CM-P00061790

Ref.Th.2858-CERN
LAPP-TH-19

GRAND UNIFICATION AND COSMOLOGY

John Ellis
CERN -- Geneva

Mary K. Gaillard
LAPP -- Annecy-le-vieux

D.V. Nanopoulos
CERN -- Geneva

ABSTRACT

We review the interfaces between cosmology and grand unified gauge theories of elementary particle interactions. Particular emphasis is placed on fermion -- especially neutrino -- spectroscopy, the generation of the net baryon asymmetry observed in the Universe, dissipative effects, and the problem of the cosmological creation of grand unified monopoles.

Contribution to the proceedings of the
Europhysics Study Conference on
"Unification of the Fundamental Interactions"
Erice - Sicily, March 17 to 24, 1980

GRAND UNIFICATION AND COSMOLOGY

John Ellis
CERN -- Geneva

Mary K. Gaillard
LAPP -- Annecy-le-vieux

D.V. Nanopoulos
CERN -- Geneva

ABSTRACT

We review the interfaces between cosmology and grand unified gauge theories of elementary particle interactions. Particular emphasis is placed on fermion -- especially neutrino -- spectroscopy, the generation of the net baryon asymmetry observed in the Universe, dissipative effects, and the problem of the cosmological creation of grand unified monopoles.

Contribution to the proceedings of the
Europhysics Study Conference on
"Unification of the Fundamental Interactions"
Erice - Sicily, March 17 to 24, 1980

GRAND UNIFICATION AND COSMOLOGY *)

John Ellis
CERN, Geneva, Switzerland

Mary K. Gaillard
LAPP, Annecy-le-Vieux, France

D.V. Nanopoulos
CERN, Geneva, Switzerland

1. A COSMIC CONNECTION?

Do cosmology and grand unified theories (GUTs) of elementary particle interactions have anything useful to say to each other? There is a great deal of theoretical work on GUTs unifying the strong, weak and electromagnetic interactions¹, now that many theorists perceive these individual interactions to be understood in principle. GUTs invoke energy scales of $O(10^{15})$ GeV which seem vertiginous to many physicists. It is not imaginable to reach these energies in laboratory experiments, and experimentalists are therefore forced to look for very indirect and feeble side-effects of grand unification such as proton decay. Even these valiant efforts may be brought to nought by a (logarithmically) modest increase in the grand unification mass-scale². However, this mass-scale may be achieved directly in cosmological and astrophysical situations. For example, black hole explosions could in principle achieve temperatures up to the Planck temperature of 10^{32} K corresponding to energies of 10^{19} GeV, while temperatures corresponding to particle energies of 10^{15} GeV or more are generally thought to have occurred very early in the Big Bang when the Universe was about 10^{-37} seconds old. It may indeed turn out that the Big Bang has been the only place to do direct experiments on GUTs (albeit rather uncontrolled, at least by us). In this paper we review progress on some interfaces between cosmology and grand unification, asking whether cosmology tells us important things about GUTs, and vice versa.

---

*) Presented by John Ellis.

Table
Important Problems

Grand Unification	Cosmology
Fermion spectroscopy	Baryon to photon ratio, and lepton numbers
Grand unification mass $m_X = ?$ and coupling constant $\alpha_{GUT} = ?$	Homogeneity and isotropy
SU(5) or a more complicated GUT?	Galaxy formation
Mechanism of symmetry breaking	Simultaneity of expansion of causally separated parts of the Universe
Connection with gravitation?	How many black holes and/or monopoles in the Universe?

Shown in the Table are a few of the important problems in grand unification and cosmology: neither listing is exhaustive. Within the rubric of fermion spectroscopy we include the absolute number of fermion species and the masses of the apparently very light neutrinos as well as the possible existence of superheavy fermions. As far as the grand unification mass m_X is concerned, we have a phenomenological lower bound from the absence to date of observations of proton decay³, and a suggestion⁴ from the observed strength of the strong interactions that m_X may lie very close^{5-7,2} to this lower bound: direct observation is, however, still lacking as yet. Simple-minded calculations⁵⁻⁷ suggest that the strength $\alpha_{GUT} = g^2/4\pi$ of the gauge coupling at the grand unification mass of order 10^{15} GeV is around 1/40, but some authors have suggested that α_{GUT} may be $O(1)$ at or near the Planck mass scale³. The biggest open problem in gauge theories is perhaps the mechanism of spontaneous symmetry breaking -- is it realized through explicit Higgs fields, or dynamical symmetry breaking, or a combination of the two? Eventually one must of course include gravitation in a truly unified theory of all the fundamental interactions³, though present attempts at grand unification generally postpone this problem *ad infinitum*.

One of the outstanding problems in cosmology has been to understand why there are just a few baryons in the Universe: $n_B/n_\gamma = 10^{-9 \pm 1}$, and why there seem to be no antibaryons¹⁰, at least on the scale of our local galactic cluster. Related unknowns are the densities (chemical potentials) of other quantities not yet known to be violated, such as electric charge and the various lepton numbers (L_e, L_μ, L_τ). Another fundamental problem is to understand why the Universe is apparently so homogeneous and isotropic, for example in the distribution of luminous matter within it. The reader is actually living (?) evidence that some inhomogeneities exist, of which the most notable are galaxies, whose formation is not yet fully understood. But the Universe looks quite regular on a large scale, and a particularly striking and baffling aspect of this homogeneity and isotropy of the Universe is the apparent necessity within the standard Big Bang model for causally separated parts of the Universe to have chosen to behave in the same way at the same time. Presumably it was not coincidence, and the causality arguments must somehow be thwarted by some deviation from our "standard" ideas on particle interactions or on the evolutionary history of the Universe. Finally, on a more mundane level, our unified theories of the fundamental interactions predict the existence of massive classical solutions¹¹, grand unified monopoles, whose abundance is unknown¹² but must be strongly suppressed in order to avoid a conflict with the standard Big Bang model. This might prove to be a serious embarrassment for simple grand unified models, though we ourselves doubt it.

In this review we will not address, let alone answer, all the above problems, but will address ourselves to four topics where the cosmology-GUT interface seems particularly active. The first is the topic of fermion spectroscopy, in which we remind the reader that the successful calculation^{5,13,14} of the bottom quark mass in GUTs imposes a phenomenological restriction to six of the number of quarks. This parallels nicely the restriction to three of the number of neutrinos due to nucleosynthesis¹⁵ in the Big Bang. The second topic is baryon number generation in grand unified theories¹⁶. These provide a qualitative framework for generating a net baryon excess through C-, CP- and B-violating interactions dropping out of thermal equilibrium very early in the Big Bang. When one tries to put this scenario on a quantitative basis one discovers¹⁷ that the grand unification mass should not be much lower than 10^{15} GeV, or α_{GUT} much larger than $1/40$, and that a theory larger¹⁸ than the minimal SU(5) GUT with just two Higgs representations (5 and $\bar{24}$) is probably required¹⁹. The third topic is that of dissipative processes very early in the Universe, due to the relatively long mean free times of elementary particles possibly providing forms of viscosity and thermal conductivity. These suppress²⁰ some forms of inhomogeneity, but

permit the growth of long wavelength compressional perturbations which have been suggested²¹ as the origin of galaxy formation. The fourth topic is the cosmological density of grand unified monopoles¹². It has been suggested²² that this may be brought acceptably low by postulating a first order phase transition with large supercooling, or that one may have to postulate²³ a GUT larger than SU(5), or that symmetry breaking may have to be dynamical²⁴. In fact, none of these may²⁵ be necessary.

It will be seen from the above synopsis that on many issues cosmology and grand unification do indeed have things to say to each other, though the utility of some of the connections is not yet clear.

