MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — December 1, 2009
Prof. Alan Guth

Lecture Notes 12
THE MAGNETIC MONOPOLE PROBLEM

THE MAGNETIC MONOPOLE PROBLEM:

In addition to the horizon and flatness problems discussed in Lecture Notes 10, the standard cosmological model potentially suffers from another problem, known as the magnetic monopole problem. If one accepts the basic ideas of grand unified theories (GUT's) in addition to those of the standard cosmological model, then one is led to the conclusion that there is a serious problem with the overproduction of particles called "magnetic monopoles".

A magnetic monopole is a particle with a net North or South magnetic charge. The magnetic field of a monopole points radially outward (or inward), with a magnitude proportional to $1/r^2$, just like the Coulomb field of a point electric charge. Such particles do not exist in the usual formulation of electromagnetism, in which all magnetic effects arise from electric currents. An ordinary bar magnet, with internal currents associated with the alignment of electronic orbits, has the form of a dipole, with North and South poles at the two ends. If a bar magnet is cut in half, one obtains two dipoles, each with a North and South pole. Grand unified theories, however, imply that magnetic monopoles necessarily exist. They are generally superheavy particles, with mass energies of approximately 10^{18} GeV.

The magnetic monopoles are created during the course of a phase transition that is predicted by the grand unified theories. This phase transition is intimately tied to the physics of "spontaneous symmetry breaking", which is one of the key features of the GUT's.

In general, symmetries of particle theories can be divided into two types. The most familiar are the "spacetime" symmetries. An example of a spacetime symmetry is rotational invariance, which is the statement that the form of the laws of physics does not change if one adopts a new coordinate frame that is rotated with respect to the first. The second type of symmetry, called an "internal" symmetry, is one that relates the behavior of one kind of particle to that of another. In addition to the usual spacetime symmetries, grand unified theories contain a large set of internal symmetries, which are collectively called the GUT symmetry.

At the fundamental level a GUT contains a single interaction, carried by a single type of vector boson. More precisely, there are a number of distinct vector bosons, but their properties are related by the underlying GUT symmetry. The vector bosons are said to form a single "multiplet", and the distinct vector bosons

are said to be "components" of the multiplet. The situation is completely analogous to the gluons of quantum chromodynamics— there are eight gluons which make up a single multiplet. In the simplest grand unified theory— the $SU(5)$ theory proposed by Howard Georgi and Sheldon Glashow in 1974— the multiplet of fundamental vector bosons contains 24 different particles.

At the fundamental level there is also no distinction in a GUT between a quark, an electron, or a neutrino. There is usually more than one multiplet of these particles, but each multiplet contains both quark and lepton components. The properties of the particles within a multiplet are again related by the GUT symmetry.

The distinction between the components of a multiplet which we observe at low energies arises from the process of spontaneous symmetry breaking. Although the fundamental theory possesses the GUT symmetry, the lowest energy state of the system does not. The analogy of crystal formation is discussed in the "Inflationary Universe" article,* which you may want to read. The spontaneous symmetry breaking is accomplished by including in the formulation of the theory a special set of fields known as Higgs fields (after Peter W. Higgs of the University of Edinburgh). These Higgs fields also form one or more multiplets, and the properties of the fields within a multiplet are related by the GUT symmetry. The symmetry is unbroken when all the Higgs fields have a value of zero, but it is spontaneously broken whenever at least one of the Higgs fields acquires a nonzero value.† The theory is formulated so that one or more Higgs fields has a nonzero value in the state of lowest energy (which is by definition the vacuum state). The vacuum state is then not unique, since another state of the same energy density can be found by replacing the values of the Higgs fields by any combination of values which is related by the GUT symmetry to the initial combination of values. The choice of which vacuum state occurs in a given region is then determined by its history, just as the orientation of the axes of a crystal is determined by how the crystal was formed.

At high temperatures these Higgs fields will undergo large thermal fluctuations. In many grand unified theories the high temperature thermal equilibrium

---

* "The Inflationary Universe", by Alan H. Guth and Paul J. Steinhardt (*Scientific American*, May 1984).

