8.286 Class 22
November 30, 2020

PROBLEMS OF THE CONVENTIONAL (NON-INFLATIONARY) HOT BIG BANG MODEL, PART 2,

and

GRAND UNIFIED THEORIES AND THE MAGNETIC MONOPOLE PROBLEM

Modified 12/27/20 to improve the discussion of electromagnetism as a gauge theory on pp. 25–26, and to mark the end of the slides reached in class.

Calendar for the Home Stretch:

NOVEMBER/DECEMBER				
MON	TUES	WED	THURS	FRI
23 **Thanksgiving Week**	24 —	25 —	26 —	27 —
30 Class 22	December 1	2 Class 23 **Quiz 3**	3	4
7 Class 24	8	9 Class 25 **PS 9 due** Last Class	10	11

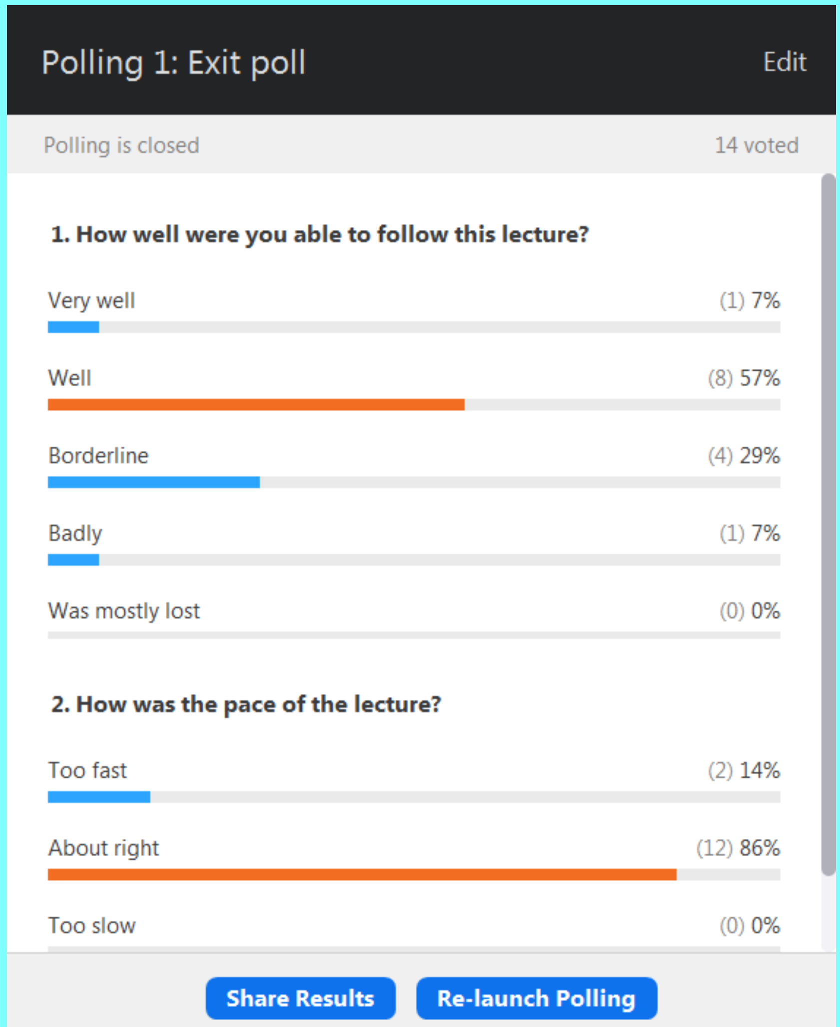
Announcements

- Quiz 3 will be this Wednesday, December 2!
- The coverage is described on the class webpage, and on the Review Problems for Quiz 3.
- If you want, you can start the quiz anytime from 11:05 am on Wednesday to 11:05 am on Thursday. If you want to start later than 11:05 am Wednesday, please send me an email by midnight Tuesday night.
- **Review Session:** this evening, 7:30 pm, run by Bruno Sheihing. Usual Zoom ID. If you have any problems or topics that you would particularly like Bruno to discuss, then email him!
- **Special office hours this week and next:**

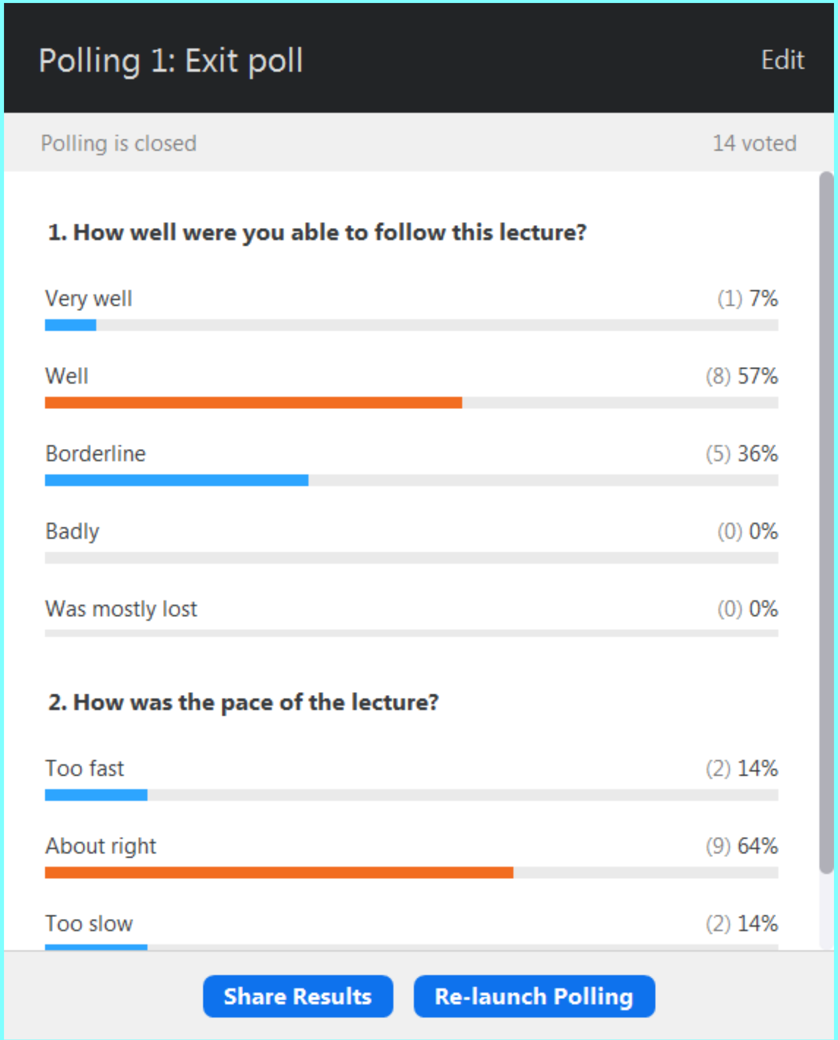
  Me: Mondays 11/30/20 and 12/7/20 at 5:00 pm.

