PHYSICS REPORTS (Review Section of Physics Letters) 67, No. 1 (1980) 183–199. North-Holland Publishing Company

Some Implications of a Cosmological Phase Transition

T.W.B. KIBBLE

Blackett Laboratory, Imperial College, London SW7 2BZ, U.K.

Abstract:
I discuss the reasons for believing that phase transitions occurred in the very early history of the universe, and the topological structures that may have been generated thereby – in particular, the strings and monopoles. The aim is mainly pedagogical, with special emphasis on unsolved problems.

1. Introduction

One of the most intriguing recent developments in particle physics is its interaction with cosmology. At least tentative answers have been found to several of the basic questions of cosmology, and cosmology in turn has been shown to yield rather stringent limits on some of the parameters of particle physics.

Many features of the early universe have analogues in condensed matter physics. It is my hope, in giving this talk, that condensed matter physicists may be encouraged to take an interest in what was, after all, the largest condensed matter system ever. Perhaps they can see how to resolve some of the outstanding problems.

I shall assume for most of this talk that the standard cosmological picture, the hot big bang, is correct. The most basic question of all – why the big bang? – remains unanswered. Moreover, we have no real explanation of the initial state. The simplest assumption is that soon after the Planck time ($t = 10^{-44}$ s) the universe was in a state of thermal equilibrium at a temperature not far below the Planck mass (10^{28} eV). Why regions that can have had no previous causal contact should have been in equilibrium is quite unclear, and for the moment must be taken as an axiom. To trace evolution back so far is audacious; to go further is, at present, impossible, for beyond the Planck mass quantum gravity effects must dominate, and as yet we have no understanding of such an environment.

Given the initial state, however, we can do quite well. There is now at least a plausible explanation for the generation of baryon number [1], yielding a numerical estimate of the baryon-to-photon number ratio that is at least in the right ballpark to fit the experimental value $10^{-9\pm1}$.

Going to a much later stage in the evolutionary history, we have a good understanding of nucleosynthesis and, in particular, the helium-to-hydrogen ratio [2]. The requirement that our fundamental theory should not spoil the agreement here gives severe limits on total numbers of particles of various types which in turn restrict the fundamental theory.

Similarly, the observed isotropy of the 3 K background radiation [3] gives strong constraints on any mechanism that generates inhomogeneities in the early universe. Some inhomogeneities are, however, badly needed, to allow us to explain the formation of galaxies which presumably evolved by gravitational condensation from an earlier near-homogenous state [4].

At first sight it may seem that discussions of the state of the universe at far above nuclear densities involve such enormous extrapolation of our theories that we can have little confidence in the results. This may be true but there is also some reason to assert that it is precisely in this situation that our theories are most trustworthy. If one believes the currently popular dogma of asymptotic freedom – and avoids a theory with too many fermions – then at the temperatures involved, all interactions should be effectively weak, so that low-order perturbation theory should be a reliable guide. Moreover, in contrast to quantum chromodynamics at laboratory energies, our theory has none of the complexities of confinement. At such densities quarks can never get far enough apart to form individual hadrons.

I shall begin (in sections 2 and 3) by reviewing, briefly, spontaneously broken gauge theories and the early history of the universe. Then, in section 4, I will try to explain why phase transitions should have occurred and what happens at such transitions. In particular, we are interested in topological singularities of various kinds – domain walls, strings or monopoles. Their classification will be discussed in section 5. Strings, and how they evolve in time are the subject of section 6. In particular, I shall examine their possible relevance to the problem of galaxy formation. Section 7 will deal with monopoles, especially possible ways of avoiding the disaster of over-population by monopoles. One specific possibility involves the assumption that the phase transitions may be strongly first-order. Finally, in section 8, I examine some implications of the effective cosmological term induced by symmetry restoration.

2. Spontaneously broken gauge theories

There is now very good reason to believe that elementary particles and their interactions are described by a spontaneously broken gauge theory. It is a very attractive hypothesis that the weak, electromagnetic and strong interactions are all united at extremely high energies in a grand unified theory [5, 6], described by a group such as SU(5). Then there are at least two distinct stages in the spontaneous symmetry breaking. At a grand unification mass, around 10^{15} GeV, we have

$$\mathrm{SU}(5) \to \underset{\text{colour}}{\mathrm{SU}(3)} \times \underset{\text{Weinberg–Salam}}{\mathrm{SU}(2) \times \mathrm{U}(1)}$$

while at the much lower scale of 10^2 GeV, the Weinberg–Salam symmetry is broken:

$$\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1) \to \underset{\text{colour}}{\mathrm{SU}(3)} \times \underset{\text{electromagnetism}}{\mathrm{U}(1)}\,.$$

This is of course only the simplest of many possible scenarios. Which one nature chooses, we do not know, but the general idea is probably correct.

The grand unification mass is obtained from a renormalization-group calculation which shows that all three coupling constants, associated with the three separate groups SU(3), SU(2) and U(1) come together at this mass at a value of about [6]

$$\alpha = g^2/4\pi \simeq \tfrac{1}{50}. \qquad (1)$$

I shall suppose for pedagogical reasons that the spontaneous symmetry breaking mechanism is via the acquisition of a non-zero vacuum expectation value by a Higgs scaler field. However, this is an inessential assumption. It would really make little difference to our arguments if this fundamental scalar field were replaced by a composite structure, like a Cooper pair [7].

Moreover, I shall ignore fermions altogether. Provided the number of species is reasonably small, they have little effect on the qualitative behaviour of the universe in its early stages. Too many fermions, however, would spoil the asymptotic freedom, changing the sign of the renormalization group function $\beta(g)$ so that $g \to \infty$ rather than tending to zero at high energies [8]. I shall assume that such a catastrophe does not occur. (If it does, all bets are off: we can calculate nothing at all about the early universe.)