3. GUTS AND FERMION SPECTROSCOPY

Before embarking on the amalgamation of grand unification and cosmology, perhaps we should first establish the credentials of GUTs for our cosmological friends -- we will, of course, take on trust their "standard" Big Bang model²⁶. GUTs^{1,27} have an immense aesthetic appeal in that they explain many previous baffling aspects of particle physics -- the quantization of electric charge $Q_e = -Q_p$ and the apparent "family resemblances" of quarks and leptons. While they have a unique gauge coupling g_{GUT} instead of the disunified $g_3 \neq g_2 \neq g_1$ of the low-energy $SU(3) \times SU(2) \times U(1)$ gauge theory, they do however have many other free parameters, notably including those associated with the Higgs fields and their couplings. These include the Higgs-fermion-fermion Yukawa couplings which fix quark and lepton masses, mixing angles and CP-violating phases²⁸. They also include the parameters of the Higgs potential which determine the largely unknown Higgs boson masses⁵, and at least one of these parameters must be fixed^{29,5} with a precision of $O(10^{-25})$ in order to yield the enormous hierarchical ratio

$$\frac{m_X}{m_W} \approx (10^{12} \text{ to } 10^{13}) \quad (1)$$

of vector boson masses expected in GUTs. Phenomenological estimates^{2,5-7,30} of the mass m_X of the baryon-number violating gauge bosons X, on the basis of the observed strengths of the strong and electromagnetic interactions at low energies, suggest that $m_X \approx 6 \times 10^{14}$ GeV with substantial uncertainties^{2,31} except in the minimal SU(5) GUT. With this value of m_X , protons and bound neutrons are expected to decay with lifetimes less than 10^{32} years which could be detected in forthcoming experiments³². However, proton decay has not yet been seen, and one would like to see at least some indirect evidence for GUTs in low-energy phenomena.

Two such pieces of evidence exist. The most celebrated is the $SU(2) \times U(1)$ mixing angle $\sin^2 \theta_w$ which has been calculated^{4,5,7} in simple GUTs such as $SU(5)$ and $SO(10)$. In these models one has

$$\sin^2 \theta_w = \frac{\frac{3}{5} g_1^2}{g_2^2 + \frac{3}{5} g_1^2} \quad (2)$$

which is $3/8$ in the symmetry limit at high energies where $g_2 = g_1$. As one comes down below m_X to low energies Q the couplings $g_1(Q)$, $g_2(Q)$ evolve differently so that the value of formula (1) also changes. When $Q \approx m_W$ one finds

$$\sin^2 \theta_w (m_W) \approx 0.20 \text{ to } 0.21 \quad (3)$$

to be compared with the experimental value³³

$$\sin^2 \theta_w (Q \ll m_W) \approx 0.230 \pm 0.015 . \quad (4)$$

To make a direct comparison between (3) and (4) requires a complete computation of radiative corrections at low energies which has not yet been done. However, the agreement between (3) and (4) for a number which *a priori* could have lain anywhere between 0 and 1 is surely impressive.

A second piece of evidence for GUTs comes from a calculation (Refs. 5, 13 and 14) of the bottom quark mass. In many GUTs there is a relation between this and the mass of the τ lepton:

$$m_b = m_\tau \quad (5)$$

valid in the high-energy symmetry limit. At low Q , where m_b and m_τ are actually measured, the prediction (5) gets renormalized^{5,13,14} to

$$\frac{m_b}{m_\tau} (Q = 10 \text{ GeV}) \approx \left[\frac{\alpha_s(Q)}{\alpha_X} \right]^{4/(11-2/3 f)} [1 \pm 0(10)\%], \quad (6)$$

where α_X is the grand unified coupling at $Q \approx m_X$ and f is the number of quark flavours. The observed mass of the τ can be used with (6) to deduce that

$$m_b \approx (5 \text{ to } 5\frac{1}{2}) \text{ GeV} \quad (7)$$

if there are only six quark flavours. The number (7) is increased by 30% (or more) if there are eight (or more) quarks. If the basis of this successful calculation is accepted, it means that there can be at most six quarks. Since there are two quarks per neutrino in simple GUTs, this limit parallels very nicely the limit¹⁵ of 3 on the number of (almost) massless neutrinos ($m_\nu \ll \ll 1$ MeV) coming from nucleosynthesis.

This parallelism is meaningful only if GUTs imply $m_\nu \ll 1$ MeV for all their neutrinos. Recall that the present laboratory limits are

$$m_{\nu_e} \lesssim 35 \text{ eV}, m_{\nu_\mu} \lesssim 600 \text{ keV}, m_{\nu_\tau} \lesssim 200 \text{ MeV} . \quad (8)$$

Most fashionable GUTs do indeed predict $m_\nu \ll 1$ MeV, with a common tendency³⁴ to find

$$\frac{m_\nu}{m_W} \approx \frac{m_W}{m_X} \Rightarrow m_\nu \sim (10^{-5} \text{ to } 1) \text{ eV} \quad (9)$$

the second hierarchy emerging naturally given the (unnatural) first one (1). Neutrino masses in the lower end of the range (9) are best detected by astrophysical means such as oscillations in cosmic ray³² or solar neutrinos³⁵, or time-delays in pulses from supernova collapses. Neutrino masses ≥ 1 eV fall within the sensitivity of present laboratory experiments³⁶, as well as having possible astrophysical implications through their aggregation in galactic haloes and/or clusters. The astrophysical and cosmological constraints on such neutrinos are reviewed by Steigman⁷ at this meeting.

3. THE EVOLUTION OF THE VERY EARLY UNIVERSE

The Universe now contains many photons, principally in the 3°K microwave background radiation, and relatively few baryons by comparison:

$$n_B/n_\gamma = 10^{-9 \pm 1} . \quad (10)$$

Furthermore, there are no known concentrations of antimatter and the absence of high-energy photons which could be the products of $B\bar{B}$ annihilation indicate that our local cluster of galaxies does not contain any significant fraction of antimatter¹⁷. The cosmic ray antiprotons recently observed³³ do not constitute evidence for cosmological antimatter, as their flux is consistent with that expected from secondary p production by primary nucleons. Before the advent of GUTs there was no satisfactory explanation

of the ratio (10) or the apparent matter-antimatter asymmetry. Some held that the Universe is in fact symmetric and contains equal amounts of matter and antimatter, but failed to explain satisfactorily how their concentrations could have got separated in the present Universe³⁹. Others hypothesized that the Universe had in fact started out with only baryons, and the large number of photons in (10) was generated subsequently by dissipative processes. Yet others just assumed that the Universe had always been slightly asymmetric. It now seems possible to believe that the matter-antimatter asymmetry and the small number (10) can be naturally explained¹⁵ in GUTs. Three crucial ingredients are necessary in order to realize this trick. One is clearly that one has a theory with baryon number B-violating interactions. Another is that some component of these interactions violates both C and CP, as making either of these transformations reverses the sign of the net baryon number which must therefore be zero if the Universe evolves from an initially symmetric state. Finally, the B-, C- and CP-violating interactions must get out of thermal equilibrium, as in such a state the arrow of time is lost and CPT invariance guarantees that the net baryon number is zero^{17,40}. We will now see how these three conditions are fulfilled with the expansion and cooling of the Universe introducing the necessary departure from thermal equilibrium while GUTs provide the necessary B-, C- and CP-violation.

We will be interested in epochs of the Universe's evolution when it was very hot with a temperature of $(10^{32} \text{ to } 10^{23})^\circ\text{K}$ equivalent to a particle energy $O(10^{19} \text{ to } 10^{10}) \text{ GeV}$. The upper limit is given by the Planck temperature T_p corresponding to a mass m_p related to Newton's gravitational constant by $m_p \equiv (G)^{-1/2} \approx 1.2 \times 10^{19} \text{ GeV}$. The quantum gravitational interactions between particles at or above T_p or m_p are $O(1)$, conventional classical approximations break down, and we cannot even speculate sensibly any more. Below 10^{19} GeV quantum gravitational effects presumably decrease as a power of the temperature T , and we can use classical solutions of the Einstein equations, typically Robertson-Walker-Friedman (RWF) metrics, to describe the large scale structure of gravitational effects. The expansion of the Universe in the presence of curvature k is then described²⁶ by

$$\frac{\ddot{R}}{R} = -4\pi G \left(\rho + \frac{p}{3} \right) \quad (11a)$$

$$\left(\frac{\dot{R}}{R} \right)^2 = \frac{8\pi G \rho}{3} - \frac{k}{R^2} \quad (11b)$$

where R is the RWF scale parameter and ρ and p are the energy density and pressure of matter in the Universe. At very high temperatures when all particles are ultrarelativistic we have $p = \frac{1}{3}\rho$ and

$$\rho = g_T B T^4, \quad (12)$$

where g_T is the effective number of particle helicity states (1 for a scalar, 2 for a massless vector, 7/4 for a two-component spinor, etc.) and B is a constant taking the value

$$B = \frac{\pi^2}{30} \quad (13)$$

if the particles are in kinetic equilibrium. Solving the evolution equations (11) with the particle energy density (12) yields an age-temperature relation for the Universe:

$$\frac{t(T)}{t_P} = \left(\frac{3}{32\pi g_T B} \right)^{\frac{1}{2}} \left(\frac{T_P}{T} \right)^2, \quad (14)$$

where t_P is the Planck time $O(10^{-42})$ sec related to the Planck energy of 10^{19} GeV by the usual uncertainty relation.