  Bruno: Tuesdays 12/1/20 and 12/8/20 at 6:00 pm.
- There will be one last problem set, Problem Set 9, due the last day of classes, Wednesday December 9. **No final exam!**

Exit Poll, Class 20 (Class Before Last)



Exit Poll, Class 21 (Last Class)



Alan Guth
Massachusetts Institute of Technology

Evidence for the Accelerating Universe

1) Supernova Data: distant SN Ia are dimmer than expected by about 20–30%.

2) Cosmic Microwave Background (CMB) anisotropies: gives $\Omega_{\rm vac}$ close to SN value. Also gives $\Omega_{\rm tot} = 1$ to 1/2% accuracy, which cannot be accounted for without dark energy.

3) Inclusion of $\Omega_{\rm vac} \approx 0.70$ makes the age of the universe consistent with the age of the oldest stars.

☆ With the 3 arguments together, the case for the accelerating universe and $\Omega_{\rm dark\ energy} \approx 0.70$ has persuaded almost everyone.

☆ The simplest explanation for dark energy is vacuum energy, but "quintessence" — a slowly evolving scalar field — is also possible.

Review from last class:

Particle Physics of a Cosmological Constant

☆ $$\boxed{u_{\rm vac} = \rho_{\rm vac} c^2 = \frac{\Lambda c^4}{8\pi G}}$$

☆ Contributions to vacuum energy density:

1) Quantum fluctuations of the photon and other bosonic fields: positive and divergent.

2) Quantum fluctuations of the electron and other fermionic fields: negative and divergent.

3) Fields with nonzero values in the vacuum, like the Higgs field.

- ☆ If infinities are cut off at the Planck scale (quantum gravity scale), then infinities become finite, but

$>$ 120 orders of magnitude too large!

- ☆ For lack of a better explanation, many cosmologists (including Steve Weinberg and yours truly) seriously discuss the possibility that the vacuum energy density is determined by **"anthropic" selection effects:** that is, maybe there are many types of vacuum (as predicted by string theory), with different vacuum energy densities, with most vacuum energy densities roughly 120 orders of magnitude larger than ours. Maybe we live in a very low energy density vacuum because that is where almost all living beings reside. A large vacuum energy density would cause the universe to rapidly fly apart (if positive) or implode (if negative), so life could not form.

The Horizon/Homogeneity Problem

- ☆ General question: how can we explain the large-scale uniformity of the universe?
- ☆ Possible answer: maybe the universe just started out uniform.
  - There is no argument that excludes this possibility, since we don't know how the universe came into being.
  - However, if possible, it seems better to explain the properties of the universe in terms of things that we can understand, rather than to attribute them to things that we don't understand.

Review from last class:

The Cosmic Microwave Background

- ☆ The strongest evidence for the uniformity of the universe comes from the CMB, since it has been measured so precisely.
- ☆ The radiation appears slightly hotter in one direction than in the opposite direction, by about one part in a thousand — but this nonuniformity can be attributed to our motion through the background radiation.
- ☆ Once this effect is subtracted out, using best-fit parameters for the velocity, it is found that the residual temperature pattern is uniform to a few parts in 10^5.
- ☆ Could this be simply the phenomenon of thermal equilibrium? If you put an ice cube on the sidewalk on a hot summer day, it melts and come sto the same temperature as the sidewalk.
- ☆ **BUT: in the conventional model of the universe, it did not have enough time for thermal equilibrium to explain the uniformity, if we assume that it did not start out uniform. If no matter, energy, or information can travel faster than light, then it is simply not possible.**

Review from last class:

Basic History of the CMB

- ☆ In conventional cosmological model, the universe at the earliest times was radiation-dominated. It started to be matter-dominated at $t_{\rm eq} \approx 50,000$ years, the time of matter-radiation equality.
- ☆ **At the time of *decoupling* $t_d \approx 380,000$ years, the universe cooled to about 3000 K, by which time the hydrogen (and some helium) combined so thoroughly that free electrons were very rare. At earlier times, the universe was in a mainly plasma phase, with many free electrons, and photons were essentially frozen with the matter. At later times, the universe was transparent, so photons have traveled on straight lines. We can say that the CMB was released at 380,000 years.**
- ☆ Since the photons have been mainly traveling on straight lines since $t = t_d$, they have all traveled the same distance. Therefore **the locations from which they were released form a sphere centered on us. This sphere is called the *surface of last scattering*,** since the photons that we receive now in the CMB was mostly scattered for the last time on or very near this surface.

Horizon Calculations

- Temperature at decoupling $T_d \approx 3000$ K. This implies the time of decoupling $t_d \approx 380,000$ yr.
- For a flat, matter-dominated universe, the horizon distance is $\ell_h(t_d) = 3ct_d \approx 1,100,000$ light-years.
- To find the radius of the surface of last-scattering at t_d, we found its radius today from the redshift $1 + z = 3000\ \mathrm{K}/2.7\ \mathrm{K}$, and then reduced it by $a(t_0)/a(t_d) = 1 + z$.
- Conclusion: the radius of the surface of last scattering, at the time t_d, was about 23 times the horizon distance.

Review from last class:

Summary of the Horizon Problem

Suppose that one detects the cosmic microwave background in a certain direction in the sky, and suppose that one also detects the radiation from precisely the opposite direction. At the time of emission, the sources of these two signals were separated from each other by about 46 horizon distances. Thus it is absolutely impossible, within the context of this model, for these two sources to have come into thermal equilibrium by any physical process.

Although our calculation ignored the dark energy phase, we have found in previous examples that such calculations are wrong by some tens of a percent. (For example we found $t_{\text{eq}} \approx 75,000$ years, when it should have been about 50,000 years.) Since $46 \gg 1$, there is no way that a more accurate calculation could cause this problem to go away.

The Flatness Problem

- A second problem of the conventional cosmological model is the *flatness problem*: why was the value of Ω in the early universe so extraordinarily close to 1?

- Today we know, according to the Planck satellite team analysis (2018), that

$$\boxed{\Omega_0 = 0.9993 \pm 0.0037}$$

at 95% confidence. I.e., $\Omega = 1$ to better than 1/2 of 1%.