Let us consider a typical gauge theory, described by the lagrangian density

$$\mathscr{L} = \tfrac{1}{8}\mathrm{tr}\, F_{\mu\nu}F^{\mu\nu} + \tfrac{1}{2}\mathrm{D}_\mu\phi \cdot \mathrm{D}^\mu\phi - U(\phi). \tag{2}$$

Here ϕ belongs to a representation of the gauge group G with real anti-symmetric generators T_a satisfying

$$\mathrm{tr}\, T_a T_b = -\tfrac{1}{2}\delta_{ab}.$$

The covariant derivative of ϕ is

$$\mathrm{D}_\mu\phi = \partial_\mu\phi - gA_\mu\phi,$$

where $A_\mu = A^a_\mu T_a$, and g is the gauge coupling constant. Also

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu + g[A_\mu, A_\nu].$$

Finally, $U(\phi)$ is a polynomial in ϕ of degree 4 (to ensure renormalizability) which is assumed invariant under G.

A pedagogically convenient, though physically unrealistic, example is obtained by taking ϕ to belong to the N-dimensional vector representation of O(N), with

$$U(\phi) = \tfrac{1}{8}h^2(\phi^2 - \eta^2)^2, \tag{3}$$

where h^2 is the Higgs coupling constant [9].

Because of the form of this potential, with a maximum at $\phi = 0$, ϕ tends to acquire a nonzero vacuum expectation value $\langle\phi\rangle$. This may be calculated by minimizing the effective potential $V(\phi)$, which is simply the minimum free-energy density for states in which $\langle\phi\rangle$ has a value equal to the argument of V. We have, in a suitably chosen gauge,

$$V(\phi) = U(\phi) + \hbar(\cdots)$$

so that in the tree approximation

$$\langle\phi\rangle^2 = \eta^2.$$

This equation does not fix the direction of $\langle\phi\rangle$, only its magnitude. We have, in fact, a degenerate set of vacuum states, labelled by $N-1$ angle variables.

In the general case, the set of degenerate vacua forms a quotient space M. If H is the subgroup of G which leaves a particular $\langle\phi\rangle$ unaltered, then $\mathrm{M}=\mathrm{G/H}$, the set of all right cosets of H in G. In our example, H is $\mathrm{O}(N-1)$ and

$$\mathrm{M}=\mathrm{O}(N)/\mathrm{O}(N-1)=\mathrm{S}^{N-1},$$

the $(N-1)$-dimensional sphere (in N-dimensional space).

The magnitude of $\langle\phi\rangle$ fixes the masses of the various particles in the theory. For instance, in the example we have one Higgs particle with mass $m_{\mathrm{S}}=h\eta$ and $(N-1)$ massive vectors with $m_{\mathrm{V}}=g\eta$, together with $\frac{1}{2}(N-1)(N-2)$ massless vectors, the gauge bosons associated with the unbroken symmetry subgroup H [9, 10].

Note that in the case $N=2$, this is the Landau–Ginzburg model [11]. Of course, the correlation length which is the chief determinant of the size of fluctuations in the magnitude of ϕ is $\xi=1/m_{\mathrm{S}}$, while the penetration depth is $\lambda=1/m_{\mathrm{V}}$. The model will correspond to a type 1 superconductor if $h<g$ or $m_{\mathrm{S}}<m_{\mathrm{V}}$ and type 2 in the contrary case. It is interesting to note that some supersymmetric theories have $h=g$ and so fall exactly on the boundary between the two.

To study the behaviour of these theories at high temperature, we need to compute the one-loop correction to the effective potential. The leading terms at high T are [12]

$$V(\phi)=U(\phi)-\mathcal{N}\frac{\pi^2}{90}T^4+\frac{1}{24}\mathcal{M}^2(\phi)T^2+\mathcal{O}(T^1), \tag{4}$$

where $\mathcal{N}$ is the total number of distinct helicity states of low-mass particles (i.e. those with $m \ll T$), counting fermions with a factor 7/8, while $\mathcal{M}^2$ is the sum of squared masses of boson helicity states plus half that for fermions. (Strictly speaking, V is gauge-dependent, but at least for our present purposes we can ignore this complication.) Though $\mathcal{N}$ is a constant, $\mathcal{M}^2$ depends on the vacuum expectation value ϕ because the masses do. For instance, in our example

$$\mathcal{N}=N^2$$

$$\mathcal{M}^2(\phi)=\tfrac{1}{2}Nh^2(\phi^2-\eta^2)+h^2\phi^2+3(N-1)g^2\phi^2.$$

Since $\mathcal{M}^2$ generally contains a positive term in ϕ^2 there will be a critical temperature T_{c} above which $\phi=0$ is a minimum of V. In the example

$$T_{\mathrm{c}}^2=\frac{12h^2\eta^2}{(N+2)h^2+6(N-1)g^2}$$

and more generally one finds (as a rough order-of-magnitude estimate)

$$T_{\mathrm{c}}\sim\eta\sim m_{\mathrm{V}}/g.$$

Then for $T < T_c$, we have the ordered phase with

$$\langle\phi\rangle^2 = \eta^2(1 - T^2/T_c^2) \tag{5}$$

while for $T > T_c$, $\langle\phi\rangle = 0$, we are in the symmetric phase.

It is possible when there are several coupling constants to so arrange matters that there are several phase transitions, sometimes going the "wrong" way [13] – more symmetry at lower temperature – but the situation described above may be regarded as the norm.

3. The early universe

Let $R(t)$ be the radius of a spherical volume expanding with the general expansion of the universe (a "comoving" sphere). Assuming isotropy and homogeneity, i.e. the standard Friedmann–Robertson–Walker universe, its time development is governed by Einstein's equation [14],

$$(\dot{R}/R)^2 = \tfrac{8}{3}\pi G\rho - K/R^2 + \Lambda \tag{6}$$

where ρ is the energy density and K and Λ are constants. For the moment let us set the cosmological constant Λ equal to zero, which is indeed the experimental value to a good approximation, at least in the present phase of the universe [15]. I shall return to the question of whether it had a different value in earlier phases in section 8.

It is interesting to rewrite (6) in the form of an energy conservation equation for a particle on the surface of our sphere [14]:

$$\tfrac{1}{2}\dot{R}^2 - G\rho(\tfrac{4}{3}\pi R^3)/R = -\tfrac{1}{2}K. \tag{7}$$

We see then that $K > 0$ corresponds to a bound orbit. This is the case of a closed universe that will eventually reverse its expansion and contract to a new singularity. Similarly the unbound case $K \leq 0$ corresponds to an open universe that will expand for ever.