The relativistic particles in the hot but cooling and expanding Universe have interactions of which the most important are $(2 \leftrightarrow 2)$ scattering processes (Fig. 1a) and $(1 \leftrightarrow 2)$ decays and inverse decays (Fig. 1b). Decays and inverse decays of a massive particle X will be characterized by a time-dilated version of the free decay time:

$$\tau \propto \frac{(T^2 + m_X^2)^{\frac{1}{2}}}{\alpha_X m_X^2} \quad (15)$$

which is very long in the limit that $T \gg$ the mass m_X . In fact, many or all of the particles in GUTs acquire their masses from spontaneous symmetry breaking, so that they are functions of temperature and are zero above some critical temperature T_c which is generally $\geq m_X$. This entails a further lengthening of τ compared with (15). At low temperatures $T \ll m_X$ the density of massive particles is suppressed by a Boltzmann factor $e^{-m_X/T}$, and there is a similar reduction in the probability for two light particles to come together to make a massive particle. This means that a true measure of the mean free time $\tau_{1 \leftrightarrow 2}$ between $1 \leftrightarrow 2$ interactions is related to τ (15) by

$$\tau_{1 \leftrightarrow 2} \approx \tau \times \exp - (m_X/T). \quad (16)$$

Fig. 1 Examples of important fundamental interactions:
 (a) $(2 \leftrightarrow 2)$ scattering, (b) $(1 \leftrightarrow 2)$ decays and
 inverse decays.

Dimensional arguments would suggest that in the ultrarelativistic limit $2 \leftrightarrow 2$ scattering in a GUT with a dimensionless coupling constant α_{GUT} would behave as

$$\sigma(2 \leftrightarrow 2) \approx \frac{\alpha_{\text{GUT}}^2}{T^2} \quad (17)$$

possibly modified by logarithmic factors, and for processes involving vector boson exchange the best guess^{5,6} is that $\alpha_{\text{GUT}} \sim 1/40$. In fact, if the scattering particles were free, one would have from the t-channel exchange of a massive boson X

$$\sigma(2 \leftrightarrow 2) \approx \frac{\alpha_{\text{GUT}}^2}{m_X^2}$$

at high energies, but in the primordial plasma we are considering there is a reduction⁴¹ of σ due to screening effects. These reduce the cross-section to at most $\alpha_{\text{GUT}}^2 \lambda_D^2$, where the Debye length⁴¹

$$\lambda_D \approx \left(\frac{T}{32\pi^2 \alpha_{GUM} n} \right)^{\frac{1}{2}} \quad (18)$$

and n is some relevant number density:

$$n = g_T^A T^3 \quad (19)$$

with $A \approx 0.12$ for a Boltzmann distribution. If (?) the relevant n corresponds to $g_T = 1$, then for $\alpha_{GUM} \approx 1/40$, $\lambda_D \approx 1/T$ and the dimensional cross-section formula (17) is applicable. If we consider $(2 \leftrightarrow 2)$ scattering (Fig. 1a) via a boson whose mass m_X is comparable to the temperature T then the cross-section (15) is suppressed by propagator effects

$$\sigma(2 \leftrightarrow 2) \approx \frac{\alpha_{GUM}^2 T^2}{(T^2 + m_X^2)^2} \quad (20)$$

which means that the scattering is greatly suppressed at temperatures $T \ll m_X$. From the cross-section (17) one can deduce a mean free time $\tau_{2 \leftrightarrow 2}$ between collisions:

$$\tau_{2 \leftrightarrow 2} = \frac{1}{n\sigma(2 \leftrightarrow 2)} \quad (21)$$

and we see from (17) and (19) that

$$\tau_{2 \leftrightarrow 2} \propto \frac{1}{\alpha_{GUM}^2} \frac{1}{T} \quad (22a)$$

at very high temperatures. At very low temperatures $T \ll m_X$, Eqs (19) and (20) imply that for $(2 \leftrightarrow 2)$ scattering through massive particle exchange Eq. (22a) is modified to become

$$\tau_{2 \leftrightarrow 2} \propto \frac{1}{\alpha_{GUM}^2} \frac{m_X^4}{T^5} \quad (22b)$$

In the neighbourhood of $T \sim m_X$ neither (22a) nor (22b) is appropriate. If the X exchange takes place in a crossed channel there will be a simple interpolating form for $\tau_{2 \leftrightarrow 2}$. However, if the X exchange takes place in the direct channel, there is a possibility

of enhancing $\sigma_{2 \leftrightarrow 2}$, and reducing $\tau_{2 \leftrightarrow 2}$ correspondingly, because of scattering through X bosons close to their mass shell⁴¹. It is convenient to explicitly remove these pole contributions from $\sigma_{2 \leftrightarrow 2}$ and lump them together with the $(1 \leftrightarrow 2)$ decays and inverse decays to which they are directly related by unitarity, leaving behind a reduced $(2 \leftrightarrow 2)$ cross-section which leads to a smooth interpolation of $\tau_{2 \leftrightarrow 2}$ between (22a) and (22b).

We can now determine approximately the epochs during which various interactions are in equilibrium by looking at the ratio between the expansion time t (14) and the various different mean free times τ . If $t/\tau < 1$, the corresponding interaction is out of equilibrium, whereas if $t/\tau > 1$ then the interaction is in equilibrium. We have several different cases to consider: the results are illustrated in Fig. 2. Hereafter unless explicitly stated to the contrary, we will use "natural" units in which Planck units (m_P , T_P , t_P , ...) are unity.

Quantum gravitational effects

If we assume that at $T < T_P = 1$, $\sigma_{\text{grav}} \leq G$ (and $\sigma_{\text{grav}} \approx G^2 T^2$ seems likely to be a realistic estimate for $(2 \leftrightarrow 2)$ scattering) then

$$\tau_{\text{grav}} \geq \frac{1}{GT^3} \quad (\text{for example } 1/G^2 T^5), \quad (23)$$

whereas $t \sim T^{-2}$ so that

$$\left. \frac{t}{\tau} \right|_{\text{grav}} \lesssim T \quad (24)$$

and quantum gravitational interactions are expected to be relatively unimportant and out of equilibrium for $T < 10^{19}$ GeV.

Strong, weak and electromagnetic interactions

Since the masses of particles mediating these interactions are negligible on the energy scales we are considering, the cross-section (17) should be applicable and we expect

$$(t/\tau)_{\text{SWEM}} \approx \frac{\alpha_{\text{GUM}}^2}{T}. \quad (25)$$

Fig. 2 The ratios of the universal expansion time (14) to the mean free times for different interactions^{4,2} discussed in the text. Note the possibility that no known interactions were in equilibrium at $10^{19} \text{ GeV} > T > 10^{15} \text{ GeV}$, the probable period of equilibrium for baryon number violating interactions at $T \sim 10^{15} \text{ GeV}$, and their subsequent non-equilibrium period.

This ratio is $O(10^{-4})$ at $T \sim 10^{19}$ GeV, meaning that these interactions should be out of thermal equilibrium at that temperature, a state of affairs which should persist until $T \sim \alpha^2 \sim 10^{15}$ GeV, after which strong, weak and electromagnetic interactions should be in equilibrium. This equilibrium state should be maintained until much lower temperatures, below those at which the weak interaction symmetry is spontaneously broken and the colour interactions are shielded by long-distance non-perturbative effects.

Grand unification interactions

By these we mean the interactions of (or involving exchanges of) the superheavy X bosons of a GUT. As far as the $(2 \leftrightarrow 2)$ interactions are concerned, we expect at $T \gg m_X$ a behaviour like that of the strong, weak and electromagnetic interactions (25), while at temperatures $T \ll m_X$ we expect from (22b) that

$$\left(\frac{t}{\tau}\right)_{2 \leftrightarrow 2 \text{ GUT}} \approx \frac{\alpha_X^2 T^3}{m_X^4} \quad (26)$$

with a smooth interpolation between (25) and (26) at $T \sim m_X$. This function is therefore $O(1)$ at $T \sim m_X \sim \alpha_X^2$. For $(1 \leftrightarrow 2)$ interactions we expect from (15) and (16) that

$$\left(\frac{t}{\tau}\right)_{1 \leftrightarrow 2 \text{ GUT}} \approx \frac{\alpha_X m_X^2}{(T^2 + m_X^2)^{1/2} T^2} e^{-m_X/T} . \quad (27)$$

This function is clearly $\ll 1$ at $T \gg m_X$ and at $T \ll m_X$, peaking for T near m_X at a value of order $(\alpha_X/m_X) \sim 10^2$. The ratios $(t/\tau)_{1 \leftrightarrow 2, 2 \leftrightarrow 2 \text{ GUT}}$ are plotted⁴² in Fig. 2, and we see that they suggest that for an epoch where T was of order m_X the interactions involving X bosons may have been in equilibrium.

