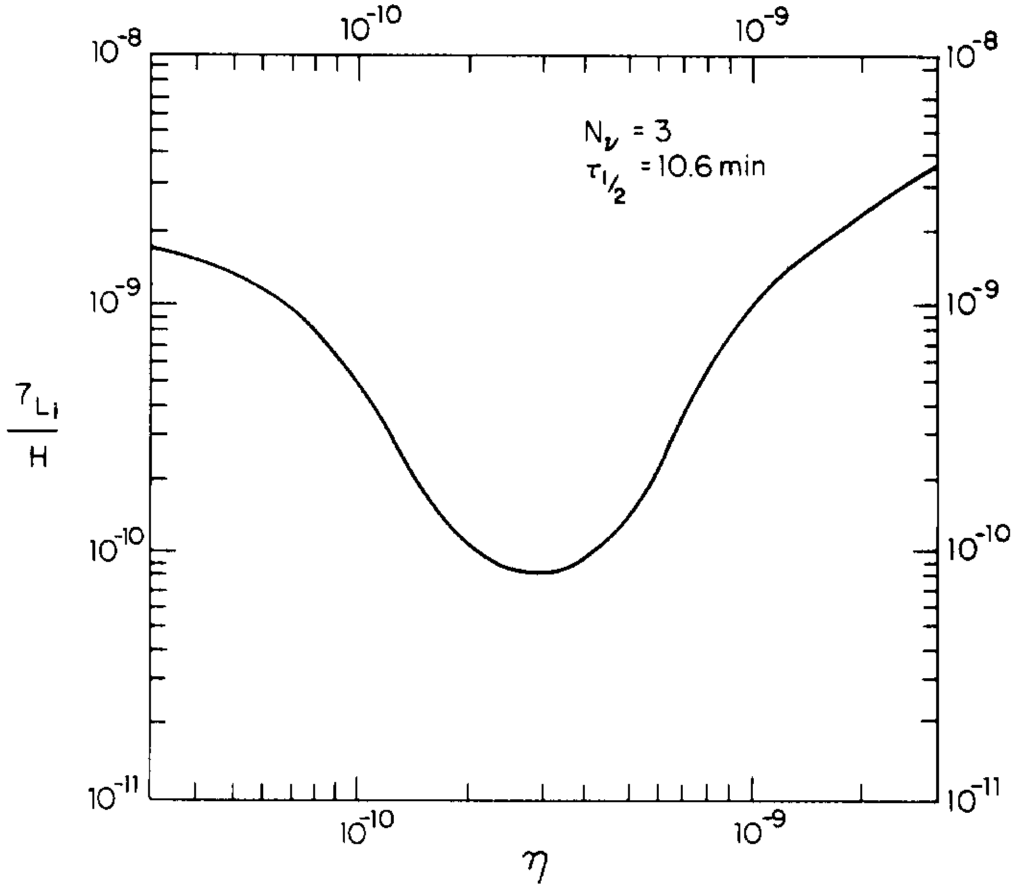

Fig. 4. The abundance (by number relative to hydrogen) of ^7Li as a function of η for $N_\nu = 3$ and $\tau_{1/2} = 10.6$ min.

^7Li is another isotope which is in principle difficult to draw solid conclusions from. The main difficulty is that ^7Li is both easily produced as well as destroyed. Recently, however, there have been some measurements[21)] of the ^7Li abundance in some very old Population II stars. Since some ^7Li might have been destroyed before the formation of these stars, we might expect $(^7Li/H)_{PopII} \leq (^7Li/H)_{BB}$. (The present ^7Li abundance would be larger still representing the contribution from stellar processing.) The observed limit on the ^7Li abundance is

$$({}^7\mathrm{Li/H})_{\mathrm{PopII}} \lesssim 1.5 \times 10^{-10}. \qquad (3.34)$$

In Fig. 4, we show[14] the calculated ratio of ^{7}Li/H as a function of η for $N_\nu = 3$ and $\tau_{1/2} = 10.6$ min. The upper limit Eq. (3.34) corresponds to bounds on η of $(2 - 5) \times 10^{-10}$. The calculated rates for ^{7}Li, however, have uncertainties which lead to uncertainties in the predicted abundance of ^{7}Li by about a factor of 2. Thus the only safe bounds on η from ^{7}Li are

$$10^{-10} < \eta < 7 \times 10^{-10} \qquad (3.35)$$

remarkably consistent with the bound Eq. (3.33) from D and ^{3}He.

This brings us to ^{4}He which is probably the most important of the isotopes studied. The main reason ^{4}He is so important is that there is so much of it. Next to hydrogen it is the most abundant element around and its abundance is quite well known. Unlike the other light elements which have observational uncertainties of $\gtrsim 100\%$, the ^{4}He abundances are measured to within a few per cent. The main problem is that it is also produced in stars and care must be taken in trying to derive the "observed" primordial abundance.

To be sure, one can place an upper limit on the primordial abundance by $Y_{4BB} < Y_{4OBS}$ (Y_4, remember is the total ^{4}He mass fraction). However, in order to use big bang nucleosynthesis to set limits on particle physics (e.g., N_ν) a much more accurate determination of Y_{4BB} is needed. Spectral measurements[22] of galactic HII regions give very accurate values of Y_4, however, there they have been contaminated with by-products of stellar processing. The observations of galaxies with low metal abundances could in principle yield an accurate value of Y_{4BB} but these measurements are difficult because these galaxies are typically very far away. It is not possible within the scope of these lectures to cover completely the discussion of Y_4. The best estimates consistent with the observations place Y_4 in the range

$$0.22 \lesssim Y_4 \lesssim 0.25. \qquad (3.36)$$

If we restrict ourselves as before to $N_\nu = 3$, $\tau_{1/2} = 10.6$ min., the upper limit on Y_4 implies an upper limit on η from Fig. 5

$$\eta \lesssim 5 \times 10^{-5} \qquad (3.37)$$

which is once again consistent with the previous limits Eq. (3.33). (The lower limit on Y_4 does not give an interesting bound on η.)

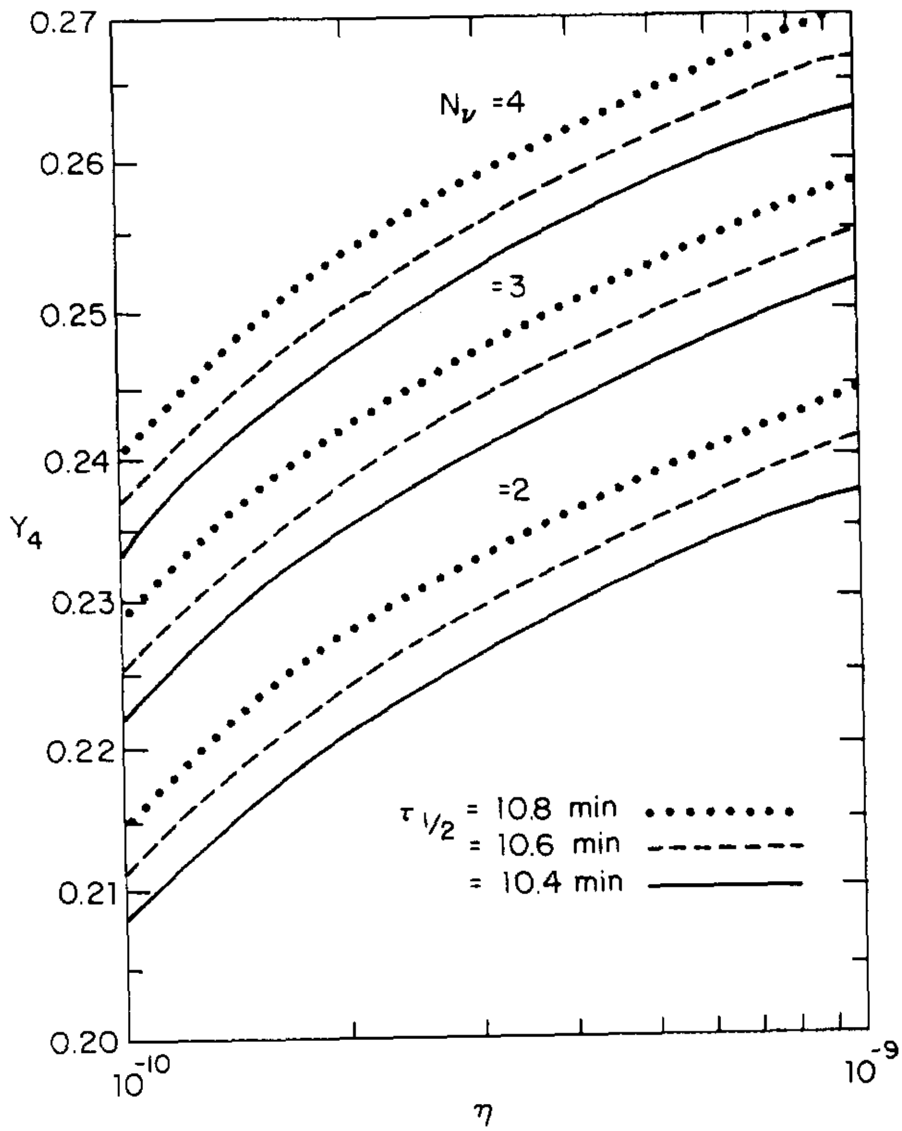

Fig. 5. The abundance (by mass) of ^{4}He as a function of η for N_ν = 2,3 and 4 and for $\tau_{1/2}$ = 10.4 min (solid), 10.6 min (dashed), and 10.8 min (dotted).

Figure 5 actually contains significantly more information than just a limit on η. In Fig. 5, we see clearly the behavior of Y_4 with respect to all three parameters: η, $\tau_{1/2}$, and N_ν. It is clear how Y_4 increases with increasing values of any of the three parameters. It is also immediately clear that we can set a limit[12-14] on N_ν provided that we have a lower limit to η. Using $\eta > 3 \times 10^{-10}$ and $Y_4 < 0.25$, we find that $N_\nu \leq 4$ with the equality being at best marginal. This implies that at most one more generation is allowed, assuming that the neutrinos associated with each generation are light and stable.

The strong dependence of Y_4 on the three parameters requires great precision to strengthen the limits due to nucleosynthesis. Strictly speaking, $\eta > 3 \times 10^{-10}$ and $\tau_{1/2} > 10.4$ min. allows $N_\nu = 4$ only if $Y_4 \gtrsim 0.253$; however, we are not yet in a position to believe the third decimal place. For $\tau_{1/2} \gtrsim 10.4$ min., the limit Eq. (3.37) on η can be relaxed so that $N_\nu \leqslant 3$, $Y_4 < 0.25$ implies $\eta < 7 \times 10^{-10}$. We can also turn the limits around and set a lower limit to the helium abundance by assuming $\eta > 3 \times 10^{-10}$ and $N_\nu \geqslant 3$ then we have $Y_4 > 0.24$. If future observations actually yield $Y_4 < 0.24$, one would have to argue that perhaps ν_τ is heavy and unstable (the present limit is only $m_{\nu_\tau} < 250$ MeV). If we only assume $N_\nu \geqslant 2$, then the lower limit on Y_4 becomes $Y_4 \geqslant 0.22$. Any observation of the primordial helium abundance less than 0.22 would indicate an inconsistency with the standard model.

The importance of the success of big bang nucleosynthesis can not be overstressed. The abundance of the predicted elements differ by about nine orders of magnitude, from ^{7}Li to ^{4}He. Furthermore, all of these predictions are consistent with the observations only for a narrow range of the baryon-to-photon ratio $(3 - 4) \times 10^{-10} \leqslant \eta \leqslant (7 - 10) \times 10^{-10}$. It is incorrect to think of this as a drawback (in the sense that one would be more comfortable with a large range for η). On the contrary, it is evidence of the predictive power of the model. Indeed, the Universe has only one value of η, Y_4, etc. If we just concentrate on a central value for η, say $\eta = 5 \times 10^{-10}$ with $N_\nu = 3$ and $\tau_{1/2} = 10.6$ min., we have $D/H = 3 \times 10^{-5}$, $^3He/H = 1.2 \times 10^{-5}$, $^7Li/H = 1.2 \times 10^{-10}$, and $Y_4 = 0.25$, all remarkably consistent with the observations.

There is still one more important consequence of the above limits, that is the limit on η can be converted to a limit on the baryon density and Ω_B. If we turn around Eq. (3.19), we have

$$\Omega_B = 3.56 \times 10^7 \eta h_0^{-2} (T_0/2.7)^3, \tag{3.38}$$

and using the limits on η Eq. (3.33), h_0 Eq. (2.35) and T_0 from (2.7 - 3)°K we find a range for Ω_B

$$0.01 \leqslant \Omega_B \leqslant 0.19. \tag{3.39}$$

Recall that for a closed Universe $\Omega > 1$, thus from Eq. (3.39) we can conclude that the Universe is not closed by baryons. This does not exclude the possibility that other forms of matter (e.g., massive neutrinos, etc.) exist in large quantities to provide for a large Ω. In fact, if large clusters of galaxies were representative of Ω the limit from nucleosynthesis would indicate that some form of dark matter must exist.

This concludes our review of big bang nucleosynthesis. We again stress its credibility and importance because it takes place at an energy scale in which we feel we understand the physics. We will not be able to make the same statement for the last three sections.

Section 4. Limits on Neutrino Masses

In this section, we will look at a typical class of constraints available from cosmology: those on particle masses and lifetimes. Most of these limits can be extended to other types of neutral particles, but we will take neutrinos as a canonical example. We will begin the discussion with stable neutrinos and then move towards the constraints on unstable neutrinos.