In any case, at early times ρ increases as $R \to 0$ like R^{-4}, so that K becomes relatively unimportant. In fact, although this curvature term in (6) increases as $R \to 0$, it does so less rapidly than the density term, so that flat space becomes a better approximation.

Let us also assume that in the very early universe we have thermal equilibrium at a temperature T much larger than all particle masses. (We ignore the alternative scenario of Hagedorn [16] which envisages an exponentially rising spectrum of particle masses and a corresponding upper limit to the temperature.) Under these conditions the matter may be treated as an ideal relativistic gas undergoing adiabatic expansion. The density is then given by

$$\rho = \frac{\pi^2}{30}\mathcal{N}T^4,$$

so that from (7)

$$R \propto T^{-1} \propto t^{1/2}.$$

It is useful to introduce the Planck mass

$$m_P = 1/G^{1/2} = 1.22 \times 10^{28}\,\text{eV}$$

in terms of which the time-temperature relation may be written

$$tT^2 = \frac{0.3}{\sqrt{\mathcal{N}}} m_P = \frac{2.4}{\sqrt{\mathcal{N}}}\,\text{s MeV}^2. \tag{8}$$

The number $\mathcal{N}$ is of order 10^2. In the simplest SU(5) grand unified theory $\mathcal{N} = 160.75$.

For each helicity state of each bosonic species, the number density is the same,

$$n_s = \frac{\zeta(3)}{\pi^2} T^3.$$

(For fermions we have an extra factor $\frac{3}{4}$.) Thus at any time a sphere whose radius is the thermal wavelength $1/T$ would contain approximately one particle of each species. Since all cross sections have a similar value,

$$\sigma \sim \alpha^2/T^2,$$

where α is given by (1), the mean free path λ is

$$\lambda = 1/n\sigma \sim 1/\mathcal{N}\alpha^2 T.$$

But $\mathcal{N}\alpha^2 \simeq 1/15$, so λ is large compared to the thermal wavelength, and *a fortiori* to the average interparticle spacing. This, of course, helps to justify the ideal-gas approximation.

It may be useful to provide a table of the most significant events in the early history of the universe, starting from the Planck time. For the purposes of constructing the table, which also lists the values of $\mathcal{N}$ at different epochs, I have assumed the simplest possible grand unified model. The two phase transitions are marked by asterisks. Before the first, at $T = T_{GU} \sim 10^{15}$ GeV, all particle states count in $\mathcal{N}$. Below it, we must exclude those particles that acquire masses of order T_{GU} which are no longer present in significant numbers. In fact, it is the decay of some of these particles that is thought to yield the baryon number asymmetry [1].

After the Weinberg–Salam transition at $T = T_{WS} \sim 10^2$ GeV, $\mathcal{N}$ drops once more: it now includes only the quarks and gluons associated with SU(3), the leptons and the photon. The ensuing hadronic era is in some ways the one of which we know least. During it, the confinement mechanism must come into play, as we pass from a quark–gluon soup to a system of separate hadrons. As the temperature falls below 1 GeV (when the universe has roughly nuclear density) the nucleon pairs annihilate, leaving only the small baryon excess. By the time we reach 100 MeV the only particles contributing to $\mathcal{N}$ are the leptons and photon. Finally when the muons and electron pairs have disappeared we are left with just photons and neutrinos, giving $\mathcal{N} = 7.25$. (Effectively, however, it would be less than this, because the neutrinos go out of equilibrium before the photons and end up with a lower temperature [17].)

The number of neutrinos is very important in determining the rate of expansion at this epoch [18], which affects the temperature T at which the neutron-to-proton ratio "freezes out" – at a value

Table 1

t (s)	T (eV)	R/R_{now}	$\mathcal{N}$	
10^{-44}	10^{28}	10^{-32}		Planck time
			160.75	
10^{-37}	10^{24}	10^{-28}	******	GU
			106.75	
10^{-11}	10^{11}	10^{-15}	******	WS
			96.75	
10^{-7}	10^{9}	10^{-13}	⋮?	N pairs ↘
			14.25	
10^{-4}	10^{8}	10^{-12}	——	$\mu^{\pm}$↘
			10.75	
1	10^{6}	10^{-10}	——	$e^{\pm}$↘
			7.25	
10^{13}	1	10^{-3}	(effect. ~5)	recombination
10^{18}	3 K	1		present

$\exp(-\Delta m/T)$ of about 1/3. More neutrinos mean faster expansion, and hence more surviving neutrons, which in turn means more helium. Roughly, each extra neutrino species would add 1% to the helium abundance of around 25%. Astrophysicists are sufficiently confident of the validity of their helium-production scenario to assert that there can be, in all, no more than four or perhaps five species of low-mass neutrinos. There are, however, some possible, though unattractive, avenues of escape: for example, more neutrinos could be accommodated if there were a large asymmetry between neutrinos and antineutrinos [19, 2].

The last important epoch in cosmic history is the time of recombination of hydrogen, when the universe is about a million years old. It is only after this point that galaxies can start to form. I shall come back to this episode in section 6.

4. Structure at the phase transitions

Let us assume for the moment, in conformity with the simple theory of section 2, that the phase transitions are second-order. As T falls and passes through T_c, ϕ will tend to acquire a nonzero expectation value. But (5) fixes only the magnitude of $\langle\phi\rangle$; its direction is arbitrary. The situation is analogous to that of a perfectly isotropic ferromagnetic cooled through its Curie point. It must acquire nonzero magnetization but the magnetization direction is arbitrary, determined in practice by any small external field or, in the absence of such fields by the random fluctuations.

In the same way the universe must choose a direction for $\langle\phi\rangle$, i.e. a point on the manifold M of equivalent vacua. The choice is random, and may be different in different regions of space. Indeed, there can surely be no correlation extending beyond the current "horizon", at a distance ct. More remote parts of the universe can have no prior causal contact, at any rate in the conventional picture. In any event, the expected range of correlations is shorter than this, for reasons I shall try to explain [20].