One can make similar estimates for the interactions of the superheavy Higgs bosons H_X expected in GUTs. Shown in Fig. 2 are the curves corresponding to $(2 \leftrightarrow 2)$ interactions via Higgs exchange and $(1 \leftrightarrow 2)$ Higgs decays and inverse decays, assuming a Higgs mass $\sim 10^{14}$ GeV and a Higgs coupling $\alpha_H \sim 10^{-3}$. Though these parameters are clearly subject to some uncertainty, we see a tendency to have a period in which H_X processes are in equilibrium.

Given the general qualitative features of the expansion time to mean free time ratios shown in Fig. 2, there are three qualitatively different scenarios⁴² for generating the baryon asymmetry observed in the Universe, illustrated in Fig. 3.

Fig. 3

Fig. 3 The various scenarios^{4,2} (A), (B) and (C) described in the text for baryon number generation in the context of GUTs

(A) The Universe might have started at $T \sim 10^{19}$ GeV with a baryon-antibaryon asymmetry $\gg 10^{-9}$, or such a large asymmetry might have been generated by interactions at $T > m_X \sim 10^{15}$ GeV. This baryon asymmetry would then have been reduced during the period of grand unified equilibrium around $T \sim 10^{15}$ GeV (in accord with general theorems^{17,40}) to the value $10^{-9 \pm 1}$ observed today^{43,17}.

(B) Again one considers the possibility that the baryon asymmetry may have been large at temperatures $> m_X$, but that grand unified interactions in equilibrium may then have pushed the number much below $10^{-9 \pm 1}$, only to have X and H_X interactions in the post-equilibrium period regenerate the required number $10^{-9 \pm 1}$.

(C) Perhaps the Universe really did start off at $T > 10^{15}$ GeV with a baryon asymmetry which was either zero or very small $\ll 10^{-9 \pm 1}$. Then as X and H_X interactions drop out of thermal equilibrium after 10^{15} GeV the required baryon number of $10^{-9 \pm 1}$ per photon could have been generated.

Examining these different alternatives, we see that options (B) and (C) give us the virtue of naturalness: the final asymmetry is determined by GUT considerations independent of the initial conditions. In contrast, option (A) requires a very delicate fine-tuning¹⁷ of the interaction rates if the total wash-out of the baryon number is to be avoided. It does not require a C- and CP-violating component in the B-violating GUT interactions, but we have strong reasons to believe that such a component exists. In the rest of this discussion we will discuss the baryon asymmetry¹⁶ generated by C- and CP-violating interactions of GUTs in options (B) and (C).

4. MECHANISMS FOR BARYON NUMBER GENERATION

The generation of the net baryon number (10) may have been quite complicated, with several different baryon-number violating processes contributing and/or competing. However, it seems likely that the dominant driving mechanism may have been the decays and inverse decays of some superheavy boson^{40,44} either a gauge vector X or a Higgs scalar H_X . In most simple^{1,27} GUTs the dominant decays of these particles are expected (Fig. 4) to be into pairs of quarks or into antilepton and antiquark:

$$X \text{ or } H_X \rightarrow (q + q) \quad \text{or} \quad (\bar{q} + \bar{q}) . \quad (28)$$

Fig. 4 Dominant decay modes of superheavy vector bosons X and Higgs bosons H_X .

The total decay rates of particles and antiparticles must be the same because of CPT:

$$\Gamma_{\text{tot}}(X) = \Gamma(X \rightarrow qq) + \Gamma(X \rightarrow \bar{q}\bar{\ell}) + \dots \quad (29)$$

$$\Gamma_{\text{tot}}(\bar{X}) = \Gamma(\bar{X} \rightarrow \bar{q}q) + \Gamma(\bar{X} \rightarrow q\ell) + \dots$$

but the partial decay rates into conjugate channels may differ if C and CP are both violated:

$$B \equiv B(X \rightarrow \bar{q}\bar{\ell}) \equiv \frac{\Gamma(X \rightarrow \bar{q}\bar{\ell})}{\Gamma_{\text{tot}}(X)} \neq \bar{B} \equiv B(\bar{X} \rightarrow q\ell) \equiv \frac{\Gamma(\bar{X} \rightarrow q\ell)}{\Gamma_{\text{tot}}(\bar{X})}. \quad (30)$$

Then if one naïvely assumes that one starts with an identical population of X and $\bar{X}$ bosons (or of H_X and $H_{\bar{X}}$) their decays would generate a net baryon to photon number density ratio⁴⁴

$$\frac{n_B}{n_\gamma} \approx \left(\frac{g_X}{g_{\text{tot}}} \right) (\bar{B} - B) \approx (10^{-2}) \Delta B \quad (31)$$

where g_X is the number of helicity states of X bosons and g_{tot} is the total number of helicity states. We will see later that this naïve decay mechanism is somewhat over-simplified⁴⁵, but it suffices for discussions of some qualitative features of the baryon generation problem.

It is clear that no difference ΔB in the branching ratios of particles and antiparticles can arise in lowest order. Such a non-zero ΔB can first arise in fourth order, through final state interactions (see Fig. 5) involving bosons which violate B conservation⁴⁴. The resulting difference in branching ratios is of order

$$\Delta B \approx \text{Im Tr}(abcd) / \sum_a \text{Tr}(aa^*) , \quad (32)$$

Fig. 5 A sample 4th order interference diagram which could (see Eq. (32)) contribute to ΔB and hence to the baryon asymmetry.

where a, b, c and d are the coupling matrices introduced in Fig. 5. In fact, in simple models the first C- and CP- violating ΔB often occurs only in higher orders. In a large class of GUTs all the CP violation originates from the fermion-fermion-Higgs Yukawa couplings. As in the simplest case (32), all leading CP-violating rate differences such as ΔB are related to products of Yukawa coupling matrices traced along fermion lines. These traces may arise from thermal averaging over initial states as well as from sums over possible final states, and occur for scattering processes as well as for $(1 \leftrightarrow 2)$ decays and inverse decays. It is then important in the context of any particular model to determine the lowest order trace of Yukawa couplings which violates C and CP. The lowest order contributions to the baryon asymmetry can then be got from this trace by connecting up the Yukawa vertices by Higgs lines in all possible ways, and then cutting the resulting diagrams in all possible ways corresponding to processes with different initial and final states and different possible arrangements of the final state interactions. As an example, we consider⁴² the "minimal" SU(5) model with the usual $\underline{5} + \underline{10}$ fermions and one $\underline{24}$ and one $(\underline{5} + \underline{\bar{5}})$ pair of Higgs representations. In this model the first CP-violating trace is 8th order in the Yukawa coupling -- see Fig. 6a. Many different ways of connecting up the Higgs vertices are possible, of which one example is shown in Fig. 6b. This diagram and its fellows may then be cut in many different ways corresponding to different physical processes, as indicated in Figs 6 c,d. Each one of these cuts corresponds to a direct CP-violating interaction rate $((1 \leftrightarrow 2) \text{ or } (2 \leftrightarrow 2) \text{ etc.})$ which may contribute to the total baryon asymmetry generated. Suppose we consider the contribution of the cut in Fig. 6c which gives a CP-violating difference ΔB in decay branching ratios. Each Yukawa vertex is $O(m_f/m_W \times g)$ and the CP-violating effect is proportional to GIM-like products of differences in quark masses. If we assume there are just three generations of $\underline{5} + \underline{10}$ fermions and that $m_t \gg m_b \gg m_c \gg \text{other fermion masses}$, then we have a CP-violating

