Fermi National Accelerator Laboratory

FERMILAB-CONF-86/139-T
October, 1986

ELEMENTARY PARTICLE PHYSICS: DISCOVERIES, INSIGHTS, AND TOOLS

CHRIS QUIGG
Fermi National Accelerator Laboratory
P. O. Box 500, Batavia, Illinois 60510

Presented at the Conference on the Teaching of Modern Physics, held at Fermilab, April, 1986.

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

ELEMENTARY PARTICLE PHYSICS: DISCOVERIES, INSIGHTS, AND TOOLS[1]

CHRIS QUIGG
Fermi National Accelerator Laboratory
P. O. Box 500, Batavia, Illinois 60510

ABSTRACT

This is a lightly edited transcript of two lectures presented at the Conference on the Teaching of Modern Physics held at Fermilab in April, 1986. The informality of the spoken word has been preserved, but some of the immediacy of the interchange with the audience is inevitably lost.

LECTURE 1: THE FUNDAMENTAL CONSTITUENTS

What I would like to talk to you about this morning is ELEMENTARY PARTICLE PHYSICS, the science of the ultimate constituents of matter and the interactions among them. Like all of physics (but in an especially immediate manner), it tries to ask and answer the questions

- What is the world made of?
- How does the world work?

In common with other physicists, we hope that by beginning to understand the laws of Nature, by codifying them, by extending the domain over which they apply, we may be able to put our new knowledge to productive use.

The questions that we pose for ourselves (see Fig. 1) are

- What are the basic constituents of matter and energy?
- What are the forces by which these constituents interact with each other?

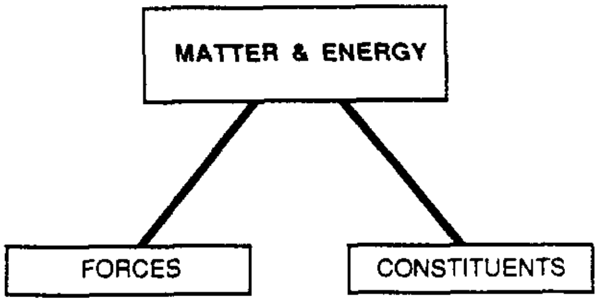


Figure 1: Goals of elementary particle physics.

What I will try to do in these two talks is to introduce you to the description of matter and energy to which we have come, and to emphasize both the simplicity and the tentativeness of that description. In the course of this, we will fill in some of the white space in Fig. 1. In a sense, it is easy to do this. It is easy because dramatic progress has been made over the last twenty years. The picture we have of fundamental physics is much simpler, much more comprehensive, and much more unified than it was a couple of decades ago. This has prompted some to say that a grand synthesis of natural law is at hand. It is unquestionably true that great progress has been made, and that the place at which we have arrived is at least a good starting point for the next great leap.

The reason we can explain our world view to students in relatively simple terms has to do with the emergence of something called THE STANDARD MODEL OF ELEMENTARY PARTICLE PHYSICS. The point of my lectures this morning will be to illustrate for you some of the prominent features of the Standard Model.

The Standard Model has a couple of aspects that I want to emphasize. One is the identification of a set of *elementary particles,* at least for our generation of scientists, called the *quarks* and the *leptons.* I'll spend much of this first lecture reminding you of some of the features of those constituents. On the other side of our chart, in trying to understand the interactions of those constituents, there has been the recognition of a grand principle and the development of a class of theories called gauge theories of the strong, weak, and electromagnetic interactions. I'll try to indicate to you in the beginning of the second lecture what is the strategy of gauge theories. We won't go through all the mathematical details of gauge theories, but as with most wonderful ideas, once someone has slogged through the details for you, you can explain it more or less simply, and I'll try to do that for you. Finally, the reason for the gleam in one's eye is that because of the simplicity of this picture, having identified the relatively small number of fundamental constituents and seen a nice mathematical framework in which to express their interactions, we see the promise of going further and gaining a more coherent understanding of all the forces of Nature. I'll try at the end of the second lecture to allude to that a little bit, and the thrust of where we go from here is what Howard Georgi is supposed to talk about in the next couple of days.[2]

Now, particularly because you are here at Fermilab, but also because I think it's important, I'd like to ask that in listening to me and in your working groups, you try to take into account the interplay between *discoveries* and *insights* and *tools,* or if you feel the need for labels, between experiment and theory and advances in technology (See Fig. 2.) One of

[2] A list of suggested readings appears at the end of Lecture 2.