- As we will see, this implies that Ω in the early universe was extaordinarily close to 1. For example, at $t = 1$ second,

$$\boxed{|\Omega - 1|_{t=1 \text{ sec}} < 10^{-18} \ .}$$

- The underlying fact is that the value $\Omega = 1$ is a point of unstable equilibrium, something like a pencil balancing on its point. If Ω is ever ***exactly*** equal to one, it will remain equal to one forever — that is, a flat ($k = 0$) universe remains flat. However, if Ω is ever slightly larger than one, it will rapidly grow toward infinity; if Ω is ever slightly smaller than one, it will rapidly fall toward zero. For Ω to be anywhere near 1 today, Ω in the early universe must have been extraordinarily close to one.
- Like the horizon problem, the flatness problem could in principle be solved by the initial conditions of the universe: maybe the universe began with $\Omega \equiv 1$.
    - But, like the horizon problem, it seems better to explain the properties of the universe, if we can, in terms of things that we can understand, rather than to attribute them to things that we don't understand.

History of the Flatness Problem

The mathematics behind the flatness problem was undoubtedly known to almost anyone who has worked on the big bang theory from the 1920's onward, but apparently the first people to consider it a problem in the sense described here were Robert Dicke and P.J.E. Peebles, who published a discussion in 1979.*

*R.H. Dicke and P.J.E. Peebles, "The big bang cosmology — enigmas and nostrums," in **General Relativity: An Einstein Centenary Survey**, eds: S.W. Hawking and W. Israel, Cambridge University Press (1979).

The Mathematics of the Flatness Problem

Start with the first-order Friedmann equation:

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \; .$$

Remembering that $\Omega = \rho/\rho_c$ and that $\rho_c = 3H^2/(8\pi G)$, one can divide both sides of the equation by H^2 to find

$$1 = \frac{\rho}{\rho_c} - \frac{kc^2}{a^2 H^2} \quad \Longrightarrow \quad \boxed{\Omega - 1 = \frac{kc^2}{a^2 H^2} \; .}$$

Evolution of $\Omega - 1$ During the Radiation-Dominated Phase

$$\boxed{\Omega - 1 = \frac{kc^2}{a^2 H^2} \ .}$$

For a (nearly) flat radiation-dominated universe, $a(t) \propto t^{1/2}$, so $H = \dot{a}/a = 1/(2t)$. So

$$\boxed{\Omega - 1 \propto \left(\frac{1}{t^{1/2}}\right)^2 \left(\frac{1}{t^{-1}}\right)^2 \propto t \quad \text{(radiation dominated).}}$$

Evolution of $\Omega - 1$ During the Matter-Dominated Phase

$$\Omega - 1 = \frac{kc^2}{a^2 H^2} \ .$$

For a (nearly) flat matter-dominated universe, $a(t) \propto t^{2/3}$, so $H = \dot{a}/a = 2/(3t)$. So

$$\Omega - 1 \propto \left(\frac{1}{t^{2/3}}\right)^2 \left(\frac{1}{t^{-1}}\right)^2 \propto t^{2/3} \quad \text{(matter-dominated).}$$

Tracing $\Omega - 1$ from Now to 1 Second

Today,

$$|\Omega_0 - 1| < .01 \ .$$

I will do a crude calculation, treating the universe as matter dominated from 50,000 years to the present, and as radiation-dominated from 1 second to 50,000 years.

During the matter-dominated phase,

$$(\Omega - 1)_{t=50{,}000\ \mathrm{yr}} \approx \left(\frac{50{,}000}{13.8 \times 10^9}\right)^{2/3} (\Omega_0 - 1) \approx 2.36 \times 10^{-4}\,(\Omega_0 - 1) \ .$$

$$|\Omega_0 - 1| < .01 \ .$$

$$(\Omega - 1)_{t=50{,}000 \text{ yr}} \approx \left(\frac{50{,}000}{13.8\times 10^9}\right)^{2/3} (\Omega_0 - 1) \approx 2.36 \times 10^{-4}\,(\Omega_0 - 1) \ .$$

During the radiation-dominated phase,

$$\begin{aligned}(\Omega - 1)_{t=1 \text{ sec}} &\approx \left(\frac{1 \text{ sec}}{50{,}000 \text{ yr}}\right) (\Omega - 1)_{t=50{,}000 \text{ yr}} \\ &\approx 1.49 \times 10^{-16}\,(\Omega_0 - 1) \ .\end{aligned}$$

The conclusion is therefore

$$|\Omega - 1|_{t=1 \text{ sec}} < 10^{-18} \ .$$

The conclusion is therefore

$$\boxed{\left|\Omega - 1\right|_{t=1\ \mathrm{sec}} < 10^{-18}\ .}$$

Even if we put ourselves mentally back into 1979, we would have said that $0.1 < \Omega_0 < 2$, so $|\Omega_0 - 1| < 1$, and would have concluded that

$$\boxed{\left|\Omega - 1\right|_{t=1\ \mathrm{sec}} < 10^{-16}\ .}$$

The Dicke & Peebles paper, that first pointed out this problem, also considered $t = 1$ second, but concluded (without showing the details) that