Fig. 6 (a) One of the lowest order C-, CP- and B-violating fermion traces in minimal SU(5). (b) The Higgs-fermion-fermion vertices of the trace connected up in one of many possible ways. (c) The previous diagram cut in such a way as to contribute a CP-violating ΔB in Higgs decays. (d) The same diagram cut so as to make a CP-violating difference between Higgs-fermion scattering cross-sections - cf. Fig. 1(a). The dashed lines are $\underline{5}$ Higgs bosons, the solid lines $\underline{10}$ fermions, the zig zag lines $\underline{5}$ fermions, and the dotted lines are unitarity cuts.

$$\Delta B \lesssim \frac{1}{10} \alpha_X^4 \frac{m_b^4 m_t^3 m_c}{m_W^8} / (m_t^2/m_W^2) \alpha_X, \quad (33)$$

where the $1/10$ is an upper limit on the product of mixing angles and phase factors, and we have neglected various numerical factors from phase space, etc., as well as combinatorial factors from the large number of similar diagrams. Putting in the known quark masses we find

$$\Delta B \lesssim 10^{-14} \frac{m_t}{m_W} \leq 10^{-14} \quad (34)$$

which Eq. (31) demonstrates to be too small to generate the observed baryon asymmetry of order $10^{-9 \pm 1}$. A model more complicated than minimal SU(5) (more Higgs representations? superheavy fermions?) is necessary^{18,44} to generate the observed number, and can easily be constructed. (Note that any process involving a gauge vector boson in this model, whether in an initial or final state or in a virtual interaction, has CP violation only in at least one order of α higher.)

Before discussing in more detail the mechanics of computing the baryon asymmetry in specific models it is perhaps worth making a few general remarks about the nature of the CP-violation necessary for generating non-zero baryon number. It is clear that one needs a "hard" form of CP-violation which persists up to high energies $\sim 10^{15}$ GeV, and therefore that models where CP-violation is generated by spontaneous breakdown on a weak interaction scale of order 100 GeV are not suitable. Furthermore, models where CP is violated spontaneously at $O(10^{15})$ GeV are also not acceptable. At these very early epochs causally connected domains in the Universe are very small, and in a scenario of spontaneous CP violation there would be no correlation between the signs and perhaps magnitudes of the baryon number generated in different domains. At subsequent epochs previously separated domains become connected, and the net baryon asymmetry we see today would tend to be a statistically averaged, and hence much diluted, version of the asymmetry possible in any individual domain. It therefore seems that a scenario of spontaneous CP-violation is difficult to maintain in the absence of acausal effects, and that some form of intrinsic CP-violation like that found^{1,28} in the minimal SU(5) model is to be preferred. One might wonder whether the sign of the baryon asymmetry can be related to the sign of CP-violation in the $K^0-\bar{K}^0$ system. Quite apart from the computational problems of getting numbers for either of these physical quantities from parameters of a fundamental Lagrangian, it seems unlikely in general that the same CP-violating

parameter controls both phenomena. For example, in the standard SU(5) with N_g (e.g. 3) generations the generalized Kobayashi-Maskawa matrix characterizing low-energy weak interactions has $(N_g-1)^2$ (e.g. 4) parameters of which $\frac{1}{2}(N_g-1)(N_g-2)$ (e.g. 1) are CP-violating phases "observable" via $K^0-\bar{K}^0$ phenomena, while at high energies there are^{2,3,4,2} (N_g-1) (e.g. 2) additional CP-violating phases, one of which determines the dominating CP-violating rate (33). In fact, even in the minimal SU(5) model CP-violation and hence a net baryon asymmetry is possible with only four quarks, though the number generated increases if more quarks or other structure is introduced, and a realistic model probably needs^{4,6} at least six quarks.

There are several obstacles remaining before a quantitative calculation of the baryon asymmetry can be made. The first is that because the minimal SU(5) model is too small, one must go^{1,3} to a more complicated model with more free parameters, which hinders the attainment of numerical precision. Secondly, the naïve decay scenario outlined above is incomplete. This is because, as can be seen from Fig. 2, $(2 \leftrightarrow 2)$ interactions are still significant and near equilibrium around the time of the heavy particle decays. Since the net baryon asymmetry must be zero in equilibrium, this suggests that these $(2 \leftrightarrow 2)$ interactions have a tendency to reduce the net asymmetry generated by $(1 \leftrightarrow 2)$ interactions. This washing-out has been computed numerically by various groups^{1,7,4,7,4,8} integrating up the Boltzmann equations using both $(1 \leftrightarrow 2)$ and $(2 \leftrightarrow 2)$ interactions. They find that the amount of wash-out is very sensitive both to the masses of the superheavy bosons and to the grand unified coupling α_X . We see from Fig. 7 that the efficiency of B generation via vector bosons X with $\alpha_X \sim 1/40$ is 100% efficient only for $m_X \geq 10^{17}$ GeV, while for Higgs bosons H_X with $\alpha_{H_X} \sim 10^{-3}$ complete efficiency persists down to $m_{H_X} \approx 10^{15}$ GeV. If one requires an efficiency $\geq 0.1\%$ in order to get a large enough n_b/n_γ , then one needs (see also the paper by Weinberg and the second paper by Yoshimura in Ref. 16):

$$m_X \geq 10^{15} \text{ GeV} \quad \text{or} \quad m_{H_X} \geq 10^{13} \text{ GeV} . \quad (35)$$

For comparison, the latest best estimates from low-energy phenomenology^{2,6,7,30} are

$$m_X \approx 6 \times 10^{14} \text{ GeV} , \quad m_{H_X}/m_X \approx 10^{0 \pm 1} \quad (36)$$

which are just consistent with the constraints (35). The greater amount of washing-out for vector boson processes is another reason (in addition to their intrinsically higher order of CP-violation) why one might expect the dominant contribution to the baryon asymmetry to come from processes involving superheavy Higgses. There

Fig. 7 A diagram from the work of Kolb and Wolfram¹⁷ illustrating how the net baryon number obtained from a CP-violating difference in decay branching ratios may be less than expected due to the wash-out by $(2 \leftrightarrow 2)$ interactions.

is one complication with these wash-out calculations that should be mentioned. It is that in models with reducible fermion representations for each generation of fermions, such as SU(5) with its $\bar{5} + 10$ representations, the vector bosons which are the dominant $(2 \leftrightarrow 2)$ interactions cannot⁴⁷ wash out all combinations of quark asymmetries because of helicity conservation. This means that in such "reducible" models the constraints (35) may be relaxed somewhat, though the degree of relaxation has not yet been computed. Another important problem is the interplay of washing out and CP-violation in a situation with several different varieties of super-heavy boson (more than one stage of GUT symmetry breaking, $m_X \neq m_{H_X}$, etc.). Under certain restrictions on α_X and m_X the heavier bosons may establish equilibrium for the lighter bosons⁴⁴, or the asymmetries from different sources may just add⁴⁷. Other calculational complications arise from possible deviations from the simple isotropic and homogeneous, monotonically expanding and cooling Universe which we have been considering. For example, there may

be significant inhomogeneities - primordial black holes (PBHs) - whose explosions at or after a time corresponding to a universal temperature of $\sim 10^{15}$ GeV might⁴³ contribute significantly to the net baryon asymmetry. (Conversely, a satisfactory computation of n_B/n_γ in a completely homogeneous and isotropic scenario might impose constraints on the spectrum of primordial black holes.) Related complications may arise from the observation⁴² (see Fig. 2) that no known particle interactions may have been in equilibrium at temperatures between 10^{13} and 10^{15} GeV. This means that particles need not have been in a Boltzmann distribution in this epoch, and that dissipative phenomena may have been important, as discussed in the next section. Finally, as discussed in Section 6 of this paper, it has recently been suggested²² in the context of grand unified monopole production early in the Universe that the Universe may have had a complicated thermal history during the grand unified phase transition. It may have supercooled in the symmetric phase down to a temperature $\ll 10^{15}$ GeV, and then been reheated by the latent heat released in a first order phase transition. This would complicate considerably the previous naïve discussion (see Fig. 2) of the expansion rate, interaction rates and thermal equilibrium. Because of all the complications mentioned here -- and doubtless others -- it seems reasonable to conclude that the GUT mechanism for generating the baryon asymmetry (10) is still in a qualitative stage, though the process of quantification is making significant advances.