For cosmological purposes, a stable particle is one with lifetime $\tau > 10^{24}$ sec. The limits on stable neutrinos can be divided up into three mass regions: 1) light neutrinos, $m_\nu < 1$ MeV; 2) massive neutrinos, $m_\nu > 1$ MeV; and 3) very massive neutrinos with a lepton asymmetry. Let us begin with the limits on light neutrinos. Neutrinos with mass less than 10^{-4}eV are still relativistic and hence, equivalent to zero mass neutrinos and are allowable. For $m_\nu > 10^{-4}$eV we can compute the total mass density in neutrinos by

$$\rho_\nu = m_\nu n_\nu, \tag{4.1}$$

where n_ν is the number density of neutrinos

$$n_\nu = (\tfrac{3}{4}) g_\nu \zeta(3)\ T_\nu^3/\pi^2, \tag{4.2}$$

where g_ν is the number of degrees of freedom for a massive neutrino. Depending on the particle physics model, there are two choices for g_ν. For Dirac mass neutrinos $g_\nu = 4$ (i.e., the mass term in the Lagrangian is similar to that for an electron $\propto \bar{\nu}\nu$). For Majorana mass neutrinos, $g_\nu = 2$ (the mass term is $\propto \nu\nu$). We can put Eq. (4.2) in terms of the photon number density Eq. (3.17)

$$n_\nu = \frac{3}{4}\left(\frac{g_\nu}{2}\right) n_\gamma \left(\frac{T_\nu}{T_\gamma}\right)^3 = \frac{3}{22}\, g_\nu n_\gamma. \tag{4.3}$$

If we compare the neutrino mass density to the critical mass density Eq. (2.23) we can write down that part of Ω which is due to neutrinos

$$\Omega_\nu = \rho_\nu/\rho_c = 0.01\ m_\nu(\mathrm{eV}) h_0^{-2}\left(\frac{g_\nu}{2}\right)(T_0/2.7)^3 \tag{4.4}$$

for neutrino masses in eV. Although the limits on Ω and h_0 taken separately yield a limit $\Omega h_0^2 \lesssim 4$ this would imply an age for the Universe $\tau_U < 8 \times 10^9$ yrs which is much too small. We regret that we are not able here to go into the details involving the age of the Universe arguments, but consistency requires* that $\Omega h_0^2 < 1$. We thus have the following limit[24] on the mass of light-stable neutrinos,

*A tighter limit assuming $\tau_u > 1.3 \times 10^{10}$ yrs gives[23] $\Omega h_0^2 \lesssim 0.25$.

$$m_\nu(\text{eV}) < 200\ g_\nu^{-1}\ \text{eV}. \tag{4.5}$$

Actually the limit Eq. (4.5) is additive in the sense that if more than one species of light neutrinos (or other types of particles at the same temperature) exist, the limit becomes

$$\sum_\nu m_\nu(\text{eV}) < 200\ g_\nu^{-1}\ \text{eV}, \tag{4.6}$$

i.e., the sum of all light masses must be less than 200 g_ν^{-1}. Thus, for Dirac-type neutrinos

$$\sum_\nu m_\nu < 50\ \text{eV} \tag{4.7}$$

while for Majorana-type neutrinos

$$\sum_\nu m_\nu < 100\ \text{eV}. \tag{4.8}$$

It is interesting to note that from Eq. (4.4) we see that to close the Universe with light neutrinos we need a total mass $\Sigma m_\nu > 200\ h_0^2 g^{-1}(2.7/T_0)^3$ which is only possible with $h_0 < 1$.

The limits on more massive neutrinos[25-26] are qualitatively different and involve a numerical calculation. The difference is that unlike the case for light neutrinos, massive neutrinos have a chance to reduce their number density through annihilations. In equilibrium, the neutrino number density is essentially governed by the Boltzmann factor $\exp(m_\nu/T)$. It is only after the rates for neutrino annihilations freeze out ($\Gamma_A < H$) does the neutrino density become fixed. For light neutrinos the rates all froze out at 1 ~ 1 MeV. For more massive neutrinos, freeze out will occur at a higher temperature (typically $T_f \sim m_\nu/20$) because the density of neutrinos has fallen so as to render $n\sigma < H$.

In order to calculate the number density of neutrinos, one must solve the Boltzmann equation

$$\frac{dn}{dt} = -3Hn - \langle\sigma v\rangle\ (n^2 - n_0^2), \tag{4.9}$$

where n_0 is the equilibrium number density, $H = -\dot{T}/T$ takes into account the Universal expansion. $\langle\sigma\upsilon\rangle$ is the thermal average of the cross section

$$\langle\sigma\upsilon\rangle \approx G_F^2 m_\nu^2\ N_A/2\pi, \tag{4.10}$$

where G_F is the Fermi weak interaction constant and N_A is the number of annihilation channels. Solving Eq. (4.9), one finds[26] for the total mass density

$$\rho_\nu = m_\nu n_\nu = (1.6 \times 10^{-27} \text{gcm}^{-3})[m_\nu(\text{GeV})]^{-1.85} \times \times N_A^{-0.95}\left(\frac{g_f}{2}\right)^{-0.52}, \tag{4.11}$$

where $g_f \equiv N(T_f)$ is the number of degrees of freedom at the temperature at which the annihilations freeze out. Dividing Eq. (4.11) by ρ_c we again have an expression for Ω_ν

$$\Omega_\nu = 2.66\ m_\nu^{-1.85}(\mathrm{GeV}), \tag{4.12}$$

where we have taken $N_A^{-0.95}\ (g_f/2)^{-0.52} \approx 3 \times 10^{-2}$ as a typical value. Once again, using the limit $\Omega h_0^2 < 1$ implies

$$m_\nu \gtrsim 1.7\ \mathrm{GeV}, \tag{4.13}$$

for each type of heavy neutrino.

The limit Eq. (4.13) can actually be strengthened by realizing that neutrinos in the GeV range (or higher) would cluster[27] with galaxies, binaries, and small groups of galaxies. In this case, we should not use the cosmological limit $\Omega h_0^2 < 1$ but rather the limit on Ω coming from binaries and small groups (see section 2) $\Omega h_0^2 < 0.15$. Thus for clustering neutrinos we have

$$m_\nu \gtrsim 4.7\ \mathrm{GeV}. \tag{4.14}$$

If we consider still larger neutrino masses, although their annihilation rates are effective enough to reduce their number density, if we assume that there is a slight excess of neutrinos over antineutrinos (or visa versa), the slight excess will remain even after the annihilations have ceased. In the simplest models of baryon generation (see the following section) we expect that a slight asymmetry on the order of the baryon asymmetry be produced. It would indeed be difficult to imagine that absolutely no asymmetry results since baryons and leptons are mixed in the processes which produce the baryon asymmetry. Thus, let us assume that

$$\eta_\nu \equiv (n_\nu - n_{\bar\nu})/n_\gamma = \delta\eta, \tag{4.15}$$

where δ is a model-dependent factor which we will suppose is $O(1)$. Independent of the annihilations, the left-over mass density will then be

$$\rho_\nu = m_\nu \eta_\nu n_\gamma = m_\nu \delta\eta n_\gamma \tag{4.16}$$

and

$$\Omega_\nu = 3.8 \times 10^7 m_\nu(\mathrm{GeV})\eta h_0^{-2}(T_0/2.7)^3\delta. \tag{4.17}$$

Taking $\eta > 3 \times 10^{-10}$ and $\Omega h_0^2 < 1$ we have an additional upper limit[28] on the neutrino mass

$$\sum_\nu m_\nu \lesssim 88\ \delta^{-1}\ \mathrm{GeV}, \tag{4.18}$$

where we have taken a sum over all species with asymmetry $\delta\eta$. As before, if we apply the limit due to the clustering around galaxies $\Omega h_0^2 < 0.15$ we have the following limit

$$m_\nu \lesssim 13\ \delta^{-1}\ \text{GeV}. \tag{4.19}$$

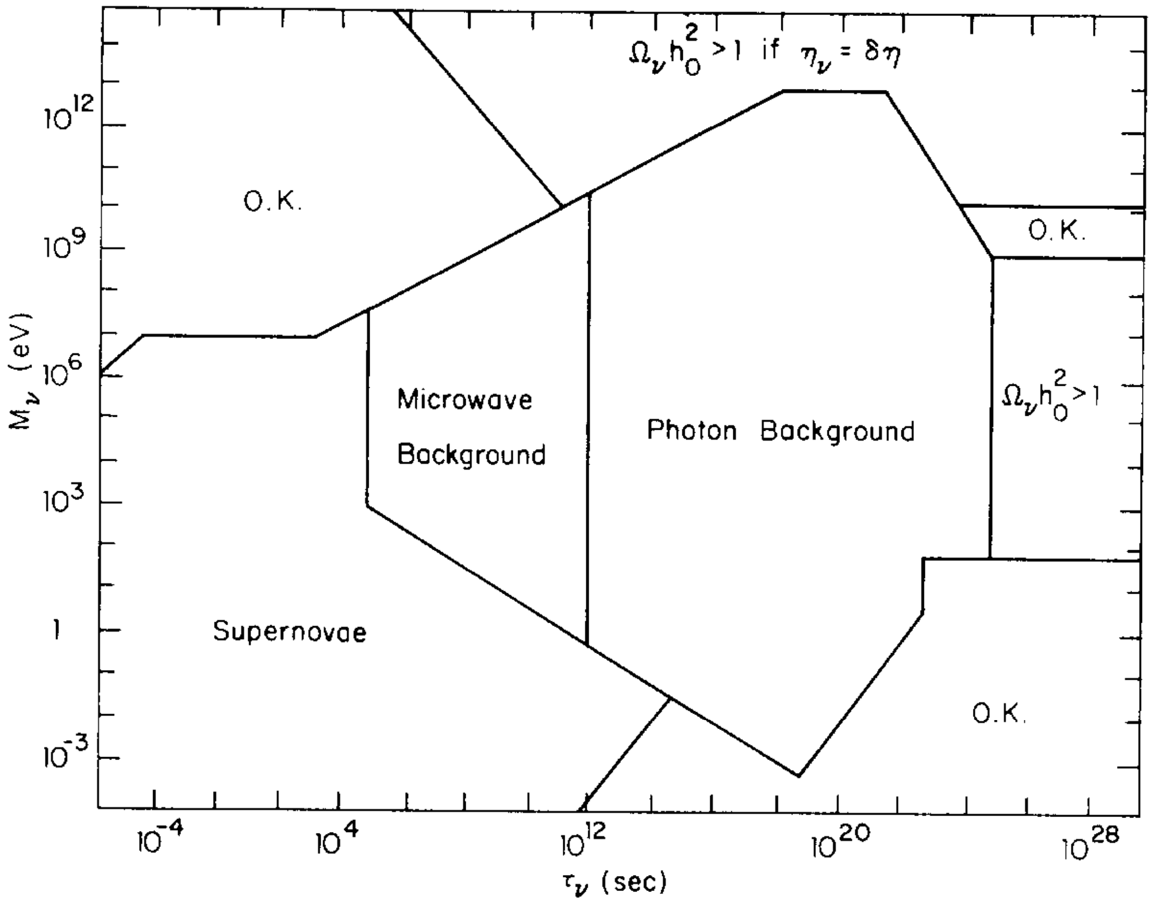

Fig. 6. Astrophysical constraints on the lifetimes and masses of neutrinos.

We will now briefly describe the limits that one can place on unstable neutrinos. These limits as well as for stable neutrinos are all summarized[27,29] in Fig. 6. The limits on ustable neutrinos [30,27] all assume that a significant portion of the decay products involve photons or charged particles. There are basically three ranges for lifetimes which use different arguments to rule them out. The first range is for τ_ν between 10^{12} sec and 10^{24} sec. During this range neutrinos decay after the decoupling of photons during the recombination period of neutral hydrogen formation. Because the decay occurs after decoupling, the decay photons remain unthermalized and would show up in the UV, X- and γ-ray backgrounds. Limits on the observed fluxes of these backgrounds rule out neutrinos with masses up to about 1 TeV (see Fig. 6). Heavier neutrinos would not be sufficiently abundant to interfere with these backgrounds unless, again, there was some asymmetry between ν and $\bar{\nu}$.

Neutrinos with intermediate lifetimes 2000 sec $\lesssim \tau \lesssim 10^{12}$ sec decay early enough that their decay products can still be thermalized with the microwave background. However, they distort the spectrum in that they produce too many photons for a given temperature. Once again, for sufficiently high masses, the abundances are sufficiently low so as to be acceptable.

The final lifetime range is for neutrinos with 10^{-3}sec $< \tau$ $<$ 2000 sec. Neutrinos with this lifetime and a mass less than 10 MeV are ruled out[31)] by supernova energetics. During the formation of a neutron star or a black hole, a supernova releases a total of 10^{53} ergs of which 10^{51} ergs is visible, the remainder being in the form of neutrinos. Neutrinos which decay in this range would yield supernovae which produce more than 10^{51} ergs in visible light and are hence ruled out.

Once again we stress that although we have limited this discussion to neutrinos, generalizations are applicable to other particle types as well. We hope only to show the power of cosmological and astrophysical limits on particle properties.

Section 5. Big Bang Baryosynthesis

As we have seen in the first three sections of these lectures, the big bang model very successfully explains the expansion of the Universe, the existence of the cosmic background radiation, and the abundances of the light elements. As we will try to show in this section, the big bang model when combined with GUTs can explain the origin of the baryon asymmetry and the value of η. Up until now, we have simply used η as a parameter without regard to its origin. Indeed, it is very strange that such a small parameter exists. On the other hand, it is difficult to understand why there is an asymmetry in the first place, i.e., why isn't $\eta = 0$?

There is a fair amount of evidence that indeed $\eta \neq 0$. There is, to begin with, no evidence within the solar system of antimatter. In addition, the cosmic rays show only evidence of $\bar{p}$'s with no $\bar{\alpha}$'s which would be necessary to definitively argue in favor of antimatter in the cosmic rays ($\bar{p}$'s can and are produced as secondaries in collisions). It has also been argued[32)] that if in clusters of galaxies certain galaxies were of one form of matter or the other, the γ-ray flux produced by $p\bar{p}$ annihilations in the cluster would exceed the limits on the observed γ-ray flux. Furthermore, it is very difficult to imagine a mechanism leading to such large separations of matter and antimatter when one accepts that at early times they were so well mixed.

Let us for the moment, assume that in fact $\eta = 0$. Then just as in the case for neutrinos (section 4) we can compute the final number density of nucleons left over after annihilations have frozen out. At very high temperatures $T > 1$ GeV, nucleons were in thermal equilibrium with the photon background and $n_N = n_{\bar{N}} = 3/2 n_\gamma$ (a factor of 2 accounts for neutrons and protons and the factor 3/4 for the difference between fermi and bose statistics). As the temperature fell below m_N, annihilations kept the nucleon density at its equilibrium value $(n_N/n_\gamma) = (m_N/T)^{3/2} \exp(-m_N/T)$ until the annihilation rate $\Gamma_A \approx n_N m_\pi^{-2}$ fell below the expansion rate. This occurred at $T \approx 20$ MeV. However, at this time the nucleon number density had already dropped to

$$n_N/n_\gamma = n_{\bar{N}}/n_\gamma \approx 10^{-18}, \tag{5.1}$$

which is eight orders of magnitude too small[32)] aside from the problem of having to separate the baryons from the antibaryons. If any separation did occur at higher temperatures (so that annihilations were as yet incomplete) the maximum distance scale on which separation could occur is the causal scale related to the age of the Universe at that time. At $T = 20$ MeV, the age of the Universe was only $t \approx 2 \times 10^{-3}$ sec. At that time, a causal region (with distance scale defined by $2ct$) could only have contained $10^{-5}\ M_\odot$ which is very far from the galactic-mass scales which we are asking for separations to occur, $10^{12}\ M_\odot$.

A final possibility might be statistical fluctuations, but in a region containing 10^{12} $M_\odot$, there are $\sim 10^{80}$ photons so that one would only expect statistical fluctuations to produce an asymmetry $\eta \sim 10^{-40}$! Thus we are left with the problem as to the origin of a small non-zero value for η. We can assume that it was an initial condition to start off with and in a baryon number conserving theory it would remain nearly constant. [The production of entropy (photons) could cause it to fall.] In this case, however, we must still ask ourselves, why is it so small? A more attractive possibility, however, is to suppose that the baryon asymmetry was in some way generated by the microphysics. Indeed, if one can show that a small non-zero value for η developed from $\eta = 0$ (or **any** other value) as an initial condition, we could consider the question solved. In the rest of this section, we will look at this second possibility for generating a non-zero value of η using GUTs.[33)]

There are three basic ingredients necessary[3)] to generate a non-zero η. They are

1. baryon number violating interactions
2. C and CP violation
3. a departure from thermal equilibrium.