$$\boxed{\left|\Omega - 1\right|_{t=1\ \mathrm{sec}} < 10^{-14}\ .}$$

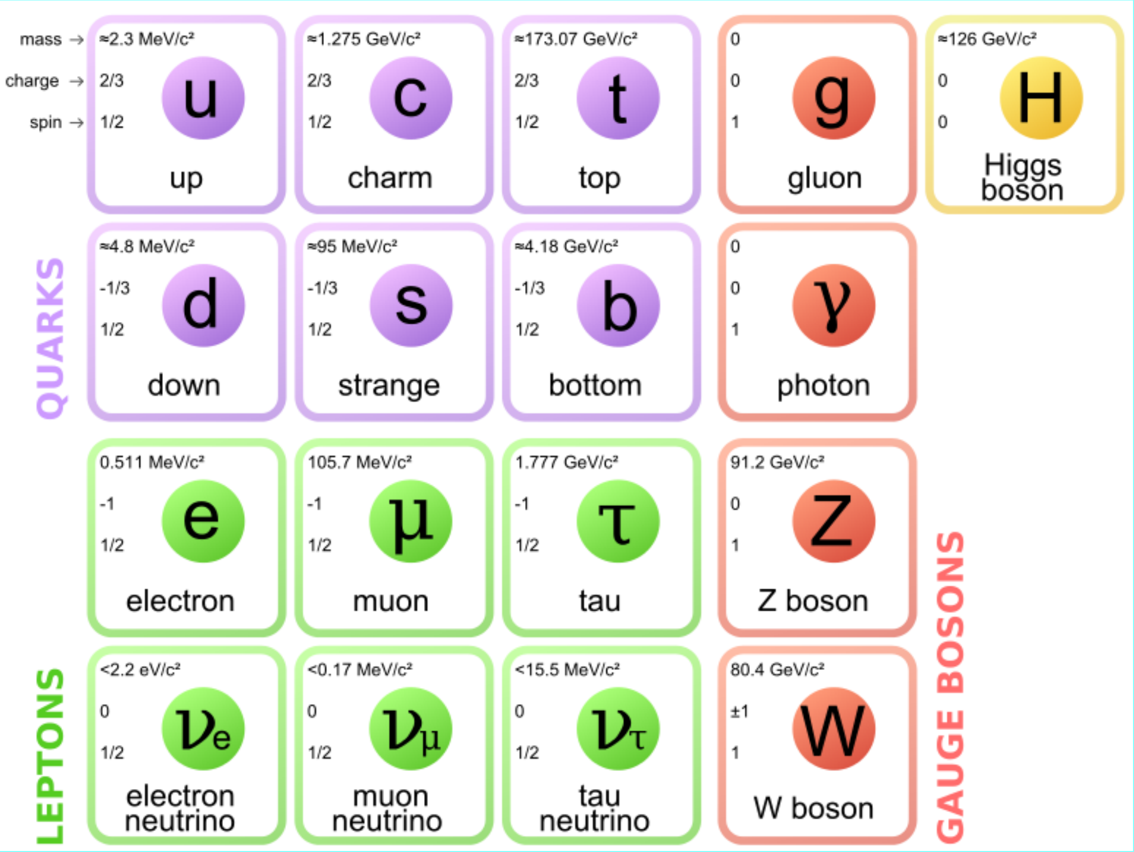
They were perhaps more conservative, but concluded nonetheless that this extreme fine-tuning cried out for an explanation.

8.286 Class 22
November 30, 2020

GRAND UNIFIED THEORIES
AND THE
MAGNETIC MONOPOLE PROBLEM

The Standard Model of Particle Physics

Particle Content:



Wikimedia Commons. Source: PBS NOVA, Fermilab, Office of Science, United States Department of Energy, Particle Data Group.

Quarks are Colored

- A quark is specified by its *flavor* [u(p), d(own), c(harmed), s(trange), t(op), b(ottom)], its spin [up or down, along any chosen z axis], whether it is a quark or antiquark, AND ITS COLOR [three choices, often red, blue, or green].
- Quarks that differ only in color are completely indistinguishable, but the color is relevant for the Pauli exclusion principle: one can't have 3 identical quarks all in the lowest energy state, but one can have one red quark, one blue quark, and one green quark.
- Color is also relevant for the way quarks interact. The colors act like a generalized form of electric charge. Two red quarks interact with each other exactly the same way as two blue quarks, but a red quark and a blue quark interact with each other differently.

Gauge Theories: Electromagnetic Example

Fields and potentials*: $\vec{E} = -\vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t}$, $\vec{B} = \vec{\nabla} \times \vec{A}$.

Four-vector notation: $A_\mu = \left(-\frac{\phi}{c}, A_i\right)$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$,

$E_i = cF_{i,0}$, $B_i = \frac{1}{2}\epsilon_{ijk}F_{jk}$.

Gauge transformations:

$$\phi'(t,\vec{x}) = \phi(t,\vec{x}) - \frac{\partial \Lambda(t,\vec{x})}{\partial t} \ , \quad \vec{A}'(t,\vec{x}) = \vec{A}(t,\vec{x}) + \vec{\nabla}\Lambda(t,\vec{x}) \ ,$$

or in four-vector notation,

$$A'_\mu(x) = A_\mu(x) + \frac{\partial \Lambda}{\partial x^\mu} \ , \ \text{where } x^\mu \equiv (ct, \vec{x}).$$

$\vec{E}$ and $\vec{B}$ are gauge-invariant (i.e., are unchanged by a gauge transformation):

$$\vec{B}' = \vec{\nabla} \times \vec{A}' = \vec{\nabla} \times (\vec{A} + \vec{\nabla}\Lambda) = \vec{\nabla} \times \vec{A} = \vec{B} \ ,$$

*Using the conventions of D.J. Griffiths, *Introduction to Electrodynamics*, Fourth Edition.

$$\phi'(t,\vec{x}) = \phi(t,\vec{x}) - \frac{\partial\Lambda(t,\vec{x})}{\partial t} \ , \quad \vec{A}'(t,\vec{x}) = \vec{A}(t,\vec{x}) + \vec{\nabla}\Lambda(t,\vec{x}) \ ,$$

$\vec{E}$ and $\vec{B}$ are gauge-invariant (i.e., are unchanged by a gauge transformation):

$$\vec{B}' = \vec{\nabla} \times \vec{A}' = \vec{\nabla} \times (\vec{A} + \vec{\nabla}\Lambda) = \vec{\nabla} \times \vec{A} = \vec{B} \ ,$$

$$\vec{E}' = -\vec{\nabla}\phi' - \frac{\partial \vec{A}'}{\partial t} = -\vec{\nabla}\left(\phi - \frac{\partial\Lambda}{\partial t}\right) - \frac{\partial}{\partial t}\left(\vec{A} + \vec{\nabla}\Lambda\right)$$

$$= -\vec{\nabla}\phi - \frac{\partial\vec{A}}{\partial t} = \vec{E} \ ,$$

where we used $\vec{\nabla} \times \vec{\nabla}\Lambda \equiv 0$ and $\vec{\nabla}\left(\frac{\partial\Lambda}{\partial t}\right) = \frac{\partial}{\partial t}\vec{\nabla}\Lambda$. So A_μ and A'_μ both satisfy the equations of motion, and describe the SAME physical situation.

Gauge transformations can be combined, forming a group:

$$\Lambda_3(x) = \Lambda_1(x) + \Lambda_2(x) \ .$$

Gauge symmetries are also called local symmetries, since the gauge function $\Lambda(x)$ is an arbitrary function of position and time.

Electromagnetism as a $U(1)$ Gauge Theory

$\Lambda(x)$ is an element of the real numbers.