5. NON EQUILIBRIUM IN THE VERY EARLY UNIVERSE?

It is interesting and perhaps significant that when the temperature of the Universe was between 10^{13} GeV and roughly 10^{15} GeV no "known" particle interactions may have been in equilibrium^{42,49}: during that epoch one has approximately

$$\frac{\text{mean free time}}{\text{expansion time}} \approx \frac{T}{\alpha^2} > 1 . \quad (37)$$

This means that particles need not have been Boltzmann distributed, and that dissipative phenomena²⁰ may have been significant. What might the primordial distribution of particles have been? In a Robertson-Walker-Friedman Universe, when the temperature was of order 10^{13} GeV, the horizon diameter $D_H = 4t$ would have been of the same order as the Planck distance $\approx 1/m_P$. In the absence of acausal effects -- which should, however, perhaps be expected in the context of quantum gravity -- it is difficult to imagine how coherence could have been established over distances $> D_H$ with the same ease as over distances $< D_H$. (It is difficult to imagine particles with a wavelength size larger than the horizon size.) In this case, one might expect⁴³ particles with short wavelengths $\lambda < 1/m_P$ or high

energies $E \gtrsim 0(\text{mp})$ to have a spectrum distorted by comparison with a naïve Boltzmann distribution. As an extreme case, one could even imagine a cut-off of the low-energy tail of the spectrum at $E < 0(\text{mp})$. If such an approximate cut-off occurred, one can estimate that $(1 \leftrightarrow 2)$ production processes would have been unlikely to fill in the spectrum at $T \gtrsim \text{few} \times 10^{16} \text{ GeV}$, while $(2 \leftrightarrow 3)$ interactions are unlikely⁴⁹ to have done the job before $T \approx 10^{16} \text{ GeV}$. Therefore, if the masses of the X bosons were significantly higher than the fashionable 10^{15} GeV , it is an open question whether their distributions were ever fully thermalized. In the absence of thermalization, various theorems^{17,40} about the absence of a net baryon number become inapplicable, and the computation of n_B/n_γ becomes more complicated. This complication is unlikely to have arisen if m_X is as low as the expected 10^{15} GeV , but could be significant if m_X were significantly larger.

Equation (37) implies that during this very early epoch particles would have had mean free paths large compared with the horizon size, and therefore that dissipative phenomena²⁰ may have occurred as illustrated in Fig. 8. Generically, the shear viscosity η of an imperfect fluid is⁵⁰ of order

$$\eta \approx T^4 \tau \quad (38a)$$

and taking the earlier estimates of the mean free time τ one gets²⁰

$$\eta \approx (10^2 \text{ to } 10^3) T^3 \quad (38b)$$

when the temperature exceeds 10^{15} GeV . Similarly, the thermal conductivity χ of an imperfect fluid is of order

$$\chi \sim T^3 \tau \quad (39a)$$

which becomes²⁰

$$\chi \approx (10^3 \text{ to } 10^4) T^2 \quad (39b)$$

in this very early epoch. By contrast, the bulk viscosity is zero for a fluid of ultrarelativistic particles. The grand unified viscosity (38) and thermal conductivity (39) may have played a role in smoothing out primordial inhomogeneities or, perhaps, correcting initial anisotropies as illustrated in Fig. 8. Because of causality,

Fig. 8 Illustration of particles with long mean free paths transferring energy and momentum and hence possibly giving rise to dissipation through viscosity and/or thermal conductivity.

their effects should be limited to within the size of a horizon: this may be sufficient²⁰ to damp down high frequency modes of oscillation²⁶ (radiative, rotational or compressional) which might otherwise have contained an energy density comparable to that in elementary particles. These dissipative effects also tend to homogenize the Universe locally preparatory to the generation of baryon number: in particular they may⁵¹ tend to hinder the production of primordial black holes $\leq$ the horizon size. However, viscous effects do not prevent the initiation and growth of very low frequency, large scale compressional inhomogeneities, which Press²¹ has speculated may have originated at the time of the grand unified phase transition and subsequently grown in intensity and evolved into galaxies. In view of the logarithmically long growth time between 10^{15} GeV and the epoch at which galaxy formation is conventionally initiated, this may be a promising way²¹ of avoiding the normal problems in explaining galaxy formation.

6. GRAND UNIFIED MONOPOLES

These appear¹¹ near the mass scale at which a primitive U(1) factor in the gauge group of the world gets absorbed into a simple group. For example, in the case of the minimal GUT SU(5): $SU(5) \rightarrow SU(3) \times SU(2) \times U(1)$ at around 10^{15} GeV, one expects to encounter topologically non-trivial and stable states (cf. Fig. 9a) called magnetic monopoles M with masses

$$m_M \approx \frac{m_X}{\alpha_X} \approx 10^{16} \text{ GeV} . \quad (40)$$

Fig. 9 Impressions of (a) Higgs fields in causally separated domains pointing in different (internal group) directions causing (b) a monopole to be formed where they meet: note that the value of the Higgs field goes to zero in the centre.

These monopoles may have been produced copiously in the very early Universe, and if so their topological stability may have caused them to survive to lower temperatures in unacceptably large numbers¹². It is convenient to introduce a dimensionless density for these objects which is

$$r \equiv n/T^3 \quad (41)$$

and is basically the monopole-to-photon ratio. From the success of the conventional calculations¹⁵ of nucleosynthesis at temperatures $T \sim 1$ MeV one deduces that monopoles did not dominate the energy density at that epoch, so that

$$r_{1 \text{ MeV}} < 10^{-19} . \quad (42a)$$

The absence of a collapse in the present Hubble expansion due to a high density of relic monopoles means that

$$r_{30\text{K}} < 10^{-25} . \quad (42b)$$

It has been estimated that conventional $M\bar{M}$ capture and annihilation in a homogeneous Universe could not have reduced r below 10^{-10} . The Universe now contains inhomogeneities (e.g. you, us and galaxies) but excess annihilation in these could not have explained the number (42a), and probably not the number (42b) either. We therefore conclude that either very few monopoles ($\leq$ Eq. (42)) were produced and/or the conventional capture and annihilation rates are not applicable.

The conventional picture^{12,22} of monopole production in the early Universe goes somewhat as follows. At very high temperatures the Higgs fields have zero vacuum expectation value and tend to run around in random directions as in Fig. 10a. At low temperatures the Higgs fields will congeal with a fixed, non-zero v.e.v. but in general pointing in different internal group directions in different domains of space, as illustrated in Fig. 9b. Where these domains come together there may be a mismatch with the Higgses pointing in different directions so that the lowest energy configuration has a zero field value in the centre, as in Fig. 9a. Such a "hole" is a topologically stable monopole¹¹ for which one must pay the energetic price (40).

A popular line of argument²² is to suggest that the number of "holes" is the same order⁵² as the number of causally separated domains at the temperature where the monopoles freeze out. To get an acceptably low number (42) one of them tries to reduce as far as possible the number of different domains by setting this freeze-out at the lowest possible temperature, postulating a first-order phase transition and strong supercooling²². When the phase

Fig. 10 Illustration of the Higgs field value as a function of temperature. (a) At high temperatures $T > T_c$, $|H|$ is expected to be zero. (b) For intermediate temperatures $T_c > T > T_G$ the minimum of the potential $V(|H|, T)$ is at $|H| \neq 0$, but the thermal fluctuations of $|H|$ are so large that it can pass over the central hump. (c) At low temperatures $|H| \neq 0$ and the Higgs field no longer fluctuates over the hump.

transition takes place the latent heat then reheats the Universe back to a temperature closer to 10^{15} GeV, after which it expands and cools in the normal way. Apart from the suppression of monopole production, an interesting feature⁵³ of this scenario is that the Universe expands exponentially during the supercooled epoch,

as the energy density and pressure are then dominated by the vacuum energy density. This has two⁵³ attractive implications, and one which is potentially catastrophic. One piece of good news is that the classic "simultaneity" problem is avoided because before this period of exponential expansion all the Universe we now see, and much more besides, was sufficiently close to be in causal contact and hence all expanding in the same way at the same time. Another piece of good news is that during this exponential expansion any initial curvature ($k \neq 0$ in the RWF solution (11)) would have got flattened out, rather like a balloon being blown up. This might explain why the present density of the Universe is so close to the critical density for eventual gravitational collapse. The bad news is that it is very difficult⁵³ to see how the transition to a broken phase of the GUT could ever have occurred. The basic problem is that while bubbles of asymmetric Universe expand at the speed of light, the Universe as a whole is expanding too fast for them to catch up. Unless this problem is solved, the scenario is doomed.