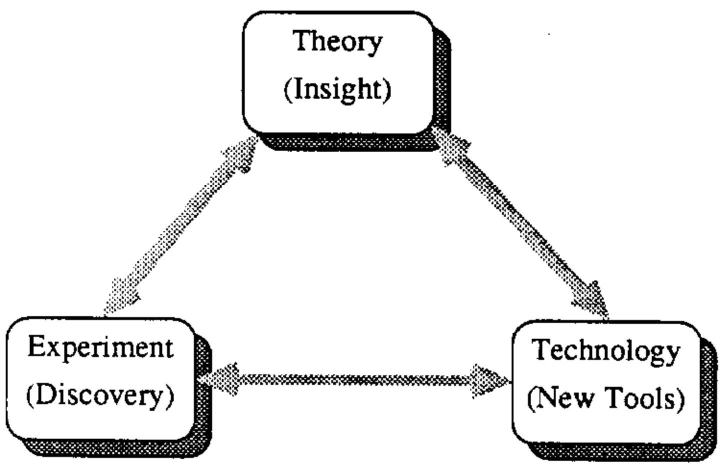


Figure 2: Synergism in basic research.

the things most disappointing to me when I look at my children's science textbooks is that there is usually simply an Aristotelean statement of "Mr. X or Ms. Y invented this or that, and this was the idea they got," and this is the way it is. The last person credited with using a technological innovation to learn something about the world is usually van Leeuwenhoek with his microscopes, and that took place three hundred years ago. The way things actually happen around here and at other great centers of science is that we do experiments, we make observations, we try to learn about things. The theory (which is not always done by theorists) leads us to catalog these observations, abstract from them, and so on. And in the long run these new insights into Nature give us the means for developing new technologies. Once a new technology is available, it is used immediately, often for the first time, in the pursuit of fundamental science to try to make new sorts of experience. You have the opportunity here to wander around and see — in addition to our buffalo herd — some of this state-of-the-art instrumentation in action. I would urge you to do that and to take that excitement of search and discovery home to your students.

One of the wonderful things Professor Weisskopf did this morning in showing those highly polished baubles of insight was to illustrate how in trying to understand things that are present in common experience we are led to retreat from common experience and make use of understanding that we gain on different levels. In particle physics we try to push always to the smallest (and we hope simplest) levels, hoping to find the most fundamental pieces of matter and the interactions among them. The whole history of science tells us that it ought to be possible to build up from those minimal parts to the larger complex systems we see around us.

In order to look at matter on fine scales and to see the interactions — to make them happen — we use particle accelerators and detectors, which together you may think of as the microscopes of high energy physics. We push to higher energies for two reasons. One is that these little things that are inside the deepest levels of matter are stuck together pretty firmly, and so to get inside and move them around and see what they do, you've got to hit them harder and harder. It is, in other words, a question of binding energy being larger as you go to deeper levels, and that requires that you hit things harder with projectiles of higher energy. The other reason we go to higher energy is related to the fact that you can listen to FM radio stations in underground parking garages, but can't listen to AM radio stations. That is, to see little things you've got to inspect them with probes of short wavelength. Short wavelength corresponds to higher energies.

Now to our main subject. The prerequisites for this lecture are the sum of human knowledge from Antiquity to twenty years ago, as represented in Fig. 3. The key idea illustrated here (and one of the enormous simplifications that physics has brought to us) is that we can explain and understand all natural phenomena in terms of a small number of fundamental forces. Since the 1930s these have been identified as the *strong force*, the *electromagnetic force* (itself the union of electricity and magnetism from a century ago), the *weak force* responsible for radioactivity, and *gravitation*. What we're going to try to do is to learn something about the properties of these forces, and to learn what are the most basic constituents upon which they act.

Thanks to a great number of experiments, principally over the last couple of decades, we have identified two classes of fundamental particles called the *leptons* and the *quarks*. I want to take a few minutes to tell you a little bit about them.

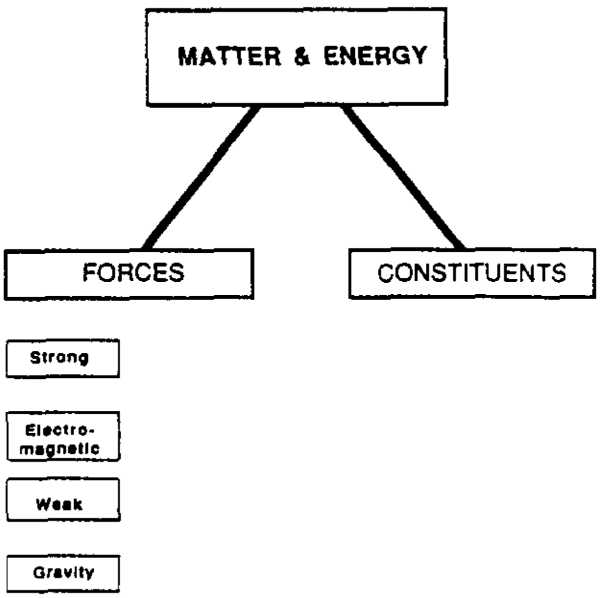


Figure 3: A starting point.