The first condition is rather obvious, unless there is some mechanism for violating baryon number conservation, baryon number will be conserved and an initial condition such as $\eta = 0$ will remain fixed. C and CP violation indicate a direction for the asymmetry. That is, should the baryon number violating interactions produce more baryons than antibaryons? If C or CP were conserved, no such direction would exist and the net baryon number would remain at zero. The final ingredient is necessary in order to insure that not all processes are actually occurring at the same rate. For example, in equilibrium if every process which produced a positive baryon number was accompanied by an equivalent process which destroyed it, again no net baryon number would be produced.

The first two of these ingredients are contained in GUTs, the third in an expanding universe where it is not uncommon that interactions come in and out of equilibrium. In SU(5), for example, as we have seen in the earlier contributions to these lectures, the fact that quarks and leptons are in the same multiplets allows for baryon non-conserving interactions such as $e^- + d \rightarrow \bar{u} + \bar{u}$, etc., or decays of the supermassive gauge bosons X and Y such as $X \rightarrow e^- + d, \bar{u} + \bar{u}$. Although today these interactions are very ineffective because of the masses of the X and Y bosons, in the early Universe when $T > M_X \sim 10^{15}$ GeV these types of interactions should have been very important.

As we have also seen in the earlier contributions, C and CP violation is very model dependent. In the minimal SU(5) model, the magnitude of C and CP violation is too small to yield a useful value of η. The C and CP violation in general comes from the interference between tree level and first loop corrections. We refer the reader to those contributions for further details.

As we have said, the departure from equilibrium is very common in the early Universe when interaction rates can not keep up with the expansion rate. In fact, the simplest (and most useful) scenario for baryon production makes use of the fact that a single decay rate goes out of equilibrium. It is commonly referred to as the out of equilibrium decay scenario.[34] The basic idea is that the gauge bosons X and Y (or Higgs bosons) may have a lifetime long enough to insure that the inverse decays have already ceased so that the baryon number is produced by their free decays.

More specifically, let us call X, either the gauge boson or Higgs boson, which produces the baryon asymmetry through decays. Let α be its coupling to fermions. For X a gauge boson, α will be the GUT fine structure constant, while for X a Higgs boson, $(4\pi\alpha)^{1/2}$ will be the Yukawa coupling to fermions. The decay rate for X will be

$$\Gamma_D \sim \alpha\, M_X. \tag{5.2}$$

However decays can only begin occurring when the age of the Universe is longer than the X lifetime Γ_D^{-1}, i.e., when $\Gamma_D \geqslant H$

$$\alpha\, M_X \gtrsim N(T)^{1/2} T^2/M_p \tag{5.3}$$

or at a temperature

$$T^2 < \alpha\, M_X M_p N(T)^{-1/2}. \tag{5.4}$$

Scatterings on the other hand proceed at a rate

$$\Gamma_s \sim \alpha^2 T^3/M_X^2 \tag{5.5}$$

and are hence not effective at lower temperatures. In equilibrium, therefore, decays must have been effective as T fell below M_X in order to track the equilibrium density of X's (and $\bar{X}$'s). Thus the condition for equilibrium is that at $T = M_X$, $\Gamma_D > H$ or

$$M_X \lesssim \alpha\, M_p(N(M_x))^{-1/2} \sim 10^{18}\alpha \text{ GeV}. \tag{5.6}$$

In this case, we would expect no net baryon asymmetry to be produced.

For masses $M_X \gtrsim 10^{18}\alpha$ GeV, the lifetime of the X bosons is longer than the age of the Universe when $T \sim M_X$. Decays finally begin to occur when $T < M_X$, however, the density of X's is still comparable to photons $n_X/n_\gamma \sim 1$ whereas the equilibrium density at $T < M_X$ is $n_X/n_\gamma \sim (M_X/T)^{3/2} \times \exp[-M_X/T] \ll 1$. Hence, the decays are occurring out of equilibrium (inverse decays are not occurring). Hence, we have the possibility for producing a net asymmetry.

Let us now look at what happens during the decay of an $X,\bar{X}$ pair. If we consider the example of the X gauge boson and its decays to $\bar{u},\bar{u}$ with branching ratio r and net baryon number change $\Delta b_1 = -2/3$ and to e^-, d with branching ratio 1-r, and net baryon number change $\Delta b_2 = +1/3$

$$X \xrightarrow[r]{} \bar{u} + \bar{u} \qquad \Delta b_1 = -2/3 \tag{5.7a}$$

$$X \xrightarrow[1-r]{} e^- + d \qquad \Delta b_2 = +1/3 \tag{5.7b}$$

A similar set of decays will occur for $\bar{X}$

$$\bar{X} \xrightarrow[\bar{r}]{} u + u \qquad \Delta b_{\bar{1}} = +2/3 \tag{5.8a}$$

$$\bar{X} \xrightarrow[1-\bar{r}]{} e^+ + \bar{d} \qquad \Delta b_{\bar{2}} = -1/3 \tag{5.8b}$$

If C and CP are violated then $r \neq \bar{r}$ and we can define the total net baryon number produced per decay of X and $\bar{X}$

$$\Delta B = (\Delta b_1)r + (\Delta b_2)(1-r) + (\Delta b_{\bar{1}})\bar{r} + (\Delta b_{\bar{2}})(1-\bar{r})$$

$$= \bar{r} - r. \tag{5.9}$$

The value of $\bar{r}-r$ will of course depend on the specific model for C and CP violation.

The total baryon density that will have been produced by the $X,\bar{X}$ pair [provided Eq. (5.6) is not satisfied] is

$$n_B = (\Delta B)n_X \tag{5.10}$$

and since we also have $n_X = n_{\bar{X}} = n_\gamma$,

$$n_B = (\Delta B)n_\gamma. \tag{5.11}$$

Although the net baryon number is conserved during the subsequent evolution of the Universe, the photon number density is not. A more useful quantity just after baryon generation is the baryon-to-specific entropy ratio, n_B/s. The entropy density, recall from Eq. (2.59) is

$$s = \frac{2\pi^2}{45} N(T)T^3 \tag{5.12}$$

and can be related to the photon number density Eq. (3.17) by

$$s = \frac{\pi^4}{45\zeta(3)} N(T)n_\gamma. \tag{5.13}$$

At $T \lesssim M_X \sim 10^{15}$ GeV, we expect $N(T) \gtrsim O(100)$ so that $s \sim O(100)\ n_\gamma$. Thus the baryon-to-entropy ratio we would expect to produce in the out-of-equilibrium decay scenario would be

$$n_B/s \sim 10^{-2}(\Delta B). \tag{5.14}$$

The value of n_B/s that we are looking for must be related to the limits on η (Eq. 3.33) found in our discussion of nucleosynthesis. From the time of nucleosynthesis to the present, the contribution to N is due only to photons and the cooler neutrinos, $N_S = 3.91$ [see footnote related to Eq. (3.6)] and hence

$$s = 7.04\ n_\gamma. \tag{5.15}$$

Thus, η in the range $(3\text{-}10)\times10^{-10}$ corresponds to a value of n_B/s in the range $(4.3\text{-}14)\times10^{-11}$. Comparing this with the expected production, Eq. (5.14) gives us a lot of hope that GUTs may provide us with a viable mechanism for generating a small (but not too small) value for η.

Although we can be encouraged by the above scenario, we must still show that given a GUT, after the full set of Boltzmann equations have been integrated, an acceptable and definite value of η emerges. In particular, most GUTs do satisfy Eq. (5.6), for $\alpha = 1/41$ and $M_X \sim 10^{15}$ GeV decays will be occurring at $T \sim M_X$ but in at best partial equilibrium. Thus the estimate, Eq. (5.14) is not a good one. We will now outline what needs to be done in order to calculate in general the baryon-to-photon ratio in a GUT, in terms of the C and CP violating parameter ΔB.

As we have just said, a full solution to this problem requires solving[35)] a set of Boltzmann equations. In Eq. (4.9) we wrote down the Boltzmann equation for annihilations. Here we will be more interested in equations governing decays

$$\frac{dn_X}{dt} = -3Hn_X - \Gamma_D\left(n_X - n_{X_0}\right), \tag{5.16}$$

where n_{X_0} is the equilibrium number density of X's. In general, we can define the number of particles of type i in a phase-space element $dVd\Pi_i$ as

$$dN_i = u_\alpha p_i^\alpha N_i(p_i^\mu, x_i^\mu)dVd\Pi_i, \tag{5.17}$$

where u_α is the velocity four-vector, N_i is the phase-space density, and

$$d\Pi_i = \frac{1}{(2\pi)^3}\, g_i d^3p_i/p_i^0 \tag{5.18}$$

is the momentum element for i with g_i degrees of freedom. The number density of type i is just

$$n_i = \int u_\alpha p_i^\alpha N_i d\Pi_i = \frac{g_i}{2\pi^2} \int N_i p^2 dp_i \qquad (5.19)$$

which reduces to Eqs. (2.45) and (3.17) for

$$N_i = [\exp[(p_i^0 - \mu_i)/T] \pm 1]^{-1}. \qquad (5.20)$$

If there were no interactions which could change the number of i's, then the density would change as

$$dn_i/dt = -3Hn_i \qquad (5.21)$$

simply due to the expansion of the Universe.

When we add interactions, we have that

$$\frac{dn_i}{dt} = -3Hn_i + \sum_{\substack{j,\ldots \\ \ell,m,\ldots}} \int d\Pi_i d\Pi_j \ldots d\Pi_\ell d\Pi_m \ldots$$
$$\times [N_\ell N_m \cdots (1 \pm N_i)(1 \pm N_j) \cdots W(p_\ell p_m \cdots \to p_i p_j \cdots)$$
$$-N_i N_j \ldots (1 \pm N_\ell)(1 \pm N_m) \ldots W(p_i p_j \ldots \to p_\ell p_m \ldots)], \qquad (5.22)$$

where the factors $(1 \pm N_i)$ are the stimulated emission and exclusion factors for bosons and fermions. The invariant transition rate is

$$W = (s/2^n)|\mathcal{M}|^2 (2\pi)^4 \, \delta^4(\sum_{in} p_i^\mu - \sum_{out} p_\ell^\mu), \qquad (5.23)$$

where s is a statistical term containing $(m!)^{-1}$ for each set of m identical incoming or outgoing particles and n is the total number of particles in the process. $\mathcal{M}$ is the invariant amplitude for the process.

For practical purposes, one generally uses only Maxwell-Boltzmann statistics so that all factors $(1 \pm N_i)$ are neglected. We now have a prescription of how to calculate the number densities of each particle type i which involves a change in baryon number. We must also include all interactions which carry a change in baryon number such as decays, annihilations, and scatterings. In the simplest case where we only consider decays and inverse decays we can reduce Eq. (5.22) and take X's as an example

$$\frac{dn_X}{dt} = -3Hn_X + \int d\Pi_X d\Pi_{\bar{u}_1} d\Pi_{\bar{u}_2} [N_{\bar{u}_1} N_{\bar{u}_2} W(\bar{u}_1 \bar{u}_2 \to X)$$
$$-N_X W(X \to \bar{u}_1 + \bar{u}_2)]$$

$$+ \int d\Pi_x d\Pi_e d\Pi_d \lfloor N_e N_d W(e + d \to X)$$

$$- N_x W(X \to e + d) \rfloor . \qquad (5.24)$$

In Fig. 7, we look at the typical results which one finds after a complete numerical integration[35)] of the Boltzmann equations. These particular resluts are for an SU(5) model, but their behavior is generic for most any GUT. What is plotted is the time development of the baryon-to-entropy ratio n_B/s normalized to the net baryon number produced by pair decay ΔB. The horizontal scale, M_x/T, is proportional to $t^{1/2}$. The three curves correspond to different choices for the mass of the boson X. In curve 1, we have chosen, a mass which we expect to satisfy the out-of-equilibrium condition $M_x \simeq 3 \times 10^{18}\alpha$ and we indeed find that the maximum asymmetry has been generated $n_B/s \approx 10^{-2}\Delta(B)$ as we expected (5.14). This in itself confirms the original idea.

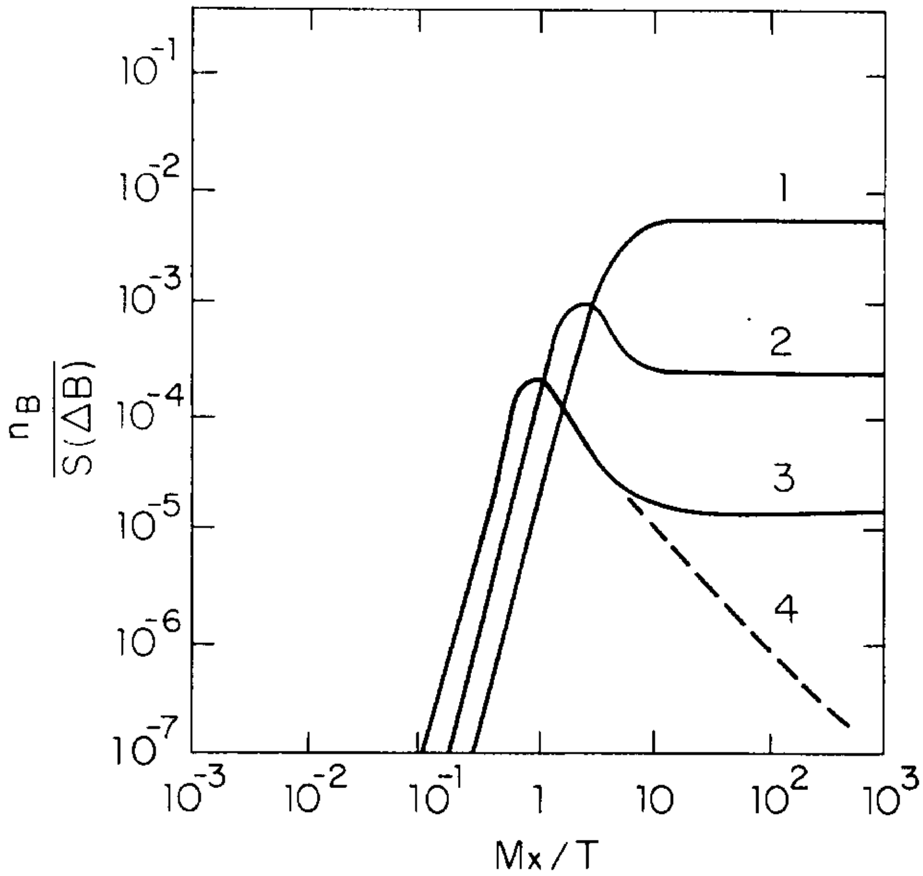

Figure 7. The time evolution of the baryon asymmetry in units of (ΔB) for 1) $M_x \simeq 3\times10^{18}\ \alpha$; 2) $M_x \simeq 3\times10^{17}\ \alpha$; 3) $M_x \simeq 3\times10^{16}\ \alpha$; and 4) if scatterings remain very effective.