The necessity for a strongly first-order phase transition has in any case been criticized²⁵. The concept of a Higgs field pointing in any given direction is gauge-dependent, and the concept of domains "pointing in a fixed direction" is not on as solid a basis as in a ferromagnet. A monopole is of course a gauge-independent object, and the proposal²⁵ is to calculate their density assuming a second-order or weakly first-order transition and the validity of thermal equilibrium arguments. In a second-order phase transition the lowest energy state starts having a non-zero value of the Higgs field at the critical temperature T_c (see Fig. 10a). However, for some range of temperatures below that, the thermal fluctuations are sufficient to knock the Higgs field out of its minimum and over the central "hump" in the Higgs potential, as indicated in Fig. 10b. Therefore, in this range of temperatures it does not even really make sense to talk of a conserved topological quantum number, or of monopoles. At a temperature $T_G < T_c$ the thermal fluctuations are no longer sufficient to jiggle the Higgs, and it congeals into the non-zero valley of the Higgs potential, as in Fig. 10c. This Ginzburg^{54,12} temperature T_G is basically determined by the energy density in a correlation volume of radius $O(1/m_{\text{Higgs}})$ being less than the free energy density of order $\langle 0|H|0 \rangle^4$:

$$T_G \approx (\text{geometrical factor}) \times \frac{1}{m_H^3} \times \langle 0|H|0 \rangle^4. \quad (43)$$

Down to this temperature the Higgs field is still moving around sufficiently fast

$$\tau \sim \frac{1}{m_H} < t \sim T_G^{-2} \quad (44)$$

that thermal equilibrium might be relevant and one would expect

$$r \approx 0 \left(\exp - \frac{m_M}{T_G} \right) = 0 \left(\exp - C \frac{m_H}{m_X} \right), \quad (45)$$

where²⁵ in the minimal SU(5) model

$$C \approx 0(30) . \quad (46)$$

Bearing in mind the uncertainties in estimating C and the possibility that $m_H/m_X > 1$ (dynamical symmetry breaking?) this argument could well give r low enough to be compatible with the limits (42). Some people dislike²⁴ this argument on the grounds that since it gives much less than one monopole per causally connected volume, it seems to require the existence of some acausal correlation in the Higgs field. However, as remarked above, gauge invariance may require some of the arguments about causally unrelated domains to be relaxed.

A final remark about the present density of monopoles concerns the classic calculations of capture and annihilation rates which assumed no correlations in the spatial distributions of monopoles and antimonopoles. It has²³ been pointed out that in some GUTs some of the monopoles may be bound together by a string of Z^0 magnetic flux which enhances the annihilation rate. An example⁵⁵ of a model where this may play a role in suppressing monopoles is in SO(10) broken down through the chain $SO(10) \rightarrow SO(6) \times SO(4) \rightarrow SU(3) \times SU(2) \rightarrow U(1) \rightarrow SU(3) \times U(1)$. For yet another mechanism for suppressing monopoles, see Ref. 56.

Our present feeling is that while none of the above arguments provides a totally convincing mechanism for disposing of the grand unified monopole problem, none of us proposes to lose any sleep worrying about it. But it would be worth improving the present experimental upper limits⁵⁷ on the relic monopole density by a few orders of magnitude: after all, perhaps ...?

7. CONCLUSIONS

There is indeed a flourishing cosmic connection between grand unification and cosmology, which has several aspects that are more or less strong. The principal points concern:

Neutrino masses and numbers

There are phenomenological reasons in GUTs to expect only three or four "light" neutrinos, and theoretical reasons to expect that they might have masses up to a few electron volts. The success of nucleosynthesis calculations seems to require three or four "light" neutrinos, and masses up to $O(100)$ eV are allowed by cosmology. Such massive "light" neutrinos may even help alleviate the missing light problem of clusters of galaxies.

Baryon number generation

GUTs provide a framework and a qualitative mechanism for generating the baryon asymmetry observed in the Universe. For it to work, the masses of the superheavy bosons should not be too small, nor their coupling constants too large. Present GUTs are more or less consistent with these constraints, but it seems that a larger GUT than the minimal SU(5) may be necessary if one is to generate a large enough number.

Grand unified monopoles

First estimates of their production rates early in the Universe gave values much above the allowed limits. Ideas for getting rid of them have been proposed, one of which has other interesting cosmological implications (simultaneity, asymptotic flatness) but also has serious problems. It may well be that the grand unified monopole problem has a relatively banal solution. In the meantime: why not look for them?

Acknowledgements

We would like to thank F.A. Bais, J.D. Barrow, B.J. Carr, G. Cocconi, M.B. Einhorn, A. Guth, S. Rudaz, G. Steigman, M.S. Turner, S.-H.H. Tye, S. Weinberg and S. Wolfram for interesting conversations about the topics discussed here.

REFERENCES

1. J.C. Pati and A. Salam, Phys. Rev. Letters 31 (1973) 661 and Phys. Rev. D8 (1973) 1240 were the first proposers of grand unified gauge theories, but the type discussed in this paper was first proposed by H. Georgi and S.L. Glashow, Phys. Rev. Letters 32 (1974) 438.
2. J. Ellis, M.K. Gaillard, D.V. Nanopoulos and S. Rudaz - LAPP preprint TH-14/CERN preprint TH.2833 (1980) and references therein.

3. J. Learned, F. Reines and A. Soni, Phys. Rev. Letters 43 (1979) 907 and H.R. Steinberg, private communication (1980), quote limits on the nucleon lifetime of 2 and 3×10^{30} years, respectively, if the decay modes resemble those expected in the SU(5) model of Ref. 1.
4. H. Georgi, H.R. Quinn and S. Weinberg, Phys. Rev. Letters 33 (1974) 451.
5. A.J. Buras, J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Nuclear Phys. B135 (1978) 66.
6. T. Goldman and D.A. Ross, Phys. Letters 84B (1979) 208 and Caltech preprint CALT-68-759 (1980).
7. W. Marciano, Phys. Rev. D20 (1979) 274 and Rockefeller Univ. preprint COO-2232B-195 (1980).
8. L. Maiani, G. Parisi and R. Petronzio, Nuclear Phys. 136B (1978) 115 and references therein.
9. J. Ellis, M.K. Gaillard, L. Maiani and B. Zumino, LAPP preprint TH-15/CERN preprint TH-2841 (1980), contribution to these proceedings;
J. Ellis, M.K. Gaillard and B. Zumino, CERN preprint TH.2842/LAPP preprint TH-16 (1980) and references therein.
10. G. Steigman, Ann. Rev. Astron. and Astrophys. 14 (1976) 339.
11. G. 't Hooft, Nuclear Phys. B79 (1974) 276;
A.M. Polyakov, JETP Letters 20 (1974) 194.
12. J.P. Preskill, Phys. Rev. Letters 43 (1979) 1365; see also Ya.B. Zeldovich and M.Y. Khlopov, Phys. Letters 79B (1979) 239.
13. M.S. Chanowitz, J. Ellis and M.K. Gaillard, Nuclear Phys. B128 (1977) 506.
14. D.V. Nanopoulos and D.A. Ross, Nuclear Phys. B157 (1979) 273.
15. J. Yang, D.N. Schramm, G. Steigman and R.T. Rood, Ap.J. 227 (1979) 697; and
G. Steigman, contribution to these proceedings.
16. A.D. Sakharov, Pis'ma Zh.Eksp.Teor.Fiz. 5 (1967) 32;
A. Yu. Ignatiev, N.V. Krosnikov, V.A. Kuzmin and A.N. Tavkhelidze, Phys. Letters 76B (1978) 436;
M. Yoshimura, Phys. Rev. Letters 41 (1978) 381; 42 (1979) 746(E).
S. Dimopoulos and L. Susskind, Phys. Rev. D18 (1978) 4500;
D. Toussaint, S.B. Treiman, F. Wilczek and A. Zee, Phys. Rev. D19 (1979) 1036;
J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Phys. Letters 80B (1979) 360, 82B (1979) 464(E),
S. Weinberg, Phys. Rev. Letters 42 (1979) 850.
A.D. Sakharov, Zh.Eksp.Teor.Fiz. 76 (1979) 1172;
S. Dimopoulos and L. Susskind, Phys. Letters 81B (1979) 416;
M. Yoshimura, Phys. Letters 88B (1979) 294.
17. E.W. Kolb and S. Wolfram, Phys. Letters 91B (1980) 217, and Caltech preprint OAP-579/CALT-68-754 (1979);
J.N. Fry, K.A. Olive and M.S. Turner, Enrico Fermi Institute preprint 80-07 (1980).