But if we included an electron field $\psi(x)$, it would transform as

$$\psi'(x) = e^{ie_0\Lambda(x)}\psi(x) \ ,$$

where e_0 is the charge of a proton and $e = 2.71828\ldots$. So we might think of $u(x) \equiv e^{ie_0\Lambda(x)}$ as describing the gauge transformation. u contains LESS information than Λ, since it defines Λ only mod $2\pi/e_0$.

But u is enough to define the gauge transformation, since

$$\frac{\partial\Lambda}{\partial x^\mu} = \frac{1}{ie_0}\, e^{-ie_0\Lambda(x)}\, \frac{\partial}{\partial x^\mu}\, e^{ie_0\Lambda(x)} \ .$$

u is an element of the group $U(1)$, the group of complex phases $u = e^{i\chi}$, where χ is real. So E&M is a $U(1)$ gauge theory.

Gauge Groups of the Standard Model

$U(1)$ is abelian (commutative), but Yang and Mills showed in 1954 how to construct a nonabelian gauge theory. The standard model contains the following gauge symmetries:

$\boldsymbol{SU(3)}$: This is the group of 3×3 complex matrices that are

S $\equiv$ Special: they have determinant 1.

U $\equiv$ Unitary: they obey $u^\dagger u = 1$, which means that when they multiply a 1×3 column vector v, they preserve the norm $|v| \equiv \sqrt{v_i^* v_i}$.

$\boldsymbol{SU(2)}$: The group of 2×2 complex matrices that are special (S) and unitary (U). As you may have learned in quantum mechanics, $SU(2)$ is almost the same as the rotation group in 3D, with a 2:1 group-preserving mapping between $SU(2)$ and the rotation group.

$\boldsymbol{U(1)}$: The group of complex phases. The $U(1)$ of the standard model is not the $U(1)$ of E&M; instead $U(1)_{\rm E\&M}$ is a linear combination of the $U(1)$ of the standard model and a rotation about one fixed direction in $SU(2)$.

Combining the groups: the gauge symmetry group of the standard model is usually described as $SU(3)\times SU(2)\times U(1)$. An element of this group is an ordered triplet (u_3, u_2, u_1), where $u_3 \in SU(3)$, $u_2 \in SU(2)$, and $u_1 \in U(1)$, so $SU(3)\times SU(2)\times U(1)$ is really no different from thinking of the 3 symmetries separately.

$SU(3)$ describes the strong interactions, and $SU(2) \times U(1)$ together describe the electromagnetic and weak interactions in a unified way, called the *electroweak interactions*.

$SU(3)$ acts on the quark fields by rotating the 3 "colors" into each other. Thus the strong interactions of the quarks are entirely due to their "colors", which act like charges.

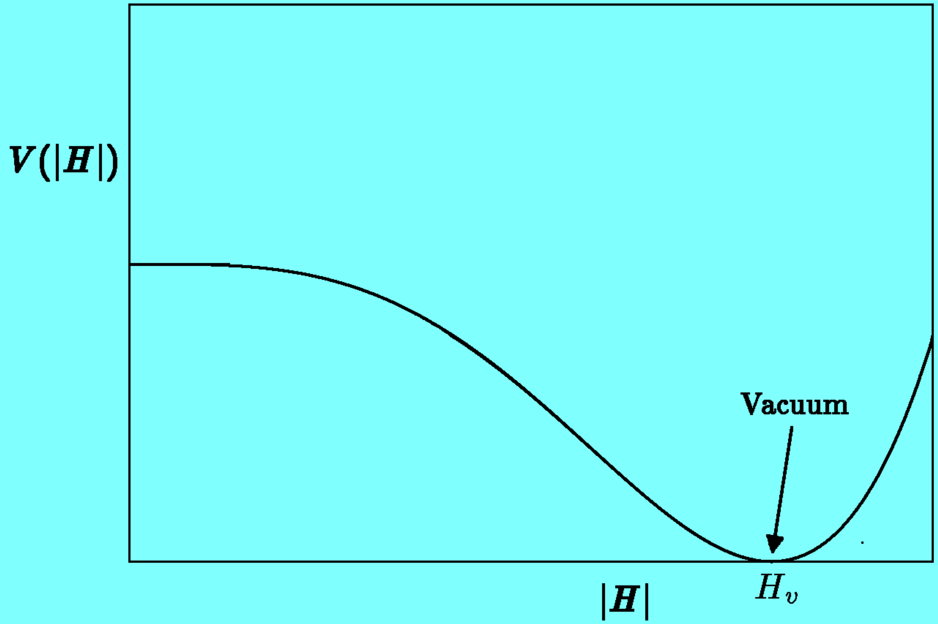
The Higgs Field and Spontaneous Symmetry Breaking

The Higgs field is a complex doublet:

$$H(x) \equiv \begin{pmatrix} h_1(x) \\ h_2(x) \end{pmatrix} \ .$$

Under $SU(2)$ transformations, $H'(x) = u_2(x)H(x)$, where $u_2(x)$ is the complex 2×2 matrix that defines the $SU(2)$ gauge transformation at x. Since the gauge symmetry implies that the potential energy density of the Higgs field $V(H)$ must be gauge-invariant, V can depend only on $|H| \equiv \sqrt{|h_1|^2 + |h_2|^2}$, which is unchanged by $SU(2)$ transformations.

Potential energy function $V(|H|)$:



The minimum is not at $|H| = 0$, but instead at $|H| = H_v$.

$|H| = 0$ is $SU(2)$ gauge-invariant, but $|H| = H_v$ is not. H randomly picks out some direction in the space of 2D complex vectors.

Spontaneous Symmetry Breaking: Whenever the ground state of a system has less symmetry than the underlying laws, it is called spontaneous symmetry breaking. Examples: crystals, ferromagnetism.

This and the following slides were not reached, but will be discussed in the next class.

Higgs Fields Give Mass to Other Particles

When $H = 0$, all the fundamental particles of the standard model are massless. Furthermore, there is no distinction between the electron e and the electron neutrino ν_e, or between μ and ν_μ, or between τ and ν_τ. (Protons, however, would not be massless — intuitively, most of the proton mass comes from the gluon field that binds the quarks.)

For $|H| \neq 0$, H randomly picks out a direction in the space of 2D complex vectors. Since all directions are otherwise equivalent, we can assume that in the vacuum,

$$H = \begin{pmatrix} H_v \\ 0 \end{pmatrix} .$$

Components of other fields that interact with $\mathrm{Re}\,(h_1)$ then start to behave differently from fields that interact with other components of H.

Mass: mc^2 of a particle is the state of lowest energy above the ground state. In a field theory, this corresponds to a homogeneous oscillation of the field, which in turn corresponds to a particle with zero momentum.