The good news that we find from Fig. 7 is that even for lower masses, an asymmetry is still produced. In curve 2, we have chosen $M_X = 3\times10^{17}\alpha$ and we find still a substantial asymmetry $n_B/s \sim 10^{-4}$ (ΔB). What is happening is that at $T \sim M_X$, inverse decays are still effective in trying to restore equilibrium. Eventually, they too freeze out and any X's and $\bar{X}$'s still present, decay freely to produce a net baryon number. If we continue to lower the mass as in curve 3, $M_X = 3\times10^{16}\alpha$, scatterings begin to play a role in driving things further towards equilibrium. Again, when they freeze out the remaining $X,\bar{X}$ pairs decay leaving an asymmetry. If scatterings become dominant, however, the resulting asymmetry in the standard model will become exponentially small with decreasing M_X as shown in the dashed curve. In Fig. 8, we have plotted the final asymmetry which is produced as a function of $K = 3\times10^{17}\alpha/M_X$ where K is defined by

$$K \equiv \Gamma_D/H\big|_{T=M_X}. \tag{5.25}$$

Depending on whether or not X is gauge or Higgs boson, the resulting final asymmetry can be approximated by

$$n_B/s \approx 2\times10^{-3}(\Delta B)/[1 + (3K)^{1.2}] \tag{5.26}$$

for Higgs bosons, and

$$n_B/s \approx 8\times10^{-3}(\Delta B)/[1 + (16K)^{1.3}] \tag{5.27}$$

for gauge bosons.

The above approximations assume that only one type of boson carries baryon-number violating interactions. In general, there may be several in which case the baryon asymmetry generated by a heavy boson may be wiped out (totally or partially) by lighter ones. The degree of damping can be approximated as[36)]

$$(n_B/s)_{\text{initial}} \exp[-O(K)] \tag{5.28}$$

for Higgs bosons and

$$(n_B/s)_{\text{initial}} \exp[-5.5K] \tag{5.29}$$

for gauge bosons. The resulting asymmetry is then computed (using the lightest boson which violates baryon number) by damping any prior asymmetries by Eq. (5.28) or Eq. (5.29) and adding to that the asymmetry generated by Eq. (5.26) or Eq. (5.27).

As supersymmetric theories are becoming ever more important (and popular) it will be worthwhile looking at what happens to the baryon asymmetry in a supersymmetric GUT. The largest effect due to the increased number of particles is that scatterings may become more important and make it harder to go out of equilibrium. In addition, the GUT coupling may be larger and the existence of dimension 5 operators all increase the effects of scatterings.[37)]

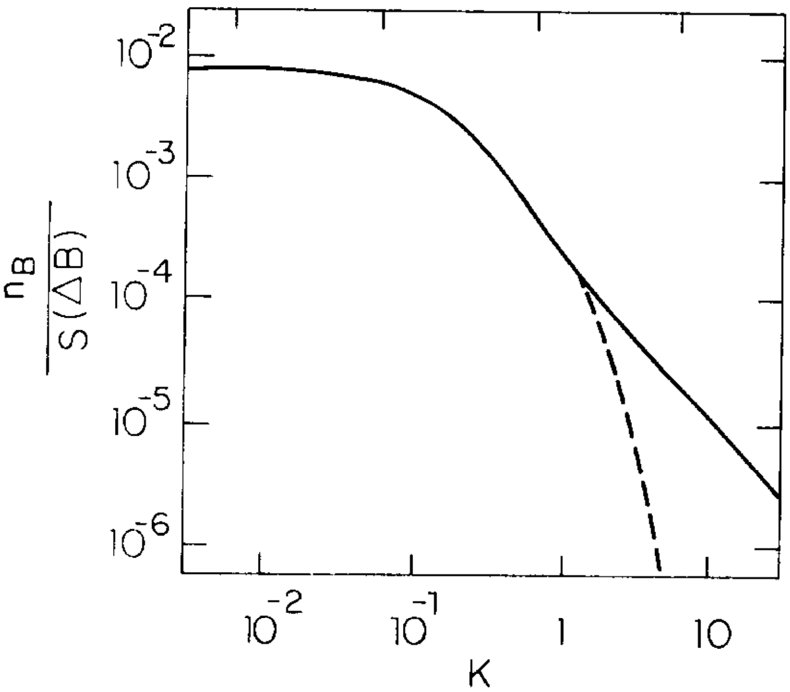

Figure 8. The final baryon asymmetry as a function of $K = 3\times10^{17}\ \alpha/M_X$ in units of (ΔB). The dashed curve assumes effective scatterings.

It is also very common in supersymmetric theories to have light Higgs bosons ($M_H \sim 10^{10}$ GeV) which violate baryon number. For a proper choice of couplings to the different generations, proton decay does not rule out such light Higgses.[38] Even in a non supersymmetric GUT, such a light Higgs boson would yield an asymmetry $n_B/s \sim 4\times10^{-8}$ (ΔB) which is probably too low to explain the baryon-to-photon ratio η. In a supersymmetric theory this number would be many orders of magnitude smaller. As we will show, however, this does not pose a serious constraint on supersymmetric theories.

Before addressing the baryon asymmetry directly, it will be useful to first address some general cosmological problems which might be encountered in a supersymmetric GUT.[39] To begin with, let us consider global supersymmetry. At zero temperature, supersymmetric GUTs may have several degenerate minima [e.g., SU(5), SU(3)×SU(2)× U(1), SU(4)×U(1), etc.] all with zero vacuum energy density. This degeneracy will however be broken at finite temperature, and we must now ask which minima is preferred.

The standard picture for a phase transition in the early Universe, e.g., SU(5) → SU(3)×SU(2)×U(1) would have a single minima at temperatures greater than some critical temperature T_c. Below T_c other minima develop with lower vacuum energy density. If there is a barrier between the minima, the transition will in general be of first order. If there is no barrier and the symmetric high temperature minimum disappears at T_c, the transition is second order. There is, however, no degeneracy. Figures 9a and 9b show schematically the possible behavior of the scalar potential as a function of a scalar field Σ.

In Figure 10, we see schematically the behavior of the scalar potential in a supersymmetric GUT, the important finite temperature corrections are just

$$V_T(\Sigma) = V(\Sigma) + C_1\Sigma^2T^2 - C_2T^4, \tag{5.30}$$

where $V(\Sigma)$ is the tree-level potential, C_1 is derived from V, and C_2 [= $(\pi^2/90)$ N(T)] simply counts degrees of freedom. The Σ^2T^2 is an expansion only relevant near the origin [for Σ>>T it is cut off by $\exp(-\Sigma/T)$] and is not all that important here in distinguishing the minima. The value of C_2, however, will be different in the different minima depending on how many light particle states there are in each vacuum. This term will break the degeneracy. Unfortunately, it does so in the wrong way. In SU(5) C_2 is larger than C_2 in either SU(3)×SU(2)×U(1) or SU(4)×U(1). Thus, it appears that at any temperature the SU(5) symmetric state would always be the lowest minimum, and a phase transition would be impossible.

The above picture is somewhat relieved when one realizes that there is some scale Λ_5 such that SU(5) becomes strong, i.e., $\alpha_{GUT} \sim 1$. Thus it is incorrect to think of SU(5) containing large numbers of (nearly) massless particles at $T < \Lambda_5$. Instead, as T drops to Λ_5, C_2 in SU(5) could become smaller than C_2 in the other minima. In this case, the vacuum energy density in the symmetric phase would be greater than those in the broken phases and hence a phase transition would become possible. If we now look at a simple example of a supersymmetric model, consider the most general (renormalizable) superpotential for the adjoint Σ,

$$f(\Sigma) = \frac{1}{2}m\mathrm{TR}(\Sigma^2) + \frac{1}{3}\lambda\ \mathrm{TR}(\Sigma^3). \tag{5.31}$$

In global supersymmetry the scalar potential for Eq. (5.31) (neglecting D terms) will be

$$V(\Sigma) = \mathrm{TR}\left|m\Sigma + \lambda\Sigma^2 - \frac{1}{5}\lambda\ \mathrm{TR}(\Sigma^2)\right|^2. \tag{5.32}$$

Now although the vacuum energy density of the broken phase is lower than in the symmetric phase, there is a barrier[40] of height $O(m^4/\lambda^2)$. The phase transition will complete itself only if the probability of tunneling per unit volume p, becomes greater than H^4. If we choose $m = M_X \sim 10^{16}$ GeV and $\lambda \sim 1$, the tunneling probably is given by[41]

$$p \approx M_X^4\ e^{-B}, \tag{5.33}$$

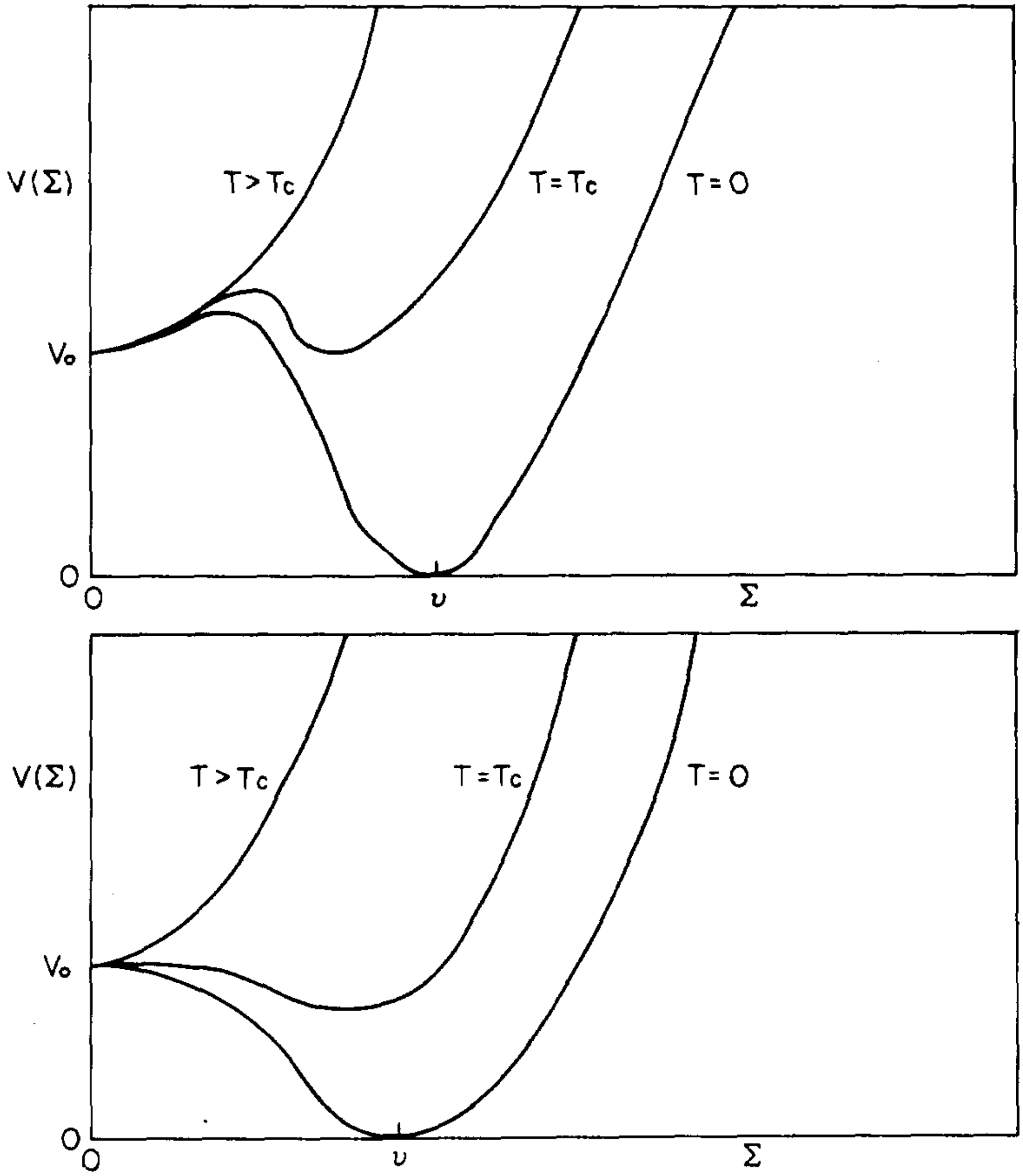

Figure 9. a) Schematic view of the scalar potential for a first-order phase transition. b) Schematic view of the scalar potential for a second-order phase transition.

where the action B is

$$B \approx 0.04(M_X/T)^9, \qquad (5.34)$$

and at $T \sim \Lambda_5$ (when the transition first becomes possible) $p \lesssim 10^{-10^{52}}$ and certainly in this case no transition occurs. One possible solution[42] (although not very attractive is to set $m \sim m_W$ and $\lambda \sim m_W/m_X \sim 10^{-14}$. In this case the barrier would be lower than the non-perturbative effects due to $\Lambda_5 \sim 10^9$–10^{10} GeV. This type of problem (small couplings) seems to have a possible solution in local supersymmetry through the use of non-renormalizable interactions. We will not here go into this solution but only refer the reader to some recent proposals.[43]

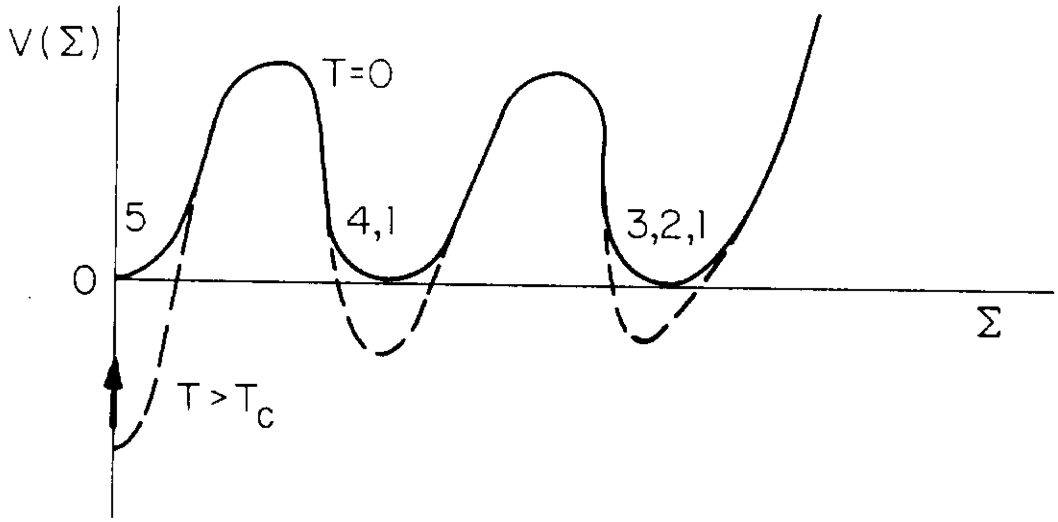

Fig. 10. Schematic view of the scalar potential in a supersymmetric theory. The dashed curves correspond to the effects of finite temperature. The arrow indicates that the SU(5) minimum is expected to move up as T approaches Λ_5.

Let us now return to the problem baryosynthesis in supersymmetric models. Consider the addition[42,43] to the superpotential Eq. (5.31)

$$f = a_1 \bar{H}\Sigma H + a_2 Y H \bar{H} + a_3 \mu Y^2 + a_4 Y^3, \qquad (5.35)$$

where H and $\bar{H}$ are in 5 and $\bar{5}$ representations, and Y is an SU(5) singlet. In order to keep the mass of the triplet in H and $\bar{H}$ light

$[m_H \sim O(10^{10})$ GeV] we will need $a_1 \sim a_2 \sim 10^{-6}$, and we can allow $a_3 \sim a_4 \sim 1$. In this case, Y will also have a mass of $O(10^{10})$ GeV if $\mu \sim 10^{10}$ GeV. It is the H and $\bar{H}$ which will eventually be responsible for producing the baryon asymmetry.