† You might think that a state of unbroken symmetry can be achieved by allowing all of the Higgs fields to have the same value. However, the symmetry involves more than just the interchange of one component field with another. Rather, the symmetry includes the possibility of replacing one component field by a linear combination of other components, exactly as a component of a rotated vector is expressed as a linear combination of the components of the vector before rotation. For the case of vectors and rotations, one knows that the only vector which is invariant under all rotations is the vector for which all three components vanish, and the same logic holds for the Higgs fields.

state is one in which the values of the fields average to zero, which means that the GUT symmetry is unbroken. As the system cools a phase transition is encountered. A phase transition is characterized by a specific temperature, called the critical temperature, at which some thermal equilibrium properties of the system change discontinuously. In this case, at temperatures below the critical temperature, some subset of the Higgs fields acquire nonzero mean values in the thermal equilibrium state— the GUT symmetry is thereby spontaneously broken. There may be one or perhaps several such phase transitions before the system reaches the lowest temperature phase— the phase which includes the vacuum. For simplicity, we will discuss the case in which there is only one such phase transition. In any case, the broken symmetry state which exists below the critical temperature is not unique, for precisely the same reason that the vacuum state is not unique.

In the standard cosmological model, it is assumed that this phase transition occurs quickly once the critical temperature is reached. Thus, in any given region of space the Higgs fields will settle into a broken symmetry state, in which some subset of the Higgs fields acquire nonzero mean values. The choice of this subset is made randomly, just as the orientation of the axes of a crystal are determined randomly when the crystal first starts to condense from a molten liquid. The other particles in the theory, such as the quarks and leptons, are also described by fields, which interact with the Higgs fields in a manner consistent with the GUT symmetry. Through these interactions, the randomly selected combination of nonzero Higgs fields determines what combination of the fields will act like an electron, what combination will act like a u-quark, etc. The same random choice determines what combination of vector boson fields will act like the photon field, and what combinations will act like the W's, Z's, or gluons. In addition, some vector bosons acquire masses of the order of 10^{16} GeV, and these vector bosons are then irrelevant to the low energy physics which we observe in present-day accelerator experiments.

The magnetic monopoles are examples of "defects" which form in the phase transition. The defects arise when regions of the high temperature symmetric phase undergo a transition to different broken-symmetry states. In the analogous situation when a liquid crystallizes, different regions may begin to crystallize with different orientations of the crystallographic axes. The domains of different crystal orientation grow and coalesce, and it is energetically favorable for them to smooth the misalignment along their boundaries. The smoothing is often imperfect, however, and localized defects remain.

The detailed nature of these defects is too complicated to explain here, so I will settle for the statement of some general facts. There are three types of defects that can occur. The simplest type is a surfacelike defect called a domain wall. This type of defect arises whenever the broken-symmetry state in one region of space cannot

be smoothly interpolated with the broken-symmetry state in a neighboring region of space. A domain wall then forms at the interface between the two regions. Some grand unified theories allow for the formation of such domain walls, and others do not. The second type is a linelike defect called a cosmic string. Again, some grand unified theories allow such defects to exist, and others do not. Finally, the third type is a pointlike defect, called a magnetic monopole. In contrast to the first two types of defects, magnetic monopoles exist in ***any*** grand unified theory.

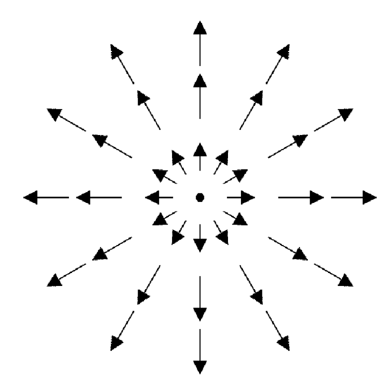
To see how a pointlike defect can arise, let us consider the simplest theory in which they occur. This theory is too simple to describe the real world, but it serves as a "toy" model which is useful to illustrate many features of spontaneously broken gauge theories. The theory has a three-component multiplet of Higgs fields, which I will denote by ϕ_a, where $a = 1, 2,$ or 3. The symmetry which operates on this multiplet is identical in its mathematical form to the transformations that describe how the three components of an ordinary vector are modified by a rotation. The potential energy density associated with the Higgs fields is then a function of the three components ϕ_a. The energy density function, however, is an ingredient of the fundamental theory which must be invariant under the symmetry. Thus, the energy density can depend only on

$$|\phi| \equiv \sqrt{\phi_1^2 + \phi_2^2 + \phi_3^2}\ . \tag{12.1}$$

The energy density will be minimized when $|\phi|$ has some particular value, say ϕ_v. Spontaneous symmetry breaking occurs if $\phi_v \neq 0$, which I will assume to be the case. Now consider the following static configuration of the Higgs field:

$$\phi_a(\vec{r}) = f(r)\hat{r}_a\ , \tag{12.2}$$

where $r \equiv |\vec{r}|$, $\hat{r}_a$ denotes the a-component of the unit vector $\hat{r} = \vec{r}/r$, and $f(r)$ is a function which vanishes when $r = 0$ and approaches ϕ_v as $r \to \infty$. This configuration is sketched below. An arrow is drawn at each point in space, and the three vector components of the arrow are used to represent the three components of the Higgs field:



If the diagram were constructed as a three-dimensional model, then all of the arrows would point radially outward from the origin. (An antimonopole is described by a similar picture, except that the arrows would point radially inward.) Note that the Higgs fields are in a vacuum state at large distances, but the fields differ from their vacuum values in the vicinity of $r = 0$, resulting in a concentration of energy. It can be proven that this configuration is "topologically stable" in the following sense: if the boundary conditions for the fields at infinity are held fixed, and if the fields are required to be continuous functions of position, then there must always be at least one point at which all three components of the Higgs field vanish. I will not attempt to prove this theorem, but I recommend that you stare at the diagram until the theorem becomes believable. Because of this topological property of the magnetic monopole configuration, it is sometimes referred to as a "knot" in the Higgs field. The configuration involves a concentration of energy localized around a point, and it behaves exactly as a particle.

So far I have not mentioned anything about magnetic fields, so the astute reader is no doubt wondering why these particles are called magnetic monopoles. The answer to this question depends on energy considerations. In the absence of any other fields, the energy of the magnetic monopole Higgs field configuration would be infinite. To understand this infinity, you must accept without proof the fact that the expression for the energy density of a Higgs field contains a term proportional to the square of the gradient. The form of Eq. (12.2) for large r (with $f(r) \to \phi_v$) then implies that the gradient of ϕ_a falls off as $1/r$ at large distances. The total energy within a large sphere therefore diverges linearly with the radius of the sphere. However, the expression for the energy density becomes more complicated when the vector boson fields are included. It is well beyond the range of this course, but it can be shown that the total energy of the Higgs field configuration of Eq. (12.2) can be made finite only if the configuration includes vector boson fields that correspond to a net magnetic charge. Even the magnitude of this magnetic charge is determined uniquely. The magnetic charge must correspond to a value $1/(2\alpha)$ times the electric charge of an electron. Here α denotes the usual fine structure constant of electrodynamics: $\alpha = e^2/\hbar c$ in cgs units, or $\alpha = e^2/4\pi\epsilon_0\hbar c$ in mks units. In any case, $\alpha \approx 1/137$. This means that the magnetic charge of a monopole is 68.5 times as large as the electric charge of an electron, and the force between two monopoles is then $(68.5)^2$ times as large as the force between electrons at the same distance.

The mass of a monopole can be estimated in these models, and it turns out to be extraordinary. The mass is approximately $1/\alpha$ times the mass scale at which the unification of forces occurs. Since the unification of forces occurs roughly at 10^{16} GeV, it follows that Mc^2 for a monopole is about 10^{18} GeV.

Having gone through the basic physics, we are now in a position to discuss how one estimates the number of magnetic monopoles that would be produced

in the GUT phase transition. I will present a crude argument which is probably accurate to within one or two orders of magnitude. Although the argument will be crude, to my knowledge no one has carried out a calculation that is more accurate. The magnetic monopole problem is so severe than an ambiguity of two orders of magnitude in the estimate is unimportant to the conclusion.

Recall that the monopoles are really knots in the Higgs field, so their number density is related to the misalignment of the Higgs field in different regions of space. This misalignment can be characterized by a "correlation length" ξ. We will need only an approximate definition of this correlation length, so it will suffice to say that ξ is the minimum length such that the Higgs field at a given point in space is almost uncorrelated with the Higgs field a distance ξ away. One then estimates that the number density of magnetic monopoles and antimonopoles is given roughly by

$$n_M \approx 1/\zeta^3 \; . \tag{12.3}$$

In words, we are estimating that every cube with a side of length ξ will have, on the average, approximately one magnetic monopole in it. This estimate was first proposed by T.W.B. Kibble of Imperial College (London).

The remaining problem is to estimate ξ. Here we will be working in the context of standard cosmology, which assumes that the phase transition occurs quickly once the critical temperature is reached. Under these assumptions the phase transition has no significant effect on the evolution of the early universe. When the universe cools below the critical temperature T_c of the GUT phase transition (with $kT_c \approx 10^{16}$ GeV), it becomes thermodynamically probable for the Higgs field to align uniformly over reasonably large distances. If the system were allowed time to reach thermal equilibrium, then very few monopoles would be present— their abundance would be suppressed by the usual factor

$$e^{-Mc^2/kT}$$

from statistical mechanics. For this case the factor is roughly $e^{-100} \approx 10^{-43}$. However, if the whole process must happen on the time scales at which the early universe evolves, then there is not enough time for this long range correlation of the Higgs field to become established. While we are not prepared to calculate the correlation length in these circumstances, we can safely say that the correlation length must be less than the horizon distance— this statement assumes only that the correlation of the Higgs field requires the transmission of information, and special relativity implies that information cannot propagate faster than the speed of light. For a radiation-dominated universe, the horizon distance is given by 2ct, where t is the time. Thus,