Just below the critical temperature, the system is still subject to large fluctuations, large enough to bring it back to $\langle\phi\rangle = 0$. Such fluctuations are probable so long as

$$\xi^3 \Delta f \leqslant T, \tag{9}$$

where ξ is the correlation length and Δf the difference in free energy density between the two phases. (Here and below, we generally ignore all factors of order unity.) The temperature at which equality holds in (9) is the Ginzburg temperature T_G, given approximately by [21]

$$(T_c - T_G)/T_c \simeq h^2.$$

For weak coupling, therefore, T_G is not far below T_c.

The scale of the initial structure formed by spatial variations in $\langle\phi\rangle$ is thus the correlation length ξ_G at the Ginzburg temperature, which is roughly [20]

$$\xi_G \simeq 1/h^2 T_c \simeq 1/hm_S. \tag{10}$$

Of course, much of this structure will rapidly disappear. For energetic reasons $\langle\phi\rangle$ will tend towards spatial uniformity unless prevented from so doing by trapped singularities of some kind. Thus the scale on which $\langle\phi\rangle$ varies will certainly grow, while the correlation length ξ falls to its zero-temperature value $\xi_0 \simeq 1/m_S$.

Before proceeding, it may be well to remark that this argument could be wrong for phase transitions at very high temperature, close to the Planck mass [22]. Formally, the correlation length ξ should become infinite at T_c, but this, of course, would mean that it is growing faster than the velocity of light. Because the universe goes through the transition at a finite rate, it cannot actually achieve an infinite correlation length. Instead, its growth is effectively cut off at the point, say ξ'_G, where $d\xi/dt = 1$. Thereafter, it may continue to grow no faster than this until it intersects the falling curve of ξ after the transition, but this further growth is, in fact, negligible. A straightforward calculation of the cut off point yields

$$\xi'_G \simeq \left(\frac{m_P}{\sqrt{\mathcal{N}} m_S^2 T_c^2}\right)^{1/3},$$

which could, in principle, be less than ξ_G, but only if m_P/T_c were close to unity. So for our present purposes it seems adequate to take ξ_G as the quantity that defines the initial scale of structure in $\langle\phi\rangle$.

As remarked above, much of this structure will rapidly disappear, with $\langle\phi\rangle$ tending to uniformity except where singularities are trapped. The possible types of singularities are governed by the topology of the manifold M of degenerate vacua.

First of all, if M contained two or more disconnected pieces, corresponding to spontaneous breaking of a discrete symmetry, we might have domain walls [23]. An example of a model with this behaviour is the model of section 2 with $N = 1$ (and no gauge fields) which has two vacuum states with $\langle\phi\rangle = \pm\eta$ (at $T = 0$). In such a case (once T has fallen well below T_c) there would be walls separating the regions with $\langle\phi\rangle = \eta$ from these with $\langle\phi\rangle = -\eta$. The thin wall where $\langle\phi\rangle \sim 0$ cannot be eliminated. Its width is of order ξ_0 so the mass per unit area of wall is approximately

$$\xi_0 \Delta f \sim h\eta^3. \tag{11}$$

As pointed out by Zel'dovich Kobzarev and Okun [23], the existence of such walls can easily be ruled out. They are so massive – even for $\eta \sim 100$ GeV – that just one such wall stretched across the universe would be over 10^8 times more massive than all other known matter. Its gravitational effect would introduce an impossibly large anisotropy into the 3 K blackbody radiation. We can, therefore, confidently rule out any model that exhibits spontaneous breaking of a discrete symmetry, unless accompanied by an explicit breaking that would preferentially select one type of domain rather than the other.

This is a remarkable constraint on our freedom to construct fundamental particle theories. It would rule out, for instance, a spontaneous symmetry breaking explanation for CP violation. This is interesting because we have other reasons for doing so. The mechanism of baryon-number generation, as currently understood, requires that CP must already be violated at the grand-unification phase transition [24].

Domain walls, then, can be ruled out, so let us pass on to the more physically interesting possibilities of singularities of lower dimension.

Linear singularities, or strings, can appear if M contains unshrinkable loops. A familiar example is the $N = 2$ case, the Landau–Ginzburg model. Here, if the angle of $\langle\phi\rangle$ changes by a nonzero multiple of 2π as one goes around the string, it cannot be eliminated: a thin tube with $\langle\phi\rangle \sim 0$ is trapped within the ordered phase. If we assume that the radius is approximately the zero-temperature correlation length ξ_0, we may estimate the mass per unit length (which in this relativistic situation is the same thing as the tension) as

$$\mu \sim \xi_0^2 \Delta f \sim \eta^2. \tag{12}$$

When $h \ll g$ it is more accurate to take the radius to be the penetration depth m_V^{-1}, but the result $\mu \sim \eta^2$ still holds. More precisely, (12) should contain a factor $f(h/g)$ of order unity.

If we compare the mass of a single string stretched across the universe, when its age or horizon distance is t, to the mass of the whole, we obtain the ratio

$$\mu t/\rho t^3 \sim \eta^2/m_P^2,$$

which even for strings arising at the grand unification transition is only in the range 10^{-8} to 10^{-10}. Significant numbers of strings could thus be accommodated without introducing unacceptable gravitational effects.

Finally, if M contains closed two-dimensional surfaces that cannot be shrunk to a point (within M), then monopoles can exist. Their mass, following the same procedure that led to (11) and (12), would be estimated as $m_{\mathrm{mon}} \sim \xi_0^3 \Delta f \sim \eta/h$. In this case, however, a better estimate is [25]

$$m_{\mathrm{mon}} \sim 4\pi\eta/g \tag{13}$$

though again there should be a factor $f(h/g)$ of order unity.

5. Classification of singularities

In each case the existence of singular structures requires that one of the homotopy groups of M be non-trivial, and the elements of that group then serve to classify the possible singularities [26]. Strings, for instance, are classified by the elements of $\pi_1(M)$, i.e. equivalence classes of loops in M. The possibilities are summarised in table 2. Textures are topologically stable structures involving no singularity [27]. Though certainly of interest, I shall not discuss them further in this talk.

One of the virtues of table 2 is that the homotopy groups of relevant spaces are well known.