18. D.V. Nanopoulos and S. Weinberg, Phys. Rev. D20 (1979) 2484;
S. Barr, G. Segrè and A. Weldon, Phys. Rev. D20 (1979) 2494;
see also
J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Ref. 16.
19. Another cosmological topic worth mentioning is the lepton number of the Universe. If lepton number L is violated as well as B , then one might expect a net n_L comparable with n_B . In many GUTs however, $B-L$ is conserved or very nearly so, and in these models neither n_B nor n_L has good reason to be small unless the net n_B and n_L were both initially zero (a symmetric Universe). In more general theories, where $(B-L)$ is violated, this problem does not arise. For discussions, see
S. Dimopoulos and G. Feinberg, Phys. Rev. D20 (1979) 1283;
D.V. Nanopoulos, D. Sutherland and A. Yildiz, Harvard Univ. preprint HUTP 79/A038 (1979), (to be published in Lettere al Nuovo Cimento).
D.N. Schramm and G. Steigman, Phys. Letters 87B (1979) 141.
20. J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Phys. Letters 90B (1980) 253.
21. W.H. Press, Physica Scripta 21 (1980) 702.
22. A. Guth and S.-H.H. Tye, Phys. Rev. Letters 44 (1980) 631;
M.B. Einhorn, D.L. Stein and D. Toussaint, Univ. of Michigan preprint UM HE 80-1 (1980).
23. G. Lazarides and Q. Shafi, CERN preprint TH-2821 (1980).
24. For reviews of the cosmological monopole problem, see
M.B. Einhorn, contribution to these proceedings and
S.-H.H. Tye, Cornell University preprint "Monopoles, Phase Transitions and the Early Universe" (1980).
25. F.A. Bais, S. Rudaz, CERN preprint TH in preparation (1980).
26. S. Weinberg, "Gravitation and Cosmology" (Wiley, New York, 1972).
27. For GUTs based on bigger groups than $SU(5)$, see
R. Barbieri, D.V. Nanopoulos, J.C. Pati, B. Stech, contributions to these proceedings.
28. J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Phys. Letters 80B (1979) 320.
29. E. Gildener, Phys. Rev. D14 (1976) 1667; see also
J. Ellis, M.K. Gaillard, A. Peterman and C.T. Sachrajda, Nuclear Phys. B164 (1980) 253, and references therein.
30. P. Binétruy and T. Schücker, CERN preprint TH-2857 (1980).
31. K.T. Mahanthappa and M.A. Sher, Colorado preprint COLO-HEP 13 (1979).
32. L. Sulak and E. Bellotti, contributions to these proceedings.
33. U. Amaldi, Proc. Neutrino 1979, Bergen, eds A. Haatuft and C. Jarlskog (Univ. of Bergen, 1979), p. 376;
F. Dydak, Proc. EPS Int. Conf. on High Energy Phys., Geneva (1979) (CERN, 1979) p. 25;
K. Winter, Proc. 1979 Int. Symp. on Lepton and Photon Interactions at High Energies, eds T.B.W. Kirk and H.D.I. Abarbanel (FNAL, Batavia, 1980), p. 258.

34. R. Barbieri, CERN preprint TH.2850 (1980), contribution to these proceedings.
  35. For a recent review, see
J.N. Bahcall, contribution to Int. Conf. on Astrophysics and Elementary Particles, Common Problems, Rome (1980).
  36. Since this talk was given more information has become available about possible neutrino masses and oscillations:
F. Reines, H.W. Sobel and E. Pasierb, U.C. Irvine preprint "Evidence for Neutrino Instability" (1980).
E.F. Tretyakov et al., reportedly see a ν_e mass between 14 and 46 eV, while the ν_e/ν_μ induced event ratio in the 1979 CERN beam dump experiments was not unity; see also
H. Wachsmuth, Proc. 1979 Int. Symp. on Lepton and Photon Interactions at High Energies, eds T.B.W. Kirk and H.D.I. Abarbanel (FNAL, Batavia, 1980), p.541.
  37. G. Steigman, contribution to these proceedings.
  38. R.L. Golden et al., Phys. Rev. Letters 43 (1979) 1196.
  39. R.L. Omnès, Phys. Rev. Letters 23 (1969) 38 and Astron. and Astrophys. 10 (1971) 228.
  40. See in particular D. Toussaint et al., and S. Weinberg, Ref. 16.
  41. See E.W. Kolb and S. Wolfram, Ref. 17.
  42. See J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Ref. 16.
  43. See the second paper of S. Dimopoulos and L. Susskind, Ref. 16.
  44. D.V. Nanopoulos and S. Weinberg, Ref. 18.
  45. For an estimate of the dilution of the baryon number generated due to subsequent entropy generation in a homogeneous and isotropic Universe, see
G. Steigman, K.A. Olive and D.N. Schramm, Phys. Rev. Letters 43 (1979) 239.
  46. D.V. Nanopoulos, Phys. Letters 91B (1980) 67, discusses anthropocentric constraints on the baryon number asymmetry.
  47. S.B. Treiman and F. Wilczek, Princeton preprint "Thermalization of Baryon Asymmetry" (1980).
  48. M.S. Turner and D.N. Schramm, Nature 279 (1979) 303;
M.S. Turner, Nature 281 (1979) 549, and Phys. Letters 89B (1979) 155;
B.J. Carr and M.S. Turner, Enrico Fermi Institute preprint 80-09 (1980); see also D. Toussaint et al., Ref. 16.
  49. J. Ellis and G. Steigman, Phys. Letters 89B (1980) 186;
J.J. Aly, Mon.Not.Roy.Astr.Soc. 189 (1979) 479 points out that the Universe could have been very chaotic initially, as long as it smoothed out before the baryon asymmetry¹⁵ was generated.
  50. J.M. Stewart, "Non-equilibrium Relativistic Kinetic Theory", Lecture Notes in Physics, Vol. 10 (Springer, Berlin, 1971);
C. Marle, Ann.Inst.H. Poincaré 10 (1969) 67, 127.
- It should be noted that the numerical factors of order 1 in the formulae for η and χ used in Ref. 20 in fact depend on the precise angular form of the cross-section σ being used, and that some formulae given in the literature are not exactly correct. See for some clarification

N. Straumann, *Helvetica Phys. Acta* 49 (1976) 269.

Also the exponent in formula (24) of Ref. 20 should be a factor of 2 larger. We thank N. Straumann for correspondence and discussions on these points.

It has recently been emphasized by H. Sato, Kyoto University preprint RIFP-390 (1980) that the conventional viscosity approximation for discussing dissipative effects is only valid when $\tau < t$. In our case (38) this is only strictly valid when the temperature is in the lower part of the range $10^{19} \text{ GeV} > T > 10^{15} \text{ GeV}$. This means that the results of Ref. 20 can only be regarded as qualitative. However, we believe that the basic physical picture of Fig. 8 remains valid even if $\tau > t$: the point is that particles are then able to "leak" out of an inhomogeneity.

51. A.F. Grillo, Frascati preprint LNF-80/21(P) (1980).
52. T.W.B. Kibble, *J. Phys. A* 9 (1976) 1387 and
M.B. Einhorn, D.L. Stein and D. Toussaint, Ref. 22.
53. A. Guth, private communication (1980).
For a recent review of cosmological phase transitions, their problems and implications, see
T.W.B. Kibble, Imperial College London preprint ICTP/79-80/23 (1980).
For discussions of the nature of the grand unified phase transition, see
P. Ginsparg, CERN Saclay preprint DPh-T/80/27 (1980);
M. Daniel and C.E. Vayonakis, CERN preprint TH-2860 (1980).
54. V.L. Ginzburg, *Fiz.Teor.Tela* 2 (1960) 2031.
55. G. Lazarides, M. Magg and Q. Shafi, CERN preprint 2856 (1980).
56. P. Langacker and S.-Y. Pi, Princeton Institute for Advanced Study preprint "Magnetic Monopoles in Grand Unified Theories" (1980) propose to reduce the cosmological density of monopoles by having electromagnetic gauge invariance spontaneously broken at temperatures $T \geq 1 \text{ TeV}$. Open problems in their scenario include its successful embedding in a grand unified theory, and the computation of the net charge density of the Universe due to C-, CP- and Q-violating effects out of thermal equilibrium. However, all other GUTs conserve Q and hence do not explain why the present charge density of the Universe is so small.
57. G. Barbiellini et al., DESY preprint "Quarks and Monopoles at LEP" (1980) contains a review of monopole searches.