Regardless of the exact details of SU(5) (perturbative or non-perturbative) the singlets Y will be unaffected. Their decay will be governed by the term linear in Y

$$4a_2a_3\mu YH\bar{H}. \tag{5.36}$$

So long as $m_Y > 2m_H$, Eq. (5.36) leads to a decay $Y \to H,\bar{H}$ with a rate

$$\Gamma_Y \sim a_2^2a_3^2\mu. \tag{5.37}$$

Once again, the decays will only begin to occur when $\Gamma_Y \gtrsim H$ or

$$T_D \lesssim a_2a_3(\mu M_p)^{1/2} \sim 10^8 \text{ GeV}. \tag{5.38}$$

At T_D, however, $n_Y \sim n_\gamma$ and hence $n_H \sim n_{\bar{H}} \sim n_\gamma$ as well. Thus, the H, $\bar{H}$'s are out of equilibrium [as at $T \sim 10^8$, GeV $n_H/n_\gamma \sim \exp(-100)$ in equilibrium]. Their subsequent decays will then produce an asymmetry

$$n_B/s \sim (\Delta B)\,\frac{T_D}{m_H} \sim 10^{-2}(\Delta B) \tag{5.39}$$

or very close to the original out-of-equilibrium decay estimate. In Eq. (5.39) ΔB is the net baryon number produced by an H, H^* ($\bar{H},\bar{H}^*$) decay and the factor T_D/m_H is due to the entropy produced by the decay. Thus we see that the original worries about baryon generation in supersymmetric models were unfounded.

Before closing this section, we would like to look at two problems which result from combining GUTs and cosmology. The first problem concerns the abundance of magnetic monopoles.[44] GUTs predict the existence if magnetic monopoles. The monopoles will be produced[45] whenever any simple group [such as SU(5)] is broken down to a gauge group which contains a U(1) factor [such as SU(3)×SU(2)×U(1)]. The mass of such a monopole would be

$$M_m \sim M_G/\alpha_G \sim 10^{16} \text{ GeV}. \tag{5.40}$$

The basic reason monopoles are produced is that in the breaking of SU(5) the adjoint can not align itself over all space. On scales larger than the horizon, for example, there is no reason to expect the direction of the Higgs field to be aligned. Because of this randomness, topological knots are expected to occur and these are the magnetic monopoles. We can then estimate that the minimum number of monopoles produced would be one per horizon volume or causally connected region at the time of the SU(5) phase transition t_c

$$N_m \sim (2t_c)^{-3}. \qquad (5.41)$$

The time t_c is related to T_c through (2.60)

$$t_c \approx 0.3M_p\, N(T_c)^{-1/2}\, T_c^{-2} \qquad (5.42)$$

so that the monopole-to-photon ratio is

$$n_m/n_\gamma \sim (10T_c/M_p)^3. \qquad (5.43)$$

Just as in the case of neutrinos, we can look at the limits on the abundance of magnetic monopoles due to the overall density of the Universe. The mass density of monopoles will be

$$\rho_m = M_m n_m \qquad (5.44)$$

and the fraction of critical density in monopoles can be expressed as

$$\Omega_m h_0^2 = 9.5\times10^4\, M_m(\mathrm{GeV})\, n_m\, (\mathrm{cm}^{-3}). \qquad (5.45)$$

Thus for $M_m \sim 10^{16}$ GeV and $\Omega h_0^2 < 1$ we have that

$$n_m/n_\gamma < O(10^{-25}). \qquad (5.46)$$

The predicted density, however, from (5.43) for $T_c \sim M_G \sim 10^{15}$ GeV yields

$$(n_m/n_\gamma) \sim 10^{-9}. \qquad (5.47)$$

Hence, we see that standard GUTs and cosmology have a monopole problem. There are basically two solutions to this problem. If instead of $T_c \sim M_G$, the SU(5) phase transition were supercooled[46] and $T_c < 10^{10}$ GeV the number of monopoles might be acceptable. Recall the types of supersymmetric GUTs described above have exactly this property, $T_c \sim \Lambda_5 \sim 10^9 - 10^{10}$ GeV. Thus they might not overproduce monopoles. The second solution involves the inflationary Universe[4] scenario and will be discussed in section 7.

The final problem we would like to discuss in relation to particle theory is only a problem in locally supersymmetric theories. This problem involves the overproduction of entropy through gravitino decay.[47] At very early times we expect that gravitinos were as abundant as photons. Gravitinos, however, only couple gravitationally and hence are decoupled from the thermal background until very late times when they decay. Their decay rate will be the gravitational rate

$$\Gamma_{3/2} \sim m_{3/2}^3/M_p^2, \qquad (5.48)$$

where $m_{3/2} \sim 100$ GeV is the gravitino mass. Because of their early decoupling, gravitinos will be at a lower temperature then the photons when they decay so that their total mass density will be

$$\rho_{3/2} \sim m_{3/2}\, n_\gamma/N(M_p). \tag{5.49}$$

When gravitinos decay, the Universe will be matter dominated (this depends of course on the gravitino mass) and the expansion rate is given by

$$H \sim m_{3/2}^{1/2}\, T^{3/2}/M_p\, N(M_p)^{1/2} \tag{5.50}$$

so gravitinos decay when $\Gamma_{3/2} \sim H$ or

$$T_D \sim m_{3/2}^{5/3}\, M_p^{-2/3}\, N(M_p)^{1/3}. \tag{5.51}$$

After the decay products of the gravitinos thermalize, they will have reheated the Universe to a temperature

$$T_A \sim (m_{3/2}\, T_D^3/N\,(M_p))^{1/4} \sim m_{3/2}^{3/2}/M_p^{1/2}. \tag{5.52}$$

In addition, the entropy increase of the Universe will be

$$(T_A/T_D)^3 \sim (M_p/m_{3/2})^{1/2}/N(M_p). \tag{5.53}$$

If we put some numbers into these equations, we find that for $m_{3/2} \sim 100$ GeV, gravitinos decay at $T_D \sim 10$ eV, i.e., after nucleosynthesis. The entropy increase is $O(10^6)$ which presents problems for both big bang nucleosynthesis as well as big bang baryosynthesis. In the next two sections, we will look at other cosmological problems and their resolution [as well as a solution to the gravitino problem[48]] in the inflationary Universe scenario.

Section 6. Cosmological Problems

In the previous five sections, we have outlined the standard big bang cosmological model from the GUT epoch at $t \sim 10^{-35}$ sec to the time of recombination at about $t \sim 10^5$ yrs. The model in its simplicity is amazingly successful. Independent of any particle physics model, it does, however, have some problems of its own which we would like to address in this section. They are 1) the horizon problem; 2) the small scale inhomogeneity; 3) the curvature or flatness problem; 4) the rotation problem; and 5) the cosmological constant. All but the last of these problems may be resolved by inflation[4)] as we will discuss in the next section.

The horizon volume or causally connected volume today, is just related to the age of the Universe $V_0 \propto t_0^3$. The microwave background radiation with temperature $T_0 \sim 3°K$ has been decoupled from itself since the epoch of recombination at $T_d \sim 10^4 °K$. The horizon volume at that time was $V_d \propto t_d^3$. Now the present horizon volume scaled back to the period of decoupling will be $V_0' = V_0(T_0/T_d)^3$ and the ratio of this volume to the horizon volume at decoupling is

$$\begin{aligned} V_0'/V_d &\sim (V_0/V_d)\ (T_0/T_d)^3 \\ &\sim (t_0/t_d)^3\ (T_0/T_d)^3 \sim 10^5, \end{aligned} \tag{6.1}$$

where we have used $t_d \sim 3 \times 10^{12}$ sec and $t_0 \sim 5 \times 10^{17}$ sec. The ratio (6.1) corresponds to the number of regions that were causally disconnected at recombination which grew into our present visible Universe.

The microwave background radiation appears to be highly isotropic. In fact, the limits on the anisotropy put

$$\Delta T/T \lesssim 10^{-4}. \tag{6.2}$$

This means that on large scales, the Universe must be very isotropic and homogeneous, (any inhomogeneities would also produce fluctuations in the microwave background). The horizon problem, therefore, is the lack of an explanation as to why 10^5 causally disconnected regions at t_d all had the same temperature to within one part in 10^4!

Although it appears that the Universe is extremely isotropic and homogeneous (in fact the standard model assumes complete isotropy and homogeneity) it is very inhomogeneous on small scales. In other words, there are planets, stars, galaxies, clusters, etc. On small scales, therefore, there are large density perturbations. The problem is to understand how such density perturbations were formed (remember that on large scales we must have $\delta\rho/\rho \sim \delta T/T \sim 10^{-4}$) and how on small scales these perturbations grew to $\delta\rho/\rho \sim O(1)$.

The curvature problem (also known as the flatness or oldness problem) stems from the fact that although the Universe is very old, we still do not know whether it is open or closed. Recall that because of the scale dependence in the Friedmann Eq. (1.25), the expansion rate will always be matter- or radiation-dominated early ($\rho \sim R^{-3}$, R^{-4}) and curvature-dominated later (k/R^2). Neglecting the cosmological constant, the curvature term was expressed in terms of Ω and H_0^2 in Eq. (2.27)

$$k/R^2 = (\Omega - 1)\ H_0^2. \tag{6.3}$$

If we now use the limits $\Omega < 4$ and $H_0 < 100$ km s^{-1} $M_{pc}{}^{-1}$ (the limit $h_0 < 1/2$ necessary when $\Omega > 1$ to be consistent with the age of the Universe will make no difference here) we can form a dimensionless constant

$$\begin{aligned} \hat{k} = k/R^2T^2 = (\Omega - 1)\ H_0^2/T^2 \lesssim 3H_0^2/T_0^2 \\ < 2\times10^{-58}, \end{aligned} \tag{6.4}$$

where we have used $T_0 > 2.7°$K. In an adiabatically expanding Universe, $\hat{k}$ is absolutely constant ($R \sim T^{-1}$) and thus the limit (6.4) represents an initial condition which must be imposed so that the Universe will have lived this long looking still so flat.

A more natural initial condition might have been $\hat{k} \sim O(1)$. In this case the Universe would have become curvature dominated at $T \sim 10^{-1}\ M_p$. For $k = +1$, this would signify the onset of recollapse. Even for $\hat{k}$ as small as $O(10^{-40})$ the Universe would have become curvature dominated when $T \sim 10$ MeV or when the age of the Universe was only $O(10^{-2})$ sec. Thus not only is (6.4) a very tight constraint, it must also be strictly obeyed. Of course, it is also possible that $k \equiv 0$ and the Universe is actually spatially flat.

Similar to the curvature problem is the rotation problem,[49] i.e., why isn't the Universe rotating? By rotation we mean an anisotropy in the Universe to which is associated a preferred direction and angular momentum. The strongest limits on the rotation are due to its possible effects on the microwave background radiation and one finds that[50]

$$\omega \lesssim 10^{-21}\ s^{-1}, \tag{6.5}$$

where ω is the associated angular velocity of the Universe. ω will scale with the expansion and the scaling depends on the equation of state

$$\omega \sim R^{3\gamma-5}, \tag{6.6}$$

where γ was defined in (2.13). Thus for a matter-dominated Universe ($\gamma = 1$) $\omega \sim R^{-2}$ and for a radiation-dominated Universe ($\gamma = 4/3$), $\omega \sim R^{-1}$.

We can now scale back ω to the planck time and see what kind of initial condition is imposed by (6.5). The limit (6.5) can be scaled back to the epoch of recombination (when we expect the change over from radiation) to matter domination by

$$\omega_d < 10^{-21}\ s^{-1}\ (\frac{t_0}{t_d})^{4/3} \sim 0(10^{-14})\ s^{-1} \qquad (6.7)$$

at recombination. At earlier times, w scales back to the planck time as

$$\omega_p < 0(10^{-14})\ s^{-1}\ (\frac{t_d}{t_p})^{1/2} \sim 0(10^{14})\ s^{-1}. \qquad (6.8)$$

We can again form a dimensionless constant which at the planck time was

$$\hat{\omega} = \frac{\omega_p}{T_p} \lesssim 2\times10^{-29}. \qquad (6.9)$$

Because a rotation term would enter into the field equations as ω^2, the limit (6.9) is amazingly similar to (6.4). Once again we would expect that initially $\hat{\omega} \sim 0(1)$.

The final problem we would like to discuss is that of the cosmological constant. As we said in section 2, the Universal expansion is not dominated by the cosmological constant. This can be put in the form of a limit on Λ and when put in dimensionless form reads

$$\Lambda/M_p^2 \lesssim 10^{-121}. \qquad (6.10)$$

As we will see in the following section, the cosmological constant might not really have been constant throughout the entire evolution of the Universe. If we associate vacuum energy densities of the various phases the Universe passed through as an effective cosmological constant, we would expect a wide variety of values. The GUT epoch would have $\hat{\Lambda} \sim (M_x/M_p)^4 \sim 10^{-12}$. While at SU(2) breaking, we would get a contribution $\hat{\Lambda} \sim (M_w/M_p)^4 \sim 10^{-68}$. How all of the phase transitions which are accompanied by a change in vacuum energy density all conspired to give such a low value for Λ today is not at all understood.

Section 7. The Inflationary Universe

In all of the problems that were discussed in the previous section (except for the cosmological constant) it was assumed that the Universe has always been expanding adiabatically. During a phase transition, however, this is not necessarily the case. If we go back to Fig. 9a, and we suppose that because of the barrier separating the two minimum, the phase transition was a supercooled first-order transition. If in addition, the transition takes place at T_c such that* $T_c^4 < V_0$, the energy stored in the form of vacuum energy will be released. If released fast enough, it will produce radiation at a temperature $T_R^4 \sim V_0$. In this reheating process entropy has been created and

$$(RT)_f \sim (T_R/T_c)\ (RT)_i \tag{7.1}$$

provided that T_c is not too low. Thus we see that during a phase transition the relation $RT \sim$ constant need not hold true and thus our "dimensionless constants" $\hat{\omega}$ and $\hat{k}$ may actually not have been constant.

The inflationary Universe scenario,[4] is based on just such a situation. If during some phase transition, the value of RT changed by a factor of $O(10^{29})$, the first, third, and fourth of our cosmological problems would be solved. The isotropy would in a sense be generated by the immense expansion; one small causal region could get blown up and hence our entire visible Universe would have been at one time in thermal contact. In addition, the parameters $\hat{k}$ and $\hat{\omega}$ could have started out $O(1)$ and have been driven small by the expansion.