In the free field limit, the field acts exactly like a harmonic oscillator. The first excited state has energy $h\nu = \hbar\omega$ above the ground state. So, $mc^2 = \hbar\omega$.

ω is determined by inertia and the restoring force. When $H = 0$, the standard model interactions provide no restoring forces. Any such restoring force would break gauge invariance.

When $H = \begin{pmatrix} H_v \\ 0 \end{pmatrix}$, the interactions with H creates a restoring force for some components of other fields, giving them a mass. This "Higgs mechanism" creates the distinction between electrons and neutrinos — the electrons are the particles that get a mass, and the neutrinos do not. (Neutrinos are exactly massless in the Standard Model of Particle Physics. There are various ways to modify the model to account for neutrino masses.)

Beyond the Standard Model

With neutrino masses added, the standard model is spectacularly successful: it agrees with all reliable particle experiments.

Nonetheless, most physicists regard it as incomplete, for at least two types of reasons:

1) It does not include gravity, and it does not include any particle to account for the dark matter. (Maybe black holes can do it, but that requires a mass distribution that we cannot explain.)

2) The theory appears too inelegant to be the final theory. It contains more arbitrary features and free parameters than one would hope for in a final theory. Why $SU(3)\times SU(2)\times U(1)$? Why three generations of fermions? The original theory had 19 free parameters, with more needed for neutrino masses and even more if *supersymmetry* is added.

Result: BSM (Beyond the Standard Model) particle physics has become a major industry.

Grand Unified Theories

Goal: Unify $SU(3)\times SU(2)\times U(1)$ by embedding all three into a single, larger group.

The breaking of the symmetry to $SU(3)\times SU(2)\times U(1)$ is accomplished by introducing new Higgs fields to spontaneously break the symmetry.

In the fundamental theory, before spontaneous symmetry breaking, there is no distinction between an electron, an electron neutrino, or an up or down quark.

The SU(5) Grand Unified Theory

In 1974, Howard Georgi and Sheldon Glashow of Harvard proposed the original grand unified theory, based on $SU(5)$. They pointed out that $SU(3)\times SU(2)\times U(1)$ fits elegantly into $SU(5)$.

To start, let the $SU(3)$ subgroup be matrices of the form

$$u_3 = \begin{pmatrix} x & x & x & 0 & 0 \\ x & x & x & 0 & 0 \\ x & x & x & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} ,$$

where the 3×3 block of x's represents an arbitrary SU(3) matrix.

Similarly let the $SU(2)$ subgroup be matrices of the form

$$u_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x & x \end{pmatrix} ,$$

where this time the 2×2 block of x's represents an arbitrary SU(2) matrix.

Note that u_3 and u_2 commute, since each acts like the identity matrix in the space in which the other is nontrivial.

Finally, the $U(1)$ subgroup can be written as

$$u_1 = \begin{pmatrix} e^{2i\theta} & 0 & 0 & 0 & 0 \\ 0 & e^{2i\theta} & 0 & 0 & 0 \\ 0 & 0 & e^{2i\theta} & 0 & 0 \\ 0 & 0 & 0 & e^{-3i\theta} & 0 \\ 0 & 0 & 0 & 0 & e^{-3i\theta} \end{pmatrix} ,$$

where the factors of 2 and 3 in the exponents were chosen so that the determinant — in this case the product of the diagonal entries — is equal to 1.

Repeating, the $U(1)$ subgroup can be written as

$$u_1 = \begin{pmatrix} e^{2i\theta} & 0 & 0 & 0 & 0 \\ 0 & e^{2i\theta} & 0 & 0 & 0 \\ 0 & 0 & e^{2i\theta} & 0 & 0 \\ 0 & 0 & 0 & e^{-3i\theta} & 0 \\ 0 & 0 & 0 & 0 & e^{-3i\theta} \end{pmatrix} .$$

---

u_1 commutes with any matrix of the form of u_2 or u_3, since within either the upper 3×3 block or within the lower 2×2 block, u_1 is proportional to the identity matrix.

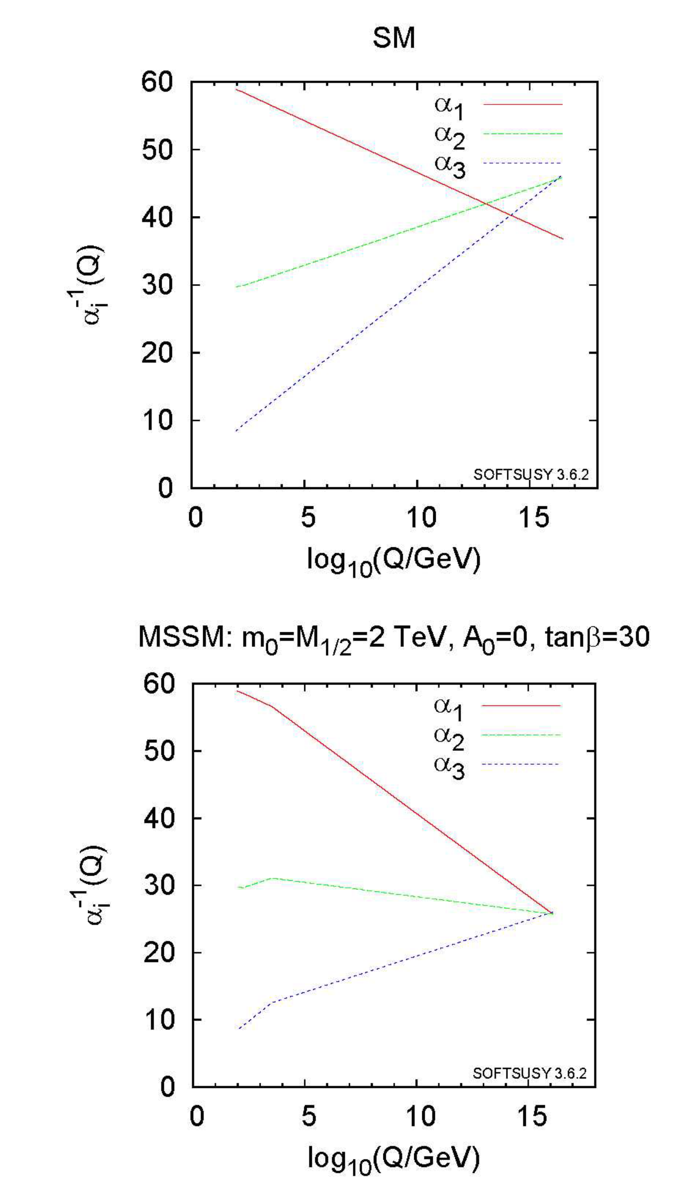
Thus, any element (u_3, u_2, u_1) of $SU(3) \times SU(2) \times U(1)$ can be written as an element u_5 of $SU(5)$, just by setting $u_5 = u_3 u_2 u_1$.