Since we are not interested in breaking discrete symmetries, we may assume that the gauge symmetry group G is connected (i.e. $\pi_0(G)$, which counts the number of disconnected pieces, is trivial, $\pi_0(G) = \mathbb{1}$). Moreover, we can always choose it to be *simply* connected ($\pi_1(G) = \mathbb{1}$), by working with the covering group – for example, using not $SO(N)$, but its two fold covering group $\widetilde{SO(N)} \cong \mathrm{Spin}(N)$. Finally, let us assume that G is a simple group. There are isomorphisms between the relevant homotopy groups of the manifold of vacuum states $M = G/H$ and the unbroken symmetry subgroup H, namely

$$\pi_1(M) \cong \pi_0(H), \qquad \pi_2(M) \cong \pi_1(H). \tag{14}$$

For the classification of monopoles, (14) gives

$$\pi_2(M) \cong \mathbb{Z}^k \times K$$

where k is the number of U(1) factors in H, $\mathbb{Z}$ denotes the group of integers, and K is a finite group. (Typically, K may be the group of integers modulo 2, or a power thereof: its elements label the mod-2 monopoles which can appear in certain cases. I shall not discuss them further in this talk.) The important thing about this result is that monopoles MUST appear if H contains at least one U(1) factor. Whatever grand unification theory we adopt, we know that the last and second last phases have symmetry groups containing a U(1) factor. So threre certainly is some phase transition, above the Weinberg–Salam one, at which monopoles make an appearance. The problems this raises I will return to.

Next let us consider the classification of strings. Here the Landau–Ginzburg model is a special case whose symmetry group U(1) is non-simple. For that case, $\pi_1(M) = \mathbb{Z}$, so the strings are labelled by an integer. When G is simple, however, $\pi_1(M)$ is usually a *finite* group. Let me illustrate with two examples, in which ϕ is chosen to belong to the five-dimensional symmetric tensor representation of SO(3), and G = SO(3). First, if the potential is chosen so that a typical value of the vacuum expectation value is

$$\langle\phi\rangle = \begin{pmatrix} \eta & \cdot & \cdot \\ \cdot & \eta & \cdot \\ \cdot & \cdot & -2\eta \end{pmatrix}, \tag{15}$$

Table 2

Structure	Dimension of singularity	Classified by
Domain walls	2	$\pi_0(M)$
Strings	1	$\pi_1(M)$
Monopoles	0	$\pi_2(M)$
Textures	–	$\pi_3(M)$

then H is isomorphic to O(2). Since this group has two disjoint pieces, $\pi_1(M) = \mathbb{Z}_2$. This model contains "mod-2 strings".

The second case illustrates an intriguing possibility raised, in the context of condensed matter physics, by Toulouse and Poénaru [28], namely the existence of non-commuting strings. Suppose that in place of (15), $\langle\phi\rangle$ takes the form

$$\langle\phi\rangle = \begin{pmatrix} a & \cdot & \cdot \\ \cdot & b & \cdot \\ \cdot & \cdot & -a-b \end{pmatrix}.$$

Then one finds that the group H is a *non-abelian* group of order 8, isomorphic to the "quaternion group" Q. Here we have $\pi_1(M) \cong H \cong Q$. Thus there are different types of strings, which cannot, in general, pass through one another. These strings can have junctions, vertices where three different types join together. Models with this feature do not seem to arise very naturally in particle physics, though they can certainly be constructed [29].

6. Evolution of strings

In this section I want to discuss how the system of strings, once formed, will evolve in time.

Consider first a section of string moving with velocity v (assumed $\ll 1$) through a medium of relativistic particles or radiation. Assuming that the string presents an effective cross section ξ_0 per unit length, it will experience a retarding force of order $\xi_0 \rho v$. Thus the effective damping time for the string velocity is [20]

$$t_d \sim \mu/\xi_0\rho \sim h\eta^3/\mathcal{N}T^4. \tag{16}$$

Suppose that initially this section of string is at rest, with a local radius of curvature r. It will experience an initial acceleration $\sim\mu/\mu r = 1/r$. If the medium is dense, so that $t_d \ll r$ (an assumption to be verified later) then the string will acquire a limiting velocity $\sim t_d/r$, and hence the kink will be straightened out in a time of order r^2/t_d.

Let us suppose that initially we have a random tangle of strings. From the arguments in section 4, we expect the length scale L to be initially of order

$$L \sim \xi_G \sim 1/h^2 T_c. \tag{17}$$

String tension will cause small kinks to straighten out. From time to time this process will lead to strings crossing, when (in the commutative case at least) they can exchange partners, thus yielding new sharp kinks which straighten out in turn. Occasionally, small loops may shrink to a point and disappear. All in all, we have a decrease in the total length of string, which means an increase in the length scale. (Roughly, the length of string per unit volume is L^{-2}.)

From (16) and (17) one sees that initially $t_d \ll L$, thus validating the assumption made earlier. It seems reasonable to assume that the time scale for growth of L is L^2/t_d, i.e.

$$\frac{1}{L}\frac{dL}{dt} \sim \frac{t_d}{L^2}. \tag{18}$$

Initially, therefore, L grows like $t^{1/2}$, rapidly increasing the ratio L/t_d. Over a longer period of time, however, (16) shows that $t_d \propto t^2$ whence (18) yields $L \propto t^{3/2}$. Eventually, therefore, t_d will catch up with L. It is not hard to check that this will happen when both t_d and L are of the same order as the age of the universe t; in fact when [20]

$$t_d \sim L \sim t \sim m_P^2/h\eta^3 = t_*, \tag{19}$$

say.

For strings appearing at the grand unification transition, we have $t_* \simeq 10^{-27}$ s, so that this stage is reached long before the Weinberg–Salam transition. However, t_* depends sensitively on η. For $\eta \sim 100$ GeV, one would get $t_* \simeq 10^7$ s. The Weinberg–Salam transition itself presumably does not generate strings, but if there were an intermediate transition not too far above it, one might get strings lasting in fair numbers to a relatively late stage. This may be relevant to theories of galaxy formation.

We see then that what happens is that the scale size L of the tangle of strings grows until it is of the same order of magnitude as the distance t to the causal horizon. Thereafter L cannot grow faster than t, but t_d continues to grow, so that strings move with little damping and presumably acquire relativistic speeds.

One interesting question that seems rather hard to answer is whether the kind of "knotted spaghetti" that would be generated if there were non-commuting strings would evolve in much the same way, or substantially more slowly.