If, in an extreme case, a barrier as in Fig. 9a caused a lot of supercooling such that $T_c^4 << V_0$, the dynamics of the expansion would have greatly changed. In the example of Fig. 9a the energy density of the symmetric vacuum, V_0, acts as a cosmological constant with

$$\Lambda = 8\pi\ V_0/M_P^2. \tag{7.2}$$

If the Universe is trapped inside the false vacuum with $\Sigma = 0$, eventually the energy density due, to say, radiation will fall below the vacuum energy density, $\rho << V_0$. When this happens, the expansion rate will be dominated by the constant V_0 and we will get the De Sitter-type expansion (2.16), (2.17)

$$R \sim \exp[Ht], \tag{7.3}$$

where

*We now refer to T_c as the temperature at which the transition actually takes place rather than when it is at first physically possible.

$$H^2 = \Lambda/3 = 8\pi\ V_0/3M_p^2. \tag{7.4}$$

The cosmological problems could be solved if

$$H\tau \gtrsim 65, \tag{7.5}$$

where τ is the duration of the phase transition and the vacuum energy density was converted to radiation so that the reheated temperature is found by

$$\frac{\pi^2}{30} N(T_R)\ T_R^4 = V_0. \tag{7.6}$$

If such a barrier persists down to low temperatures, the phase transition must proceed via the formation of bubbles of the broken phase. The bubble formation rate per unit volume is given by[41] (5.33)

$$p \sim Ae^{-B}, \tag{7.7}$$

where $A^{1/4}$ is generally taken to be the overall mass scale in the problem ($A \sim T^4$ or $A \sim M^4$) and B is tunneling action. The transition will take place in such a way so as to minimize the action. There are in general several possible forms for B of which the form yielding the lowest value will be realized. One possibility[51] is the Einstein action

$$B = -4\pi \int d^4x \sqrt{-g}\ (R_c - 2\Lambda), \tag{7.8}$$

where $g = \det g_{\mu\nu}$ and $R_c = R^\mu_\mu$ is the curvature scalar. Another possible action would be the Coleman-De Luccia[52] bounce action for the formation of a bubble including gravitational effects. The phase transition will be completed if bubbles form fast enough or $p \gtrsim H^4$. More specifically, the fraction of space in which the transition has not occurred can be expressed as[46]

$$f(t) = \exp[-\textstyle\int_0^t dt'\ p(t)\ R^3(t)\ V\ (t,t')], \tag{7.9a}$$

where

$$V(t,t') = \frac{4\pi}{3}[\textstyle\int_{t'}^{t} dt''/R(t'')]^3. \tag{7.9b}$$

The transition is finished at time τ when $f(\tau) \approx 0$.

The scenario just described is the original idea of Guth[4] for cosmological inflation. In this scenario, the Universe would undergo a phase transition, say SU(5) → SU(3)×SU(2)×U(1) in which the potential resembled that in Fig. 9a. The Universe would then get hung up in the SU(5) phase down to a very low temperature (and may, therefore, solve the monopole problem). After completion of the phase transition, the Universe would reheat to

$$T_R \sim M_X/[N(T_R)]^{1/4}. \qquad (7.10)$$

Baryon generation would then follow so long as T_R was not too low.

It is now known that there is a problem with Guth's original idea for inflation.[53] It turns out that the requirement that the Universe supercool for a long enough time ($H\tau > 65$) is not compatible with $f(\tau) \to 0$, i.e., the phase transition does not finish. In order to have a long infationary time scale, a large barrier was necessary so as to be sure that the action for tunneling was also large. It is necessary in this scheme that the initial probability for tunneling be very small. The problem is that under these conditions the tunneling probability never catches up with the expansion rate, which is exponential at this point. As a whole, the Universe remains in the De Sitter state trapped in the symmetric SU(5) vacuum with only a few isolated bubbles containing the true SU(3)×SU(2)×U(1) vacuum. Not only is the resulting Universe very inhomogeneous, but each bubble remains empty as all of the energy is stored in the bubble walls and is only released through collisions which in this case do not occur.

The solution to this problem is called the new inflationary Universe[54] and its basic and simple idea is this: tunnel first and inflate later. To realize this type of inflation, one must have a long flat scalar potential. If one can argue (e.g., by thermal effects) that at early times or high tempertures the Universe was in the symmetric phase $\Sigma = 0$ and then at some lower temperature $T << T_c$ a bubble is formed. The supercooling may be due to either a barrier as in the previous case or a suppression of thermal fluctations so that the field Σ rests near the origin. In the case of a barrier, once a bubble is formed, if the potential is very long and flat at values of Σ past the barrier, the potential energy density (approximately constant) will again act like a cosmological constant. If a single bubble were to expand by 29 orders of magnitude, the phase transition need not be completed as in the previous case. The entire visible Universe would be contained within one bubble. The bubble would be filled in this case not by bubble collisions, but by dissipation of the kinetic energy of the scalar field as it finally reaches its global minimum. A generic example of such a potential is shown in Fig. 11.

Popular examples of flat potentials considered for inflation have been the Coleman-Weinberg[55] potentials which are derived by taking first-order radiative corrections to the tree potential. If scalar self couplings are small enough, the tree potential can be neglected and we can concentrate on the corrections. In general, we can write the C-W potential as

$$V(\phi) = A\phi^4 \left(\ln \frac{\phi^2}{v^2} - 1/2\right) + D\phi^2 + 1/2\ A\ v^4, \qquad (7.11)$$

where ϕ is some scalar field [it may be the adjoint in the case of SU(5) or any other scalar field which is appropriate]. The ϕ^4 coupling A is given by

$$A = \frac{1}{64\ \pi^2\ v^4}\ [\Sigma_B g_B M_B^4 - \Sigma_F g_F M_F^4], \tag{7.12}$$

where $g_{B(F)}$ is the number of boson (fermion) helicity states of mass $M_{B(F)}$. The expression (7.12) takes into account all possible first-order corrections, the relative minus signifies that it is a fermion loop correction rather than one due to bosons. The effective mass^2 is given by

$$D = 1/2(M_0^2 + cT^2 + bR_c - 3\ \lambda < \phi^2>), \tag{7.13}$$

where M_0 is a possible bare mass term, cT^2 is a gauge group dependent finite temperature correction, bR_c is a possible coupling to the scalar curvature R_c and the final term is an effect of ϕ^2 fluctuations in curved space[56,57], $-\lambda/4$ is the ϕ^4 self coupling.

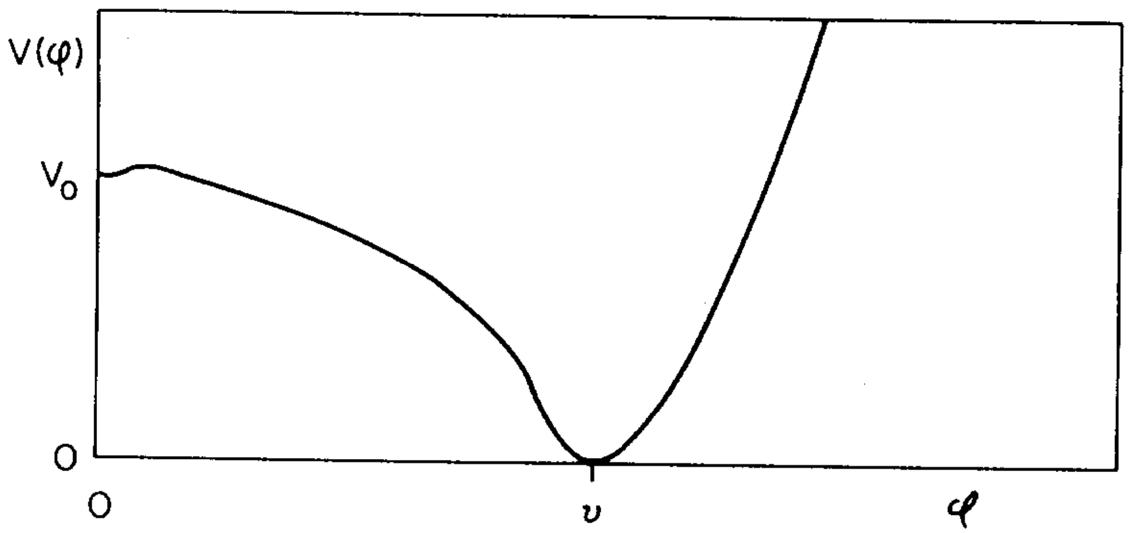

Fig. 11. Typical shape of a scalar potential needed for the new inflationary scenario.

In standard SU(5) the potential (7.11) is determined. The X and Y gauge bosons dominate the loop diagrams and we have $g_B = 36$, $M_X^2 = M_Y^2 = 25/8\ g^2\ v^2$, where g is the SU(5) gauge coupling, $g^2/4\pi = 1/41$, and v is the vacuum expectation value for the adjoint. In this case

$$A = \frac{5625}{1024\pi^2}\ g^4 \approx 5\times10^{-2}. \tag{7.14}$$

The constant $c \simeq (75/8)\ g^2$ is valid for $0 < \phi < T < v$. In order for inflation to occur during SU(5) Coleman-Weinberg breaking the tunneling action must be large.[58] If not for a barrier in this model, thermal fluctuations would drive the transition too early resulting in insufficient inflation.

To determine whether or not inflation actually occurs in this model, let us first look at the time scale for the field ϕ to go from its initial post barrier position to its global minimum at $\langle \phi \rangle = v$. The roll-over time scale is determined by the equation-of-motion for ϕ

$$\ddot{\phi} + (3H + \Gamma)\ \dot{\phi} + \partial V/\partial\phi = 0, \tag{7.15}$$

where Γ is the rate of interactions of the ϕ field, and controls particle creation. The Γ term is only relevant when $\Gamma \gtrsim H$. Initially Γ must be small for inflation to occur and we will neglect it for the time being. Initially $\ddot{\phi}$ will also be small so that the roll-over time scale can be derived from

$$3H\dot{\phi} + (\partial V/\partial\phi) = 0 \tag{7.16}$$

$$\tau^{-1} \approx \dot{\phi}/\phi \sim (\partial^2 V/\partial\phi^2)/3H. \tag{7.17}$$

For $\phi << v$ the roll-over time scale is

$$\tau \sim 3H/2D. \tag{7.18}$$

In the case of the Coleman-Weinberg potential (7.11) we have

$$H^2 = \frac{8\pi V_0}{3M_p^2} = \frac{4\pi A v^4}{3M_p^2} \tag{7.19}$$

and the condition $H\tau > 65$ translates into an upper limit on D

$$D < (4\pi/130)\ Av^4/M_p^2. \tag{7.20}$$

For $v \sim 10^{15}$ GeV we find that

$$D < O(10^{19})\mathrm{GeV}^2. \tag{7.21}$$

This is the first drawback on the new inflationary scenario with Coleman-Weinberg type SU(5) breaking. The limit (7.21) implies that each mass term in D must be fine-tuned down to $O(10^9)$ GeV. This is, however, technically unnatural as scalar mass will tend to get radiative corrections to their mass of $O(10^{15})$ GeV.

A fine-tuned mass term as in (7.21) will also make it very difficult for the Universe to remain in the symmetric state down to low temperatures. If we look, for example, at the action given by

(7.8), this can be rewritten in the form[51)]

$$B = (3M_p^4/8)\ [1/V_0 - 1/V_1]$$

$$\approx (3M_p^4/8)\ [(V_1 - V_0)/V_0^2],\ V_1 - V_0 << V_0, \tag{7.22}$$

where $V_0 = 1/2\ Av^4$ and V_1 is the value of the potential at the local maximum

$$V_1 \approx V_0 + D^2/4A\ \ln(4v^2A/D). \tag{7.23}$$

Thus the action is

$$B \approx 3M_p^4D^2/8A^3v^8\ln(4v^2A/D). \tag{7.24}$$

For SU(5) and D satisfying (7.21), we find $B \sim O(10^{-2}) << 1$ and hence does not at all prevent a transition from occurring. In order to make B large we would need $A < O(10^{-4})$ corresponding to $D < O(10^{16})\mathrm{GeV}^2$ and hence mass terms fine-tuned down to $O(10^8)$GeV. Unfortunately, A is not adjustable in SU(5) so that it is not clear that ϕ will be constrained near the origin long enough so that inflation will actually occur.[59)]

This model however appears to be a good one in the sense that so many things go wrong it gives one a list of things to watch for. In addition to those we just mentioned, it was also pointed out that during inflation scalar field fluctuations[57)] would drive the phase transition unless $\lambda < 5\times10^{-3}$. In the present case, however, the effective ϕ^4 coupling for $\phi \sim H$ is about two orders of magnitude larger.

The most serious blow to Coleman-Weinberg type inflation comes from the density perturbations which are produced during ther rollover.[60)] The isotropy of the microwave background radiation tells us that any perturbations produced on large scales must have $\delta\rho/\rho \lesssim O(10^{-4})$. Ideally, what one would want from inflation is what is known as the Harrison-Zeldovich[61)] spectrum of density fluctuations. They are also known as scale independent perturbations which are the type most desired for the purposes of galaxy formation. Their magnitude, however, must be $O(10^{-4})$. Any perturbations stronger than this would produce visible anisotropies in the microwave background radiation while weaker perturbations would not have had enough time to grow during the present period of matter domination (since decoupling).

The initial spectrum of perturbations can be classified by their magnitude on a given length scale

$$\left(\frac{\delta\rho}{\rho}\right)_{\ell}^{\text{initial}} \sim 1/\ell^{3n} \sim 1/M^n, \tag{7.25}$$

where M is the mass contained within the volume ℓ^3. Perturbations on scales larger than the horizon grow in magnitude. Inside the horizon they oscillate until further growth is possible when the Universe becomes matter dominated. Perturbations on larger scales have a longer time to grow and for $n = 2/3$ it turns out that as each scale ℓ enters the horizon, the magnitude of the perturbations are equal. This is what is meant by scale-independent perturbations. In addition, for $n > 2/3$, perturbations are too strong and tend to "close up" and form individual Friedmann "Universes," i.e., they will be described by an independent metric. For $n < 2/3$, the perturbations are too weak to form galaxies. Therefore, the $n = 2/3$ spectrum is the one preferred for galaxy formation.

As it turns out, phase transitions, such as the SU(5) transition described above, produce[62] very nearly the $n = 2/3$ spectrum which is desired. The perturbations are formed because the field ϕ does not roll down to its global minimum homogeneously. There will, in general, be a time spread over which certain regions roll down faster or slower than others. The density perturbations have been calculated in terms of this time spread[60]

$$\frac{\delta\rho}{\rho} = 2\sqrt{2}\ H\ \delta\tau, \tag{7.26}$$

where $\delta\rho/\rho$ is the magnitude of the perturbation as it enters the horizon. The time spread $\Delta\tau$ has been estimated to give

$$\Delta\tau \approx \delta\phi/\dot{\phi}, \tag{7.27}$$

where the scalar fluctuations are taken in a De Sitter space to be

$$\delta\phi \approx H/4\pi^{3/2} \tag{7.28}$$

and $\dot{\phi}$ is found from the homogeneous equation of motion (neglecting $\ddot{\phi}$)

$$3H\dot{\phi} = -(\partial V/\partial\phi) \tag{7.29}$$

at $t \approx -\ln\ (Hk^{-1})/H$ where k is the wave number of the perturbation.