How Can Three Different Types of Interaction Look Like One?

In the standard model, each type of gauge interaction — $SU(3)$, $SU(2)$, and $U(1)$ — has its own interaction strength, described by "coupling constants" g_3, g_2, and g_1. Their values of are different from each other! How can they be one interaction?

BUT: the interaction strength varies with energy in a calculable way. When the calculations are extended to superhigh energies, of the order of 10^{16} GeV, the three interaction strengths become about equal!

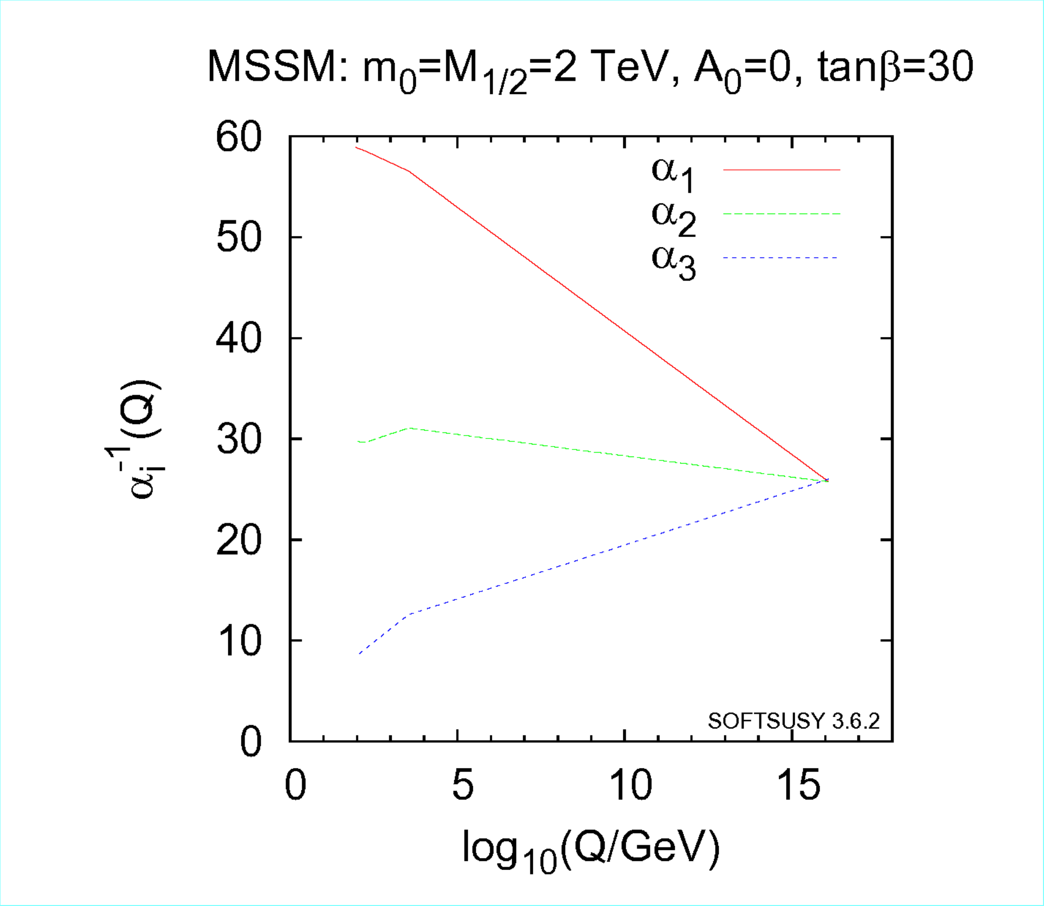
Running Coupling Constants



The top graph shows the running of coupling constants in the standard model, showing that the three coupling constants do not quite meet. The bottom graph shows the running of coupling constants in the MSSM — the Minimal Supersymmetric Standard Model, in which the meeting of the couplings is almost perfect. $\alpha_i = g_i^2/4\pi$.

Source: Particle Data Group 2016 Review of Particle Physics, Chapter 16, *Grand Unified Theories*, Revised January 2016 by A. Hebecker and J. Hisano.

Running Couplings
Minimal Supersymmetric Standard Model



Bottom line: An $SU(5)$ grand unified theory can be constructed by introducing a Higgs field that breaks the $SU(5)$ symmetry to $SU(3)\times SU(2)\times U(1)$ at an energy of about 10^{16} GeV. At energies above 10^{16} GeV, the theory behaves like a fully unified $SU(5)$ gauge theory. At lower energies, it behaves like the standard model. The gauge particles that are part of $SU(5)$ but not part of $SU(3)\times SU(2)\times U(1)$ acquire masses of order 10^{16} GeV.

GUTs (Grand Unified Theories) allow two unique phenomena at low energies, neither of which have been seen:

1) Proton decay. The superheavy gauge particles can mediate proton decay. The minimal $SU(5)$ model — with the simplest conceivable particle content — predicts a proton lifetime of about 10^{31} years, which is ruled out by experiments, which imply a lifetime $\gtrsim 10^{34}$ years.

2) Magnetic monopoles. All grand unified theories imply that magnetic monopoles should be a possible kind of particle. None have been seen.

The absence of evidence does not imply that GUTs are wrong, but we don't know.

The Grand Unified Theory Phase Transition

When $kT \gg 10^{16}$ GeV, the Higgs fields of the GUT undergo large fluctuations, and average to zero. The GUT symmetry is unbroken, and the theory behaves as an SU(5) gauge theory.

As kT falls to about 10^{16} GeV, the matter filling the universe would go through a phase transition, in which some of the components of the GUT Higgs field acquire nonzero values in the thermal equilibrium state, breaking the GUT symmetry. The breaking to $SU(3)\times SU(2)\times U(1)$ might occur in one phase transition, or in a series of phase transitions. We'll assume a single phase transition.

The Higgs fields start to randomly acquire nonzero values, but the nonzero values that form in one region may not align with those in another.

The expression for the energy density contains a term proportional to $|\nabla\Phi|^2$, so the fields tend to fall into low energy states with small gradients. But sometimes the fields in one region acquire a pattern that cannot be smoothly joined with the pattern in a neighboring region, so the smoothing is imperfect, leaving "defects".