The most intriguing possibility raised by this idea is that it might provide the basis for a theory of galaxy formation. This is at present a major unsolved problem in cosmology. Let me briefly review some features of the problem.

The basic mechanism of gravitational condensation was discussed by Jeans who showed that in any gravitating system there is a minimum length scale required for a density perturbation to grow [30]. This is the Jeans length

$$L_J \simeq c_s/\sqrt{G\rho} \simeq c_s t,$$

where c_s is the sound velocity. In the early radiation-dominated era, and before the time of electron–proton recombination, $c_s \simeq 1/\sqrt{3}$, so that $c_s t$ is a substantial fraction of the radius of the universe. Thus only very large-scale perturbations could start to grow in amplitude. They can grow essentially linearly before coming within the causal horizon, but will continue to grow for only a short time thereafter, until the Jeans length becomes too large.

After the recombination era, c_s drops quite suddenly to the value

$$c_s \simeq (5T/3m_H)^{1/2}$$

typical of hot hydrogen gas. From then on masses larger than 10^5 solar masses can start to contract.

What one needs to trigger the process of galaxy formation are initial perturbations, with $\delta\rho/\rho \sim 10^{-2}$, present at the recombination era [4]. The problem is to find a mechanism that will generate such initial perturbations – of course without affecting the isotropy of the 3 K background radiation, which shows that at this epoch temperature fluctuations were limited to $\delta T/T \lesssim 10^{-3}$.

In one respect this may not be as hard to achieve as it sounds. During the plasma era, which lasts from about $t \sim 1$ s to 10^{13} s, photon scattering maintains isothermal conditions. The "adiabatic" part of any initial perturbation on a galactic scale or less will rapidly die out, but any "isothermal" part – a pure density fluctuation – will remain. There is no damping mechanism for such fluctuations that operates on a short enough time scale to be relevant [31].

The earliest possible time at which an initial perturbation can be created by any local mechanism (rather than simply inserted in the initial conditions) is the time at which the relevant mass comes within the causal horizon. This is about 1 year for a galactic mass and 10 s for a stellar cluster. However, nonlocal mechanisms may exist. A very interesting possibility has been suggested by Press [38].

It is certainly possible that strings might provide the essential mechanism for generating these initial perturbations. The difficulty is to make their interactions sufficiently effective. The heavy strings typical of the grand unification transition would by this time be interacting rather weakly; whereas strings generated later may be too light. However it may well be that these conclusions would be changed by a better understanding of the string interactions.

7. Monopoles

The mass of a monopole generated in the grand unification transition is more than 10^{15} GeV. The initial density of monopoles may be estimated to be roughly of order

$$n_{\text{mon}} \sim 1/\xi_G^3 \sim h^6 \eta^3 \tag{20}$$

corresponding to the same initial scale size as in the case of strings.

The subsequent evolution, however, is very different. Monopoles (at least the lightest ones) are stable particles which can be removed only by annihilation. Since monopoles and anti-monopoles attract one another strongly, any that are close should rapidly annihilate. Later annihilation depends, however, on the diffusion of monopoles to anti-monopoles through the surrounding medium. This is a slow process. It is difficult to avoid the conclusion that at the time relevant for helium synthesis, the total mass density of monopoles would exceed that of all other matter by many orders of magnitude [32]. This is a disaster because it would completely destroy the agreement between observed and calculated abundances for helium and other light elements. (It is, perhaps, also puzzling that no evidence of monopoles has been seen if they are in fact so common, but it may well be possible to show that they would preferentially collect in stellar cores where they could hardly be seen. This is not, therefore, a very strong argument.) One possible answer to this problem is to make the phase transition strongly first order [33]. This can be done in at least two ways.

Firstly, if the coupling constants are of very different magnitudes, radiative corrections may do the trick. For example, if $h \sim g^2$ then the single vector loop contribution, of order g^4, should be included along with terms of order h^2. This contribution is of the form [34]

$$(m_V^2/8\pi)^2 \ln(m_V^2/\mu^2), \tag{21}$$

where μ is a renormalization point. It is easily seen that such a term will lead to an effective potential typical of a first-order transition, with two minima separated by a barrier.

Secondly, we may consider models with an explicit ϕ^3 term. Consider for example, an SU(5) theory with ϕ in the 24-dimensional adjoint representation. Here

$$U(\phi) = a\,(\text{tr}\,\phi^2)^2 + b\,\text{tr}\,\phi^4 + c\,\text{tr}\,\phi^3 + d\,\text{tr}\,\phi^2 + e. \tag{22}$$

This model (with $d<0$) can give rise to two distinct phase transitions [35, 6, 33] provided that $b>0$ and $a>-b/5$. At the higher transition, SU(5) is broken to

$$H_1 = SU(4) \times U(1)$$

while at the second H_1 changes to

$$H_2 = SU(3) \times SU(2) \times U(1).$$

Note that H_2 is NOT a subgroup of H_1. The situation is curious because both the second and third phases have monopoles, but of different types, with different expressions for the monopole charge.

One thing that is not immediately clear is whether monopoles of the first type, which necessarily disappear at the second transition, tend, in doing so, to generate monopoles of the second type. If so, then having two distinct transitions does not help much. On the other hand, the first-order character of the transition may help, by delaying the time of the transition.

This is actually a rather unusual kind of first-order transition. As always there are 3 relevant temperatures. Highest is the temperature T_b at which an asymmetric minimum first appears in the effective potential. At the true critical temperature T_c this minimum has dropped to the level of the central minimum. Finally at T_a the central minimum itself disappears.

As the universe cools, nothing happens until after it has passed T_c. Even then, none of the new phase will appear until it can be nucleated. However, the probability of tunneling through the barrier, which depends on the exponential of the action integrated along a "most probable escape path" [36], is exceedingly small until we are almost at T_a. In fact, there is essentially zero probability of generating a bubble of ordered phase by tunneling. The system in effect reaches T_a in a state of extreme supercooling. At T_a there is a sudden transition, almost simultaneously in all parts of the universe, a transition that releases a huge latent heat, reheating the universe, perhaps nearly to T_c. It seems to be difficult in this situation to estimate the length scale of the resulting structure, and hence the monopole density. Thus it is not clear whether such a mechanism can allow us to escape the problem of overproduction of monopoles.