If we now go back to the Coleman-Weinberg potential (7.11), we can compute the magnitude of density perturbations that one finds. Equation (7.29) becomes

$$3H\dot{\phi} \approx -8A(\ln\ H/v)\ \phi^3 \tag{7.30}$$

for $\phi \sim H$, where we have used the fact that $D << H^2$. For SU(5), the solution to (7.30) is

$$\dot{\phi} \approx \sqrt{3/8\lambda}\ H^2\ \ln^{-3/2}\ (Hk^{-1}), \tag{7.31}$$

where

$$\lambda \approx -8\ A\ \ln\ H/v. \tag{7.32}$$

From Eqs. (7.26), (7.27), (7.28), and (7.31) the final magnitude of the density perturbations becomes

$$\delta\rho/\rho \sim (4\lambda/3\pi^3)^{1/2}\ \ln^{3/2}(Hk^{-1}). \tag{7.33}$$

We see, therefore, that although it is not completely scale independent, the logarithmic variations are not enough to disturb its usefulness for galaxy formation. If we now take k on galactic or horizon scales today we find

$$\delta\rho/\rho \sim 50 \tag{7.34}$$

i.e., about 5 orders of magnitude too large. Had this number turned out to be $O(10^{-4})$, inflation would have solved the cosmological problem of small scale perturbations as well as the others.

Clearly the list of problems with the Coleman-Weinberg SU(5) inflationary model is long enough. Before moving to the brighter possibilities we note that it has also been shown[63)] that under reasonable circumstances, the above model does not even break to SU(3)×SU(2)×U(1), but rather SU(4)×U(1). Not only do we have a small lumpy Universe, but we're in the wrong vacuum as well.

In the remainder of these lectures, we will consider the effects of supersymmetry on cosmological inflation. Let us recall one of the most powerful tools that supersymmetry puts in our hand, namely, the non-renormalization theorems.[64)] In the previous contribution, we saw how these theorems led to the stability of gauge hierarchy, that is, if we set the mass scale for a scalar field at, say, $m_0 \sim 10^2$ GeV, a non-supersymmetric model would have corrections $\delta m^2 \sim (10^{15}\ \mathrm{GeV})^2$ so that

$$m^2 \sim m_0^2 + \delta m^2 \sim (10^{15}\ \mathrm{GeV})^2, \tag{7.35}$$

whereas in a supersymmetric model

$$\delta m^2 = 0. \tag{7.36}$$

Because we know that the Universe is not exactly supersymmetric (there are no charged scalars with mass 0.511 MeV) there will be some radiative corrections

$$\delta m^2 \sim \varepsilon\, M_s^2, \tag{7.37}$$

where M_s is the scale of supersymmetry breaking and ε is some coupling constant. In locally supersymmetric models

$$\varepsilon \sim m_{3/2}/M_P \sim 10^{-16}, \tag{7.38}$$

and hence the smallest corrections are typically

$$\delta m^2 \sim m_{3/2}^2 \tag{7.39}$$

where we have used the relation

$$m_{3/2} \sim M_S^2/M_P. \tag{7.40}$$

Thus for $m_{3/2} \lesssim M_W$, the stability of the gauge hierarchy is guaranteed.

Because of the cancellations in radiative corrections, one might expect that supersymmetry would have a big effect on our previous discussion involving Coleman-Weinberg potentials.[59,65] If we go back to (7.12), in an exactly supersymmetric model, for every M_B, g_B there is a $M_F = M_B$ and $g_F = g_B$ so that $A \equiv 0$, i.e., there is no Coleman-Weinberg potential. In broken supersymmetry, we might have some splitting between the bose and fermion states

$$M_B^2 - M_F^2 \sim \varepsilon M_S^2 \tag{7.41}$$

so that

$$\begin{aligned} A = \frac{g_{B(F)}}{64\pi^2 v^4} \lfloor M_B^4 - M_F^4 \rfloor &\approx \frac{g_{B(F)}}{32\pi^2 v^4} \lfloor M_B^2 - M_F^2 \rfloor M_B^2 \\ &\approx \frac{g_B}{32\pi^2 v^4} M_B^2 \, M_s^2 \, \varepsilon \end{aligned} \tag{7.42}$$

and in SU(5) if we take $M_B^2 = (25/8)\, g^2 v^2$, $g_B = 24$

$$A \approx (75/32\pi^2)\, g^2\, (M_S^2/v^2)\, \varepsilon. \tag{7.43}$$

The most serious constraint on A came from the magnitude of density fluctuations $A \lesssim O(10^{-14})$. This translates to

$$M_S^2\, \varepsilon < O(10^{17})\ \text{GeV}^2. \tag{7.44}$$

For $\varepsilon \sim 10^{-16}$ this requires $M_S \lesssim O(10^{16})$ GeV which is not at all a serious constraint.

The above exercise is of course not a model, but only gives one an idea that supersymmetry might be very important for inflation. For interesting models, we must have $M_S \sim 10^{10}$ GeV so that $m_{3/2} \sim M_S^2/M_P \sim 10^2$ GeV. In addition, if we want to consider GUTs or inflation around that epoch we can in fact neglect the effects of supersymmetry breaking because $M_S \ll M_X$ and thus work in the context of exact supersymmetry. Therefore, in the following, we will be able to confidently neglect radiative corrections and work entirely at the tree level.

The scalar potential in supersymmetric models, recall, is derived from a superpotential f by

$$V = \sum_i |\partial f/\partial \phi_i|^2 \tag{7.45}$$

for globally supersymmetric models[66] and

$$V = e^{\Sigma|\phi_i|^2/M^2}\left[\sum_i\left|\frac{\partial f}{\partial\phi_i} + \phi_i^* f/M^2\right|^2 - 3|f|^2\right] \tag{7.46}$$

in minimal N = 1 local supersymmetry.[67] In Eqs. (7.45) and (7.46) we are summing over all chiral supermultiplets and $M = M_p/\sqrt{8\pi} \approx 2.4\times10^{18}$ GeV. Near the origin these potentials can be expanded so that they can always be put in the form

$$V(\phi) = \delta + \gamma\phi^2 - \beta\phi^3 + \alpha\phi^4 + \ldots \tag{7.47}$$

where we have left open the possibility for including non-renormalizable term in V as these will frequently appear in the supergravity models given by (7.46). So long as we restrict ourselves to scales $<< M_p$ our theory will still be well defined.

To begin with, let us consider ϕ to be a chiral supermultiplet gauge singlet and that ϕ picks up a vacuum expectation value $\langle 0|\phi|0\rangle = \mu$. In the following we will no longer restrict ourselves to $\mu \sim M_X$ i.e., ϕ and inflation need not be related to GUTs. As we will see shortly, for $\mu >> M_X$, inflation is easier to achieve. The basic properties of $V(\phi)$ must include that at $\phi = \mu$, the potential have a minimum so that $(\partial V/\partial\phi)|_{\phi=\mu} = 0$ and $(\partial^2 V/\partial\phi^2)|_{\phi=\mu} > 0$ and we will want the vacuum energy density at the minimum to vanish so that we have no cosmological constant. This requires $V(\mu) = 0$ as well. At the origin on the other hand, we want a flat potential with positive energy density $\delta > 0$.

As in the case of the Coleman-Weinberg potential, we will imagine that initially the field ϕ is near the origin. We will not yet specify whether this is due to finite temperature effects or not. We will demand, however, that fluctuations do not drive the field away from the origin at $T > H$. The Hubble parameter, when $T \lesssim H$, i.e., when the Universe becomes dominated by the vacuum energy density δ is given by

$$H^2 = \frac{8\pi}{3}\,\delta/M_p^2 = (1/3)\,\delta/M^2. \tag{7.48}$$

If we scale the parameters as

$$\delta = \hat{\delta}\mu^4,\ \gamma = \hat{\gamma}\mu^2,\ \beta = \hat{\beta}\mu \tag{7.49}$$

we can show the constraints on the dimensionless parameter $\hat{\delta}$, $\hat{\gamma}$ and $\hat{\beta}$ in terms of μ.

Depending on the couplings, the potential may or may not have a barrier. For $\hat{\gamma} > 0$ and $\hat{\beta} > 0$ there will be a barrier with a maximum at $\phi_1 \sim (2\hat{\gamma}/3\hat{\beta})\,\mu$. Without a doubt, the strongest constraint* on the

*For constraints on the scalar potential for inflation see Ref. 68.

parameters comes from the density perturbations. The starting point of the calculation should be taken to $\phi_0 \sim H$ or in the case of a barrier, the larger of $\phi_0 \sim H$ or $\phi_0 \sim 2\phi_1$ assuming again that initially we are at $\phi \simeq 0$. If at ϕ_0, $\partial V/\partial\phi$ is dominated by the quartic term, the density fluctuations will be very similar to those in (7.33)

$$\delta\rho/\rho \simeq (16\alpha/3\pi^3)^{1/2} \ln^{3/2} (Hk^{-1}). \tag{7.50}$$

If, on the other hand, it is the cubic term dominating we then have[69]

$$\delta\rho/\rho \simeq (2\pi^3)^{-1/2} (\beta/H) \ln^2 (Hk^{-1}). \tag{7.51}$$

In this case (which is probably the most interesting physically) the constraint $\delta\rho/\rho \sim 10^{-4}$ becomes

$$\hat{\beta} \sim O(10^{-7}) (\mu/M) \hat{\delta}^{1/2}. \tag{7.52}$$

In addition, the long roll-over time scale gives a constraint on the value of $\hat{\gamma}$ in terms of $\hat{\delta}$ and μ

$$\hat{\gamma} < O(10^{-2}) (\mu/M)^2 \hat{\delta}. \tag{7.53}$$

We can now see the benefit of keeping ϕ independent of GUTs. If ϕ was taken to be the adjoint and we related $\mu \sim M_X$ we find that $\hat{\beta} \sim 10^{-10} \hat{\delta}^{1/2}$ and $\hat{\gamma} < 10^{-9} \hat{\delta}$. Although this type of fine tuning is allowable in supersymmetric theories, it again represents an unnatural set of parameters which must be imposed on the model. If instead we let $\mu \sim M$ we have $\hat{\beta} \sim 10^{-7} \hat{\delta}^{1/2}$ and $\hat{\gamma} < 10^{-2} \hat{\delta}$, which is a considerable improvement. This situation, i.e., where the scalar driving inflation picks up a vacuum expectation value $\phi \sim \mu >> M_X$ is called primordial inflation.[65]

At this point, it is worth noting two points in which inflation is facilitated by primordial supersymmetric inflation. If we go back to Fig. 11, there are at least two obvious ways in which we can make the scalar potential flatter. One way is to decrease the value of V (leaving v fixed) and the second is to increase the value of v (leaving V fixed) or, of course, both. As we saw in the exercise with the Coleman-Weinberg potential, supersymmetry can accomplish the first while the second is defined to be primordial inflation. As v approaches M_p, we must start to worry if first-order gravitational (FOG) effects do not come in and change the picture. One possibility, however, is that these effects will be incorporated if one works in the framework of supergravity. The overall hope in this case is that if all quantum gravitational effects are contained in extended N = 8 supergravity, perhaps the physics at, or below, the planck scale are correctly understood in an N = 1 supergravity. We will make this assumption to close our discussion on inflation.

In order to construct a model for inflation in N = 1 supergravity,[70] let us start with the most general superpotential for a single scalar field ϕ which we will now call the inflaton single scalar field ϕ which we will now call the inflaton

$$f_I = m^3\left(\sum_n \frac{\lambda n}{n+1}\left(\frac{\phi}{M}\right)^{n+1} + \lambda\right), \qquad (7.54)$$

where m is an as yet undetermined mass scale. The scalar potential is determined by (7.46) and can be put in the form (7.47). Note that for a certain range of parameters ($\lambda^2_1 > 4\,\lambda_0\,\lambda_2$) we can restrict our attention to the real axis. We can then make the identifications between the couplings λ_i and α, β,γ,δ

$$\alpha = \left(\frac{\lambda_0^2}{2} + \frac{9}{8}\ \lambda_1^2 + \lambda_2^2 + \frac{8}{3}\lambda_0\lambda_2 + \frac{9}{4}\lambda_1\lambda_3 + 2\,\lambda_0\lambda_4\right) m^6/M^6 \qquad (7.55a)$$

$$-\beta = 2(\lambda_0\lambda_3 + \lambda_1\lambda_2 + \lambda_0\lambda_1/2)m^6/M^5 \qquad (7.55b)$$

$$\gamma = 2\lambda_0\lambda_2\ m^6/M^4 \qquad (7.55c)$$

$$\delta = (\lambda_0^2 - 3\lambda_1^2/4)\ M^6/M^2, \qquad (7.55d)$$

where we have used $\lambda_1 = 2\lambda$ in order to cancel the linear term at the origin.

In the spirit of primordial inflation, we will take $\mu = M$. We will then require that supersymmetry remain unbroken at $\phi = \mu$. One reason as we said earlier is that radiative corrections can be neglected in this case. More importantly, however, is that if supersymmetry were broken at $\phi = \mu$ by f (7.54), the gravitino would pick up a mass $m_{3/2} \sim m^3/M^2$. Although we have not said what the value of m is, as we will see shortly $m \sim 10^{-2}M$. To preserve the gauge hierarchy, we would need $m \lesssim 10^{-5}M$, this however would never reheat the Universe so as to produce a baryon asymmetry or large enough density perturbations. Thus demanding exact supersymmetry at $\phi = \mu$ implies that

$$f_\phi(\mu) = \left\lfloor\frac{\partial f}{\partial\phi} + \phi^* f/M^2\right\rfloor\Big|_{\phi=\mu} = 0. \qquad (7.56)$$

In addition to preserving supersymmetry we must also cancel the cosmological constant from Eqs. (7.56) and (7.46) this requires

$$f(\mu) = 0. \qquad (7.57)$$

These two conditions reduce to two constraints on the superpotential f (7.54)

$$\sum_n \frac{\lambda n}{n+1} + \lambda = 0 \qquad (7.58a)$$

$$\sum_n \lambda n = 0. \qquad (7.58b)$$

The final constraint on f comes from the fact that $\phi = \mu$ must also be a minimum. It is not difficult to show that any point which preserves supersymmetry and cancels the cosmological constant is a minimum so long as

$$\partial^2 f/\partial\varphi^2 \neq 0 \tag{7.59}$$

a trivial constraint to satisfy.

Let us now look at the simplest form of f which satisfies these constraints [we will neglect any temperature corrections on this potential[71]]

$$f = m^3(\lambda + \lambda_0\phi + \lambda_1/2\ \phi^2) \tag{7.60}$$

and the constraints yield

$$\lambda = \lambda_1/2 = -\lambda_0/2. \tag{7.61}$$

If we choose $\lambda = 1$ then $\lambda_0 = 2$ and $\lambda_1 = 2$. The mass scale m is fixed from the magnitude of the density perturbations (7.52)

$$\begin{aligned} \hat{\beta} &\simeq -(m/M)^6\ \lambda_0\lambda_1 \sim 0(10^{-7})\hat{\delta}^{1/2} \\ &\simeq 0(10^{-7})\ (m/M)^3\ (\lambda_0{}^2 - 3\lambda^2)^{1/2} \end{aligned} \tag{7.62}$$

or

$$m^3 \sim 0(10^{-8})M^3. \tag{7.63}$$

The condition for a long roll-over time scale is now automatically satisfied because $\hat{\gamma} = 0$ and at $\phi \sim H$, $\partial^2 V/\partial\phi^2 \sim -6\beta H$ and $3H^2/6\beta H \sim H/2\beta \sim 1/\lfloor 8\sqrt{3}\ (m/M)^3 \rfloor >> 65$.