Topological Defects

There are three types of defects:

1) Domain walls. For example, imagine a single real scalar field ϕ for which the potential energy function has two local minima, at ϕ_1 and ϕ_2. Then, as the system cools, some regions will have $\phi \approx \phi_1$ and others will have $\phi \approx \phi_2$. The boundaries between these regions will be surfaces of high energy density: domain walls. Some GUTS allow domain walls, others do not. The energy density of a domain wall is so high that none can exist in the visible universe.

2) Cosmic strings. Linelike defects, which exist in some GUTs but not all.

3) Magnetic monopoles: Pointlike defects, which exist in all GUTs.

Magnetic Monopoles

We'll consider the simplest theory in which monopoles arise. It has a 3-component (real) Higgs field, ϕ_a, where $a = 1, 2$ or 3. Gauge symmetry acting on ϕ_a has the same mathematical form as the rotations of an ordinary Cartesian 3-vector.

To be gauge-invariant, the energy density function can depend only on

$$|\phi| \equiv \sqrt{\phi_1^2 + \phi_2^2 + \phi_3^2} \ ,$$

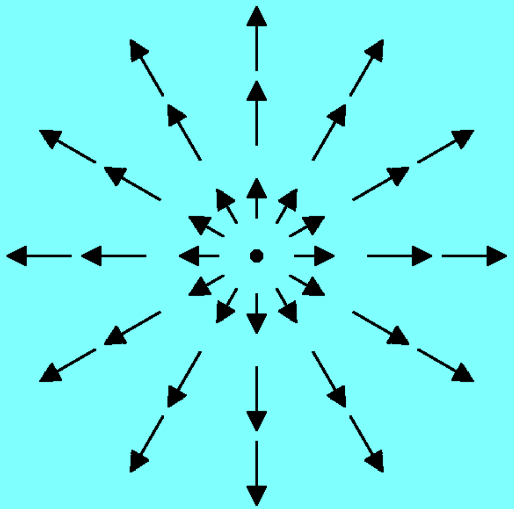
and we assume that it looks qualitatively like the graph for the standard model, with a minimum at H_v.

Now consider the following static configuration,

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

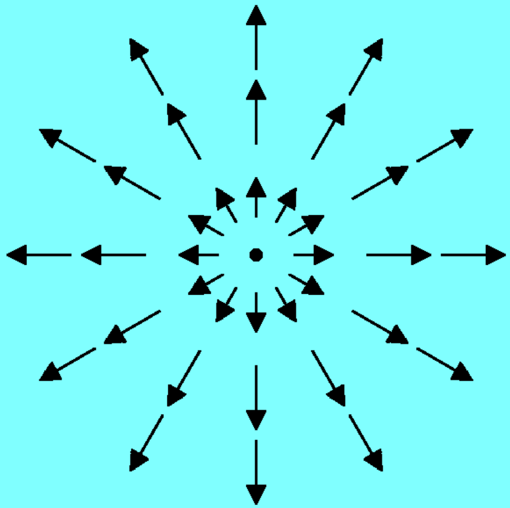
where $r \equiv |\vec{r}\,|$, $\hat{r}_a$ denotes the a-component of the unit vector $\hat{r} = \vec{r}/r$, and $f(r)$ is a function which vanishes when $r = 0$ and approaches H_v as $r \to \infty$.

Pictorially,



where the 3 components of the arrow at each point describe the 3 Higgs field components.

Repeating,



where the 3 components of the arrow at each point describe the 3 Higgs field components.

The directions in gauge space ϕ_a really have nothing to do with directions in physical space, but there is nothing that prevents the fields from existing in this configuration.

The configuration is topologically stable in the following sense: if the boundary conditions at infinity are fixed, and the fields are continuous, then there is always at least one point where $\phi_1 = \phi_2 = \phi_3 = 0$.

Thus, the monopoles are topologically stable knots in the Higgs field.

Why Are These Things Magnetic Monopoles?

Definition: A magnetic monopole is an object with a net magnetic charge, north or south, with a radial magnetic field of the same form as the electric field of a point charge.

Known magnets are always dipoles, with a north end and a south end. If such a magnet is cut in half, one gets two dipoles, each with a north and south end.

Energy of the configuration: the energy density contains a term $\sum_a \vec{\nabla}\phi_a \cdot \vec{\nabla}\phi_a$, but the changing direction of ϕ_a (always radially outward) implies $|\nabla\phi_a| \propto 1/r$. The total energy in a sphere of radius R is proportional to

$$4\pi \int^R r^2 \mathrm{d}r \left(\frac{1}{r}\right)^2 \ ,$$

which diverges as R for large R.

With the vector gauge fields, however, the energy density is more complicated. It can be made finite only if the gauge field configuration corresponds to a net magnetic charge.

Prediction of Magnetic Charge

The magnetic charge is uniquely determined, and is equal to $1/(2\alpha)$ times the electric charge of an electron, where $\alpha \simeq 1/137$ (α = fine-structure constant $= e^2/\hbar c$ in Gaussian units, or $e^2/(4\pi\epsilon_0\hbar c)$ in SI.)

The force between two monopoles is therefore $(68.5)^2$ times as large as the force between two electrons at the same distance. I.e., large!

Kibble Estimate of Magnetic Monopole Production

Since magnetic monopoles are knots in the GUT Higgs fields, they form at the GUT phase transition, when the Higgs fields acquire nonzero mean values. ("Mean" = average over time, not space.)

The density of these knots will be related to the misalignment of the Higgs field in different regions.

Define a correlation length ξ, crudely, as the minimum distance such that the Higgs field at point is almost uncorrelated with the Higgs field a distance ξ away.

T.W.B. Kibble of Imperial College (London) proposed that the number density of magnetic monopoles (and antimonopoles) can be estimated as

$$n_M \approx 1/\xi^3 \ .$$

Estimate of Correlation Length ξ

In the context of conventional (non-inflationary) cosmology, we can assume

1) that the Higgs field well before the GUT phase transition is in a thermal state, with no long-range correlations.

2) that the universe before the phase transition is well-approximated by a flat radiation-dominated Friedman-Robertson-Walker description.

3) phase transition happens promptly when the temperature of the GUT phase transition is reached, at $kT \approx 10^{16}$ GeV.

Under these assumptions, we are confident that the correlation length ξ must be less than or equal to the horizon length at the time of the phase transition. This seemingly mild limit turns out to have huge implications.

On Problem Set 9, you will calculate the contribution to Ω today, from the monopoles. I won't give away the answer, but you should find that it is greater than 10^{20}.