Note too, that such a strongly first-order transition generates a lot of entropy, which could drastically change the computed photon-to-baryon ratio (or entropy per baryon).

There is yet another effect, related to the induced cosmological term, to which I now turn.

8. Cosmological term

Observations on the present rate of recession of galaxies place rather stringent limits [15] on the magnitude of the cosmological constant Λ appearing in (6). Since a constant added to the potential $U(\phi)$ would effectively contribute to Λ, we must assume that the minimum value of U is in fact close to zero. It seems reasonable to suppose that at $T=0$ the minimum value of the effective potential (i.e. the energy density of the vacuum) should be precisely zero. This then fixes the constant in U to have the value chosen in (3).

However, it is clear that with this choice, there is in the high-temperature phase a nonzero constant term in U, namely $\frac{1}{8}h^2\eta^4$. This term contributes a constant to the free energy density f, or the energy density ρ. Lorentz invariance of the vacuum suggests that such a term must represent part of a contribution to the energy-momentum tensor $T_{\mu\nu}$ that is proportional to $g_{\mu\nu}$. This means in particular a negative contribution to the pressure. Such a term would be indistinguishable in its effects from a cosmological term Λ.

It is easy to verify that such a term is present. If we calculate the free energy densities f_- and f_+ of the lower and upper phase (i.e. the minimum values of V) for the model in section 2, we obtain expressions of the form

$$f_- = -b'T^2 - c'T^4, \tag{23}$$

$$f_+ = a - bT^2 - cT^4, \tag{24}$$

where $a = \frac{1}{8}h^2\eta^4$. Correspondingly the energy density in the upper phase is

$$\rho_+ = a + bT^2 + 3cT^4. \tag{25}$$

Recalling that the pressure is simply $p = -f$, we see that a does indeed yield a contribution proportional to $+ag_{\mu\nu}$, i.e. in effect a cosmological term.

There is no objection to such a term. The observational evidence requires Λ to be zero or near zero in the present phase of the universe, but there is no reason to exclude a large cosmological term in earlier phases. However, Bludman [37] has pointed out an interesting dynamical effect that such a term might have, in the case where the relevant transition is strongly first-order.

Suppose we follow R back in time. Once we have reached the radiation-dominated era, but still below the critical temperature, (6) yields

$$(\dot{R}/R)^2 = \gamma'/R^4 - K/R^2.$$

The first term dominates increasingly as R decreases. If K is positive (i.e. for a closed universe), there is a time, still far in the future, at which $\dot{R}$ vanishes, but as we go back in time it increases monotonically.

Now consider what happens when we go through a first-order transition, at $R = R_c$. Corresponding to the sudden change from (23) to (24) we find that the equation for $\dot{R}$ changes to

$$(\dot{R}/R)^2 = \gamma/R^4 - K/R^2 + \alpha \tag{26}$$

where α and γ come respectively from a and c in (25). (We ignore for simplicity the effect of the change from b' to b in the coefficient of T^2.) By energy conservation one has

$$\alpha = (\gamma' - \gamma)/R_c^4.$$

The intriguing thing is that if γ is small enough, then the right hand side of (26) may vanish for some $R < R_c$. This would mean that as we follow R back in time we never reach $R = 0$. Instead, at some minimum value, the universe "bounces".

This possibility would only be realised for rather extreme values of the parameters. Firstly one needs $K > 0$ which means that the density in the universe now must exceed the critical density required for closure. Current observational limits do not favour such a high density universe. Moreover, one needs a very strongly first-order transition, caused for instance by very unequal couplings, corresponding to the case where the Higgs particles are much lighter than the gauge particles. Nevertheless, the possibility clearly deserves further study.

Acknowledgements

I wish to acknowledge helpful discussions with M.J. Duff, P. Ginsparg and D. Olive. I am indebted to F.R. Klinkhamer for pointing out errors in an earlier draft of this paper, and for drawing my attention to the important work of Press [38].

References

[1] A.D. Sakharov, Pis'ma Zh. Eksp. Teor. Fiz. 5 (1967) 32 [JETP Lett. 5 (1967) 25];
Ya.B. Zel'dovich, Pis'ma Zh. Eksp. Teor. Fiz. 24 (1976) 29 [JETP Lett. 24 (1976) 25];
M. Yoshimura, Phys. Rev. Lett. 41 (1978) 281; 42 (1979) 746 (E); Phys. Lett. 88B (1979) 294;
S. Dimopoulos and L. Susskind, Phys. Rev. D18 (1978) 4500; Phys. Lett. 81B (1979) 416;
D. Toussaint, S.G. Treiman, F. Wilczek and A. Zee, Phys. Rev. D19 (1979) 1036:
S. Weinberg, Phys. Rev. Lett. 42 (1979) 850;
J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Phys. Lett. 80B (1979) 360; 82B (1979) 464;
A.Yu. Ignatiev, V.A. Kuzmin and M.E. Shaposhnikov, Phys. Lett. 87B (1979) 114;
D.V. Nanopoulos and S. Weinberg, Harvard Preprint HUTP 79/A023;
E.W. Kolb and S. Wolfram, Caltech preprint OAP-579 [CALT-68-754]; See also:
R.N. Mohapatra and J.C. Pati, Bhubaneswar preprint IP/BBSR/79-32;
J. Ellis and G. Steigman, Phys. Lett. 89B (1980) 186;
A. Hosoya and K. Vehara, Osaka preprint OU-HET-29 (1979).
[2] D.N. Schramm and R.V. Wagoner, Ann. Rev. Nuc. Sci. 27 (1977) 37;
Y. David and H. Reeves, Phil. Trans. Roy. Soc. A296 (1980) 415.
[3] R.B. Partridge and D.T. Wilkinson, Phys. Rev. Lett. 18 (1967) 557;
E.S. Chang, P.R. Saulson, D.T. Wilkinson and B.E. Corey, Ap. J. Lett. 232 (1979) L139.
[4] B.J.T. Jones, Rev. Mod. Phys. 48 (1976) 107;
M.J. Rees, Physica Scripta 17 (1978) 371.
[5] J. Pati and A. Salam, Phys. Rev. D8 (1973) 1240;
H. Georgi and S.L. Glashow, Phys. Rev. Lett. 32 (1974) 438;
H. Georgi, H.R. Quinn and S. Weinberg, Phys. Rev. Lett. 33 (1974) 451.
[6] A. Buras, J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Nucl. Phys. B135 (1978) 66.
[7] E.J. Eichten and F.L. Feinberg, Phys. Rev. D10 (1974) 3254;
B.J. Harrington and A. Yildiz, Phys. Rev. D11 (1975) 779;
K. Huang, Contribution to the 1980 Guangchou (Canton) Conf. on Theoretical Particle Physics; MIT Preprint CTP 823 (1979).
[8] H.D. Politzer, Phys. Reports C14 (1974) 129.
[9] P.W. Higgs, Phys. Rev. Lett. 12 (1964) 132;
R. Brout and F. Englert, Phys. Rev. Lett. 13 (1964) 32.
[10] T.W.B. Kibble, Phys. Rev. 155 (1967) 1554.
[11] V.L. Ginzburg and L.D. Landau, Zh. Eksp. Teor. Fiz. 20 (1950) 1064.
[12] D.A. Kirzhnits, Zh. Eksp. Teor. Fiz. Pis'ma Red. 15 (1972) 745 [JETP Lett. 15 (1972) 529];
D.A. Kirzhnits and A.D. Linde, Phys. Lett. 42B (1972) 471; Zh. Eksp. Teor. Fiz. 67 (1974) 1263 [JETP 40 (1974) 628];
S. Weinberg, Phys. Rev. D9 (1974) 3357;
C. Bernard, Phys. Rev. D9 (1974) 3312;
L. Dolan and R. Jackiw, Phys. Rev. D9 (1974) 3320.
[13] R.B. Mohapatra and G. Senjanović, Phys. Lett. 89B (1979) 571.
[14] See for example: S. Weinberg, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity (New York, Wiley, 1972).
[15] J.R. Gott, J.E. Gunn, D.N. Schramm and B.M. Tinsley, Ap. J. 194 (1974) 543.
[16] R. Hagedorn, Nucl. Phys. B24 (1970) 93.