Let us now close this discussion by looking at what happens to the Universe, in a more-or-less chronological sequence.[72] If there is a region in the Universe in which ϕ is near the origin, this region will inflate (regions in which ϕ is not near the origin do not and will hence be overshadowed by those which do inflate). As ϕ approaches its minimum at $\langle\phi\rangle = M$ it will begin to oscillate until the decay rate of ϕ

$$\Gamma_D \sim m_\phi^3/M^2 \tag{7.64}$$

becomes comparable to the expansion rate of the Universe which is governed by non-relativistic matter. When the ϕ's decay the Universe will reheat to a temperature

$$T_R \sim m_\phi^{3/2}/M^{1/2}. \tag{7.65}$$

In the case described above, the mass of the inflaton is given by

$$m_\phi = \frac{m^3}{M^2}\ \Sigma\ (n\lambda_n) = \frac{m^3\lambda_1}{M^2} \sim 2\times10^{-8}M \tag{7.66}$$

$$\sim 5\times10^{10}\ \text{GeV}$$

and hence the reheat temperature is $T_R \sim 2.8\times10^{-12}M \sim 7\times10^6$GeV. Although this may seem rather low for generating a baryon number, it turns out that an acceptable number will still arise. The inflaton will decay via gravitational interactions which are found in the cross terms of the scalar potential such as

$$m_\phi\, a_2\, \phi\, Y\, H\, \bar{H} \tag{7.67a}$$

$$m_\phi\, a_4\, \phi\, Y^3 \tag{7.67b}$$

etc., where we have included the superpotential (5.35). If we suppose that the branching ratio for ϕ into Y or H, $\bar{H}$ (or anything which leads to H, $\bar{H}$) is O(1/10) then the total baryon-to-entropy ratio produced will be

$$\begin{aligned} n_B/s &\sim O(1/10)\ (T_R/m_H)\Delta B \\ &\sim O(10^{-4})\Delta B \end{aligned} \tag{7.68}$$

which is still a sufficient baryon excess.

Looking back on our goals in this section, we have seen that of the five problems listed in section 6 we have been able to solve four of them by inflation. The horizon, curvature, and rotation problems were all solved similarly by the exponential increase in R. The small scale inhomogeneities are also supplied by inflation in that density perturbations $\delta\rho/\rho \sim O(10^{-4})$ can be generated. We are left however with the problem of the cosmological constant which is tuned by hand to be zero today.

In section 5, we saw two additional problems arise due to GUTs and supersymmetry. The monopole problem can also be solved by inflation. During the inflationary period, we expect SU(5) to break, the density of monopoles will then be exponentially suppressed and the problem disappears. The gravitino problem disappears in much the same way.[48] The initial abundance of gravitinos is exponentially suppressed; however, during the reheating they may be reproduced. Typically the abundance relative to photons is just T_R/M and is at an acceptable level if $T_R \lesssim O(10^{12})$GeV, clearly satisfied in the above example.

As we have seen through these lectures, cosmology has come a long way in the past twenty years. In the four series of lectures in this volume, we have tried to bring the reader more-or-less up-to-date in GUTs, supersymmetry, and cosmology. Our hope, of course, is that the reader can now digest it and expand upon it.

BIBLIOGRAPHY

S. W. Hawking and G. F. R. Ellis, **The Large Scale Structure of Space-Time** (Cambridge University Press, 1973).

Physical Cosmology, Proc. of the 1979 Les Houches Summer School ed. R. Balian, J. Audouze, and D. N. Schramm. (North-Holland Pub. Co., Amsterdam 1980) in particular see contributions by, J. Lequeux p. 3. G. A. Tammann, A. Sandage, and A. Yahil, p. 53. R. V. Wagoner p. 395.

Raychauduri, A. K. **Theoretical Cosmology** (Oxford University Press, 1979)

S. Weinberg, **Gravitation and Cosmology** (J. Wiley and Sons, New York 1972).

REFERENCES

1) G. Gamow, Phys. Rev. **70**, 572 (1946); R. A. Alpher, H. Bethe, and G. Gamow, Phys. Rev. **73**, 803 (1948).

2) A. A. Penzias and R. W. Wilson, Ap. J. **142**, 419, (1965); see also "The Universe at Large Redshifts," Proceedings of the Copenhagen Symposium, Physica Scripta **21**, 599 (1980).

3) A. D. Sakharov, Zh. Eksp. Teor. Fiz. Pisma Red. **5**, 32 (1967).

4) A. H. Guth, Phys. Rev. **D23**, 347 (1981).

5) See M. P. Ryan and L. C. Sheply, **Homogeneous Relativistic Cosmologies** (Princeton University Press, 1975).

6) S. W. Hawking and R. Penrose, Proc. Roy. Soc. Lond. **A314**, 529 (1970); see also S. W. Hawking and G. F. R. Ellis **The Large Scale Structure of Space-Time** (Cambridge University Press, 1973).

7) R. P. Kirshner, A. Oemler, and P. L. Schechter, Astron. J. **84**, 951 (1979).

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

9) J. Ellis and G. Steigman, Phys. Lett. **89B**, 186 (1980).

10) K. A. Olive, D. N. Schramm, and G. Steigman, Nucl. Phys. **B180** [FS2], 497 (1981).

11) R. V. Wagoner, W. A. Fowler, and F. Hoyle, Ap. J. **148**, 3 (1967); R. V. Wagoner, Ap. J. Supp. **18**, 247 (1969); R. V. Wagoner, Ap. J. **179**, 343 (1973); D. N. Schramm and R. V. Wagoner, Ann. Rev. Nucl. Part. Sci. **27**, 37 (1977).

12) G. Steigman, D. N. Schramm, and J. E. Gunn, Phys. Lett **B66**, 202 (1977); J. Yang, D. N. Schramm, G. Steigman, and R. T. Rood, Ap. J. **227**, 697 (1979).

13) K. A. Olive, D. N. Schramm, G. Steigman, M. S. Turner, and J. Yang, Ap. J. **246**, 547 (1981).

14) J. Yang, M. S. Turner, G. Steigman, D. N. Schramm, and K. A. Olive, Ap. J. (in press), 1984.

14a) Y. David and H. Reeves, in **Physical Cosmology** ed. R. Balian, J. Audouze, and D. N. Schramm (North-Holland Pub. Co. Amsterdam, 1980) 443.

15) For a recent compilation of observations see: the **Proceedings of the ESO Workshop on Primordial Helium**, Eds. P. A. Shaver, D. Kunth, and K. Kajar, Garching, Germany, 1983.

16) D. C. Black, Nature Phys. Sci. **234**, 148 (1971); D. C. Black, Geochim. Cosmochim. **36**, 347 (1972); J. Geiss and H. Reeves, Astron. Astrophys. **18**, 126 (1972).

17) J. T. Trauger et al., Ap. J. Lett. **184**, L137 (1973); V. Kunde et al., Ap. J. **263**, 443 (1982).

18) D. G. York and J. B. Rogerson, Jr., Ap. J. **203**, 378 (1976); A. Vidal-Madjar et al., Ap. J. **211**, 91 (1977); A. Vidal-Madjar et al., Astron. Astrophys. **120**, 58 (1983)

19) E. Anders, D. Heymann, and E. Mazor, Geochim. Cosmochim. **34**, 127 (1970).

20) I. Iben, Ap. J. **147**, 624 (1967); R. T. Rood, Ap. J. **177**, 681 (1972); I. Iben and J. W. Truran, Ap. J. **220**, 980 (1978); D. S. P. Dearborn, J. B. Blake, K. L. Hainebach, and D. N. Schramm, Ap. J. **223**, 552 (1978); W. Brunish and J. W. Truran (in preparation), 1983.

21) F. Spite and M. Spite, Aston. Astrophys. **115**, 357 (1982); M. Spite and F. Spite, Nature **297**, 483 (1982).

22) See Ref. 15.

23) K. Freese and D. N. Schramm, Nucl. Phys. B. (in press) 1984.

24) R. Cowsik and J. McClelland, Phys. Rev. Lett. **29**, 669 (1972); A. S. Szalay and G. Marx, Astron. Astrophys. **49**, 437 (1976).

25) P. Hut, Phys. Lett. **69B**, 85 (1977).

26) B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977).

27) J. E. Gunn, B. W. Lee, I. Lerche, D. N. Schramm, and G. Steigman, Ap. J. **223**, 101 (1978).

28) P. Hut and K. A. Olive, Phys. Lett. **87B**, 144 (1979).

29) M. S. Turner, Fermilab preprint Conf-83/109-AST, to be published in the Proceedings of the APS Division of Particles and Fields Meeting, Sept. 1983, Blaksburg, Virginia.

30) D. A. Dicus, E.W.Kolb, and V. L. Teplitz, Phys. Rev. Lett. **39**, 168 (1977); D. A. Dicus, E. W. Kolb, and V. L. Teplitz, Ap. J. **221**, 327 (1978); D. A. Dicus, E. W. Kolb, V. L. Teplitz, and R. V. Wagoner, Phys. Rev. **D17**, 1529 (1978).

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

32) G. Steigman, Ann. Rev. Astron. Astrophys. **14**, 339 (1976).

33) For a review see: E. W. Kolb and M. S. Turner, Ann. Rev. Nucl. Part. Sci. **33**, 645 (1983).

34) S. Weinberg, Phys. Rev. Lett. **42**, 850 (1979); D. Toussaint, S. B. Treiman, F. Wilczek, and A. Zee, Phys. Rev. **D19**, 1036 (1979).

35) E. W. Kolb and S. Wolfram, Phys. Lett. **B91**, 217 (1980); E. W. Kolb and S. Wolfram, Nucl. Phys. **B172**, 224 (1980); J. N. Fry, K. A. Olive, and M. S. Turner, Phys. Rev. **D22**, 2953, 2977 (1980).

36) J. N. Fry, K. A. Olive, and M. S. Turner, Phys. Rev. Lett. **45**, 2074 (1980).

37) J. N. Fry and M. S. Turner, Phys. Lett. **125B**, 379 (1983).

38) D. V. Nanopoulos and K. Tamvakis, Phys. Lett. **114B**, 235 (1982).

39) M. Srednicki, Nucl. Phys. **B202**, 327 (1982); D. V. Nanopoulos and K. Tamvakis, Phys. Lett. **110B**, 449 (1982).

40) M. Srednicki, Nucl. Phys. **B206**, 132 (1982).

41) S. Coleman, Phys. Rev. **D15**, 2929 (1977); C. Callan and S. Coleman, Phys. Rev. **D16**, 1762 (1977).

42) D. V. Nanopoulos, K. A. Olive, and K. Tamvakis, Phys. Lett. **115B**, 15 (1982).

43) D. V. Nanopoulos, K. A. Olive, M. Srednicki, and K. Tamvakis, Phys. Lett. **124B**, 171 (1983); C. Kounnas, D. V. Nanopoulos, and M. Quiros, Phys. Lett. **129B**, 223 (1983).

44) Ya. B. Zel'dovich and M. Y. Khlopov, Phys. Lett. **79B**, 239 (1979); J. P. Preskill, Phys. Rev. Lett. **43**, 1365 (1979).

45) G. 't Hooft, Nucl. Phys. **B79**, 276 (1974); A. M. Polyakov, Zh. Eskp. Teor. Fiz. Pisma Red. **20**, 194 (1974).

46) A. H. Guth and S. H. Tye, Phys. Rev. Lett. **44**, 631 (1980).

47) S. Weinberg, Phys. Rev. Lett. **48**, 1303 (1982).

48) J. Ellis, A. D. Linde, and D. V. Nanopoulos, Phys. Lett. **118B**, 323 (1982).

49) J. Ellis and K. A. Olive, Nature **303**, 379 (1983).

50) S. W. Hawking, Mon. Not. R. Ast. Soc. **142**, 129 (1969); C. B. Collins and S. W. Hawking, Mon. Not. R. Ast. Soc. **162**, 307 (1973).

51) S. W. Hawking and I. G. Moss, Phys. Lett. **110B**, 35 (1982).

52) S. Coleman and F. De Luccia, Phys. Rev. **D21**, 3305 (1980).

53) A. H. Guth and E. Weinberg, Phys. Rev. **D23**, 826 (1981); Nucl. Phys. **B212**, 321 (1983).

54) A. D. Linde, Phys. Lett. **108B**, 389 (1982); A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).

55) S. Coleman and E. Weinberg, Phys. Rev. **D7**, 1888 (1973).

56) A. Vilenkin, Phys. Lett. **115B**, 91 (1982); A. Vilenkin and L. H. Ford, Phys. Rev. **D26**, 1231 (1982).

57) A. D. Linde, Phys. Lett. **116B**, 335 (1982).

58) P. J. Steinhardt, Univ. of Penn. preprint UPR-0192T, 1982.

59) J. Ellis, D. V. Nanopoulos, K. A. Olive, and K. Tamvakis, Phys. Lett. **118B**, 335 (1982).

60) S. W. Hawking, Phys. Lett. **115B**, 295 (1982); A. H. Guth and S.-Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982); A. A. Starobinski, Phys. Lett. **117B**, 175 (1982); J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. **D28**, 679 (1983).

61) E. R. Harrison, Phys. Rev. **D1**, 2726 (1970); Ya. B. Zel'dovich, Mon. Not. R. Ast. Soc. **160**, 1P (1972).

62) W. H. Press, Phys. Scr. **21**, 702 (1980).

63) J. Breit, S. Gupta, and A. Zaks, Phys. Rev. Lett. **51**. 1007 (1983).

64) J. Wess and B. Zumino, Phys. Lett. **49B**, 52 (1974); J. Iliopoulos and B. Zumino, Nucl. Phys. **B76**, 310 (1974); S. Ferara, J. Iliopoulos and B. Zumino, Nucl. Phys. **B77**, 413 (1974); M. T. Grisaru, W. Siegel and M. Rocek, Nucl. Phys. **B159**, 420 (1979).

65) J. Ellis, D. V. Nanopoulos, K. A. Olive, and K. Tamvakis, Nucl. Phys. **B221**, 224 (1983).

66) Ya. A. Gol'fand and E. P. Likhtman, Pisma Zh. Eksp. Theor. Fiz. **13**, 323, (1971); D. Volkov and V. P. Akulov, Phys. Lett. **46B**, 109 (1973); J. Wess and B. Zumino, Nucl. Phys. **B70**, 39 (1974).

67) E. Cremmer, B. Julia, J. Scherk, S. Ferrara, L. Girardello, and P. van Nieuwenhuizen, Nucl. Phys. **B147**, 105 (1979); E. Cremmer, S. Ferrara, L. Girardello, and A. Van Proeyen, Phys. Lett. **116B**, 231 (1982) and Nucl. Phys. **B212**, 413 (1983).

68) P. J. Steinhardt and M. S. Turner, Fermilab preprint Pub-84/19-A.

69) J. Ellis, D. V. Nanopoulos, K. A. Olive, and K. Tamvakis, Phys. Let. **120B**, 331 (1983).

70) D. V. Nanopoulos, K. A. Olive, M. Srednicki, and K. Tamvakis, Phys. Lett. **123B**, 41 (1983).

71) R. Holman, P. Ramond, and G. G. Ross, Univ. of Florida preprint, 1984.

72) D. V. Nanopoulos, K. A. Olive, and M. Srednicki, Phys. Lett. **127B**, 30 (1983).