$$n_M > \frac{1}{8c^3t^3} \; . \tag{12.4}$$

If we apply the above bound immediately after the phase transition, then we can re-express it in terms of the critical temperature of the phase transition by using the time-temperature relation of Eq. (7.62). The result is then

$$n_M > \left(\frac{4\pi^3 g G}{45\hbar^3 c^7}\right)^{3/2} (kT_c)^6 \ . \tag{12.5}$$

It will prove useful to compare this density to the number density of photons at the same temperature, which is given by Eq. (7.52):

$$n_\gamma = 0.24\frac{(kT)^3}{(\hbar c)^3} \ . \tag{12.6}$$

Taking the ratio and rounding off the numerical coefficient,

$$\frac{n_M}{n_\gamma} > 20\left(\frac{gG}{\hbar c^5}\right)^{3/2} (kT_c)^3 \ . \tag{12.7}$$

For a typical grand unified theory $g \approx 10^2$, and $kT_c \approx 10^{16}$ GeV, which gives

$$n_M/n_\gamma > 10^{-5} \ . \tag{12.8}$$

This may seem like a small density of monopoles, but because of their tremendous mass a density of this magnitude is intolerable.

Magnetic charge is conserved, but magnetic monopoles can still disappear by colliding with antimonopoles and annihilating. However, one can estimate the cross section for this process, and it is very small. The problem has been investigated by John Preskill (then at Harvard, currently at CalTech), who concluded that the ratio given in Eq. (12.8) would not be reduced significantly by the annihilation of monopoles. As the universe expands the number densities of both monopoles and photons will decrease rapidly, but the expansion of the universe will not affect their ratio. The ratio would be affected, however, by the photon production which takes place when the other types of particles "freeze out" as kT falls below mc^2. This effect will change the ratio by about one order of magnitude, which is not nearly enough to alter the conclusions. For simplicity I will therefore continue by assuming that Eq. (12.8) holds in the present universe.

For a temperature of 2.7 K today, the number density of photons is about 400 cm^{-3}. Given Eq. (12.8), one has a magnetic monopole density

$$n_M > 4 \times 10^{-3}\,\mathrm{cm}^{-3} \ . \tag{12.9}$$

Using a monopole mass given by $Mc^2 \approx 10^{18}$ GeV, the mass density of monopoles and antimonopoles in the universe becomes

$$\rho > 7 \times 10^{-9}\,\text{gm-cm}^{-3} \ . \tag{12.10}$$

For comparison, we found in Lecture Notes 4 that the critical mass density is given by $\rho_c \approx 9.7 \times 10^{-30}$ gm-cm^{-3}. Thus, the contribution of the magnetic monopoles to Ω is given by

$$\Omega_M > 7.2 \times 10^{20} \ , \tag{12.11}$$

which is clearly out of the question.

The easiest way to convince yourself that an Ω of 10^{21} cannot be accepted is to consider the age of the universe. As you recall, a large value of Ω implies that the universe slowed down rapidly to its present expansion rate, giving a low predicted age for the universe. The formula for the age of the universe was derived in Lecture Notes 5, and for large Ω (during the expanding phase) it is given by

$$t = \frac{\Omega}{2H(\Omega-1)^{3/2}}\left\{\sin^{-1}\left(\frac{2\sqrt{\Omega-1}}{\Omega}\right) - x\frac{2\sqrt{\Omega-1}}{\Omega}\right\} \ ,$$

where the inverse sine function is to be evaluated in the range $\frac{\pi}{2}$ to π. For very large Ω the inverse sine function approaches π, and the age is approximated by

$$t = \frac{\pi}{2H\sqrt{\Omega}} \ . \tag{12.12}$$

Inserting numbers, one gets an age of about one year. Numbers of this order are clearly unacceptable to anyone who wants to develop a scientific account of the evolution of the universe in which we are living. Thus, any successful union of grand unified theories and the big-bang picture must incorporate some mechanism to drastically suppress the production of magnetic monopoles.

The diligent reader will notice that the "Inflationary Universe" *Scientific American* article mentions a predicted age of 30,000 years, rather than the one-year figure obtained above. The main difference is the value assumed for the unification scale of grand unified theories. At the time the *Scientific American* article was written, this energy scale was believed to be about 10^{14} GeV, rather than the currently favored number of 10^{16} GeV. In any case the calculation is only a rough estimate.