[17] See S. Weinberg, ref. [14] chapter 15.
[18] G. Steigman, D.N. Schramm and J.E. Gunn, Phys. Lett. 66B (1977) 202;
J. Yang, D.N. Schramm, G. Steigman and R.T. Rood, Ap. J. 227 (1979) 697.
[19] A.D. Linde, Phys. Lett. 83B (1979) 311;
D. Schramm and G. Steigman, Phys. Lett. 87B (1979) 141;
S. Dimopoulos and G. Feinberg, Columbia preprint UC-TP 159 (1979);
D.V. Nanopoulos, D. Sutherland and A. Yildiz, Harvard preprint HUTP-79/A03;
D.V. Nanopoulos, CERN preprint TH-2737-CERN (1979).
[20] T.W.B. Kibble, J. Phys. A9 (1976) 1387.
[21] V.L. Ginzburg, Fiz. Tverdogo Tela 2 (1960) 2031 [Sov. Phys. Solid State 2 (1960) 1824].
[22] M.B. Einhorn, A.L. Stein and D. Toussaint, Michigan preprint UM-HE80-1 (1980).
[23] Ya.B. Zel'dovich, I.Ya. Kobzarev and L.B. Okun, Zh. Eksp. Teor. Fiz. 67 (1974) 3 [JETP 40 (1974) 1].
[24] S. Dimopoulos and L. Susskind, ref. [1].
[25] E.B. Bogomol'nyĭ and M.S. Marinov, Yad. Fiz. 23 (1976) 676 [Sov. J. Nuc. Phys. 23 (1976) 355].
[26] S. Coleman, Classical lumps and their quantum descendants, Proc. 1975 School of Subnuclear Physics "Ettore Majorana", ed. A. Zidhichi (New York, Plenum, 1977) p. 297;
T.W.B. Kibble, Procs., Ecole Internationale de Physique des Particules Elementaires, Baško Polje Makarska, 1975;
G.Toulouse and M. Kléman, J. Physique Lettres 37 (1976) L-149;
G.E. Volovik and V.P. Mineev, Pis'ma Zh. Eksp. Teor. Fiz. 24 (1976) 605 [JETP Lett. 24 (1977) 561].
[27] G.E. Volovik and V.P. Mineev, Zh. Eksp. Teor. Fiz. 72 (1977) 2256 [JETP 45 (1977) 1186].
[28] G. Toulouse, J. Physique Lettres 38 (1976) L-67;
V. Poénaru and G. Toulouse, J. Physique 38 (1977) 887.
[29] G. McCauley, University of Birmingham preprint, May 1978.
[30] J.H. Jeans, Astronomy and Cosmogony (Cambridge: CUP, 1928) p. 337.
[31] P.J.E. Peebles, Ap. J. 142 (1965) 1317;
J.D. Barrow, Phil. Trans. Roy. Soc. Lond. A296 (1980) 273.
[32] P.J. Adams, V. Canuto and H.-Y. Chiu, Phys. Lett. 61B (1976) 397;
Ya.B. Zel'dovich and M.Yu. Khlopov, Phys. Lett. 79B (1978) 239;
J.P. Preskill, Phys. Rev. Lett. 43 (1979) 1365;
M.B. Einhorn, D.L. Stein and D. Toussaint, Michigan preprint U. Mich. HE80-1.
[33] A.H. Guth and S.-H.H. Tye, preprint SLAC-PUB-2448 [CLNS-79/441].
[34] D.A. Kirzhnits and A.D. Linde, Ann. Phys. N.Y. 101 (1976) 195;
A.D. Linde, Rep. Prog. Phys. 42 (1979) 389.
[35] L.F. Li, Phys. Rev. D9 (1974) 1723.
[36] S. Coleman, Phys. Rev. D15 (1979) 2929; D16 (1977) 1248 (E);
C.G. Callan and S. Coleman, Phys. Rev. D16 (1977) 1762.
[37] S.A. Bludman, Pennsylvania preprint UPR-0143T (1979).
[38] W.H. Press, Physica Scripta 21 (1979) 702.