The first class is made up of particles like the electron. These are called *leptons* because the electron is a very light particle. Other members of the class turn out not to be very light, but the name persists. The leptons are particles which experience weak and electromagnetic interactions but not the strong force, not the force that binds protons and neutrons together in nuclei. We know of six such particles, shown in Fig. 4. The electron and its heavier cousins the muon and the tau lepton

LEPTONS (COLOR NEUTRAL)			
Particle name	Symbol	Mass (MeV/c^2)	Electric Charge
electron neutrino	ν_e	~ 0	0
electron	e or e^-	0.511	-1
muon neutrino	ν_μ	~ 0	0
muon	μ or μ^-	106.6	-1
tau neutrino	ν_τ	< 70	0
tau	τ or τ^-	1784	-1

Figure 4: Some characteristics of the leptons.

all carry the same electric charge; they all have spin-$\frac{1}{2}$; and, as far as we can tell, they are all pointlike. They have no extent, no gears and wheels running around inside. As far as we can tell by means of the resolution of our present "microscopes," which is down to a distance of 10^{-16} cm, these objects are just geometrical points. It is interesting to wonder whether, as we look more closely, they will develop structure inside. Are there little tiny things in there, or will the leptons remain forever truly elementary particles, structureless and indivisible? Together with the charged leptons there are three neutral particles called neutrinos, which experience weak interactions and form family patterns with the charged leptons, as we'll see in a moment.

All the known leptons can be made readily in accelerator laboratories, and they can be studied directly in the laboratory. When the charged leptons are produced at high energies, they fly out of the reaction for macroscopic distances. [The electron is absolutely stable, the muon lives for a couple of microseconds, and the tau lives for about a third of a picosecond.] We can measure their tracks by ionization and see where they have been, and so measure them readily. The neutrinos are more difficult to measure because they are neutral and don't cause ionization, but we can see the effects of their interactions when they hit other objects. A lot is known about them. The neutrinos, so far as we know, could be exactly massless, although what we have so far is upper limits on their masses. Because we can study the charged leptons in great detail, making beams of them and even storing them for long periods, we know quite a lot about their properties. The simplest of these is their mass, indicated in the chart in Fig. 4.

In observing the interactions of the leptons, we find that there are well defined families. The electron always goes in partnership with the electron's neutrino. That is to say that there are interactions which transform one into the other, but they always go back and forth. There is no interaction that we know that changes an electron into a muon or an electron into a muon's neutrino. We thus observe these rather rigid family patterns, which are suggestive that there is some deep relationship between the members.

The other class of particles we can study in the laboratory includes the proton and neutron. These are particles which experience the strong interaction (the nuclear force), in addition to the other forces. The proton and neutron are the most familiar. The pion, or π-meson, which is grossly speaking responsible for the nuclear force, is another. And then there are tables and tables ... Just yesterday I received in the mail this year's edition of the Particle Data Tables which runs to 350 pages and has everything that you want to know about all the hundreds of species of these ***hadrons***. That's quite a thick book just listing numbers and references and properties.

Now, unlike the leptons, which all were of one general kind, all spin-$\frac{1}{2}$ particles, these are particles that have integer spins, half-integer spins, small spins, large spins. All of them are composite particles. You can see that by scattering electrons from them, for example. You find that they are big and squishy inside, and typically have a size of about 10^{-13} cm. At a certain resolution, the proton resembles a Nerf basketball.

Hadrons range in stability from the proton, which has a lifetime of 10^{31} years or more, down to the Δ (Delta) and other resonances, which have lifetimes on the order of 10^{-24} to 10^{-25} seconds. The lifetime of the proton, you will notice, is many orders of magnitude longer than the age of the Universe, which is of order 10^{10} years. So obviously we have not derived the limit by watching one proton for a very long time — there *isn't* that much time — but by watching many protons for a much shorter time, on the order of a year.

The hadrons make up a great zoo of particles, in which we can recognize a certain taxonomy. A large step to bringing order and understanding to this diverse collection of beasts came in the mid-1960s with the proposal that these hadrons, these composite objects, were made up of a small number of more fundamental objects called *quarks*. Like the leptons, the quarks would be spin-$\frac{1}{2}$, pointlike particles. And we now know, as I'll try to convince you in the next few moments, that these quarks really exist, and that they are smaller than about 10^{-16} cm.

The essential distinction between the quarks and the leptons, and indeed between the quarks and most of the other constructs that we use in science, is that we don't get to see the quarks in the laboratory. We have not been able to isolate them. As a matter of fact, we now have a strong conviction that you can't isolate them. Because of that it's helpful, I think, to spend some time reminding you why we believe in quarks. Since they are not seen directly, one is entitled to ask whether this whole story about the quark model is not just so much making of myths. So what I'd like to do in the next few minutes is to try to evoke for you some of the experimental bases for our belief in quarks. The evidence will have to be circumstantial because we can't remove a quark from a hadron and hold it in our hands, but there's so much of it, it's so consistent, and it's so overwhelming that you will be led ineluctably to the belief that quarks are real!

Why do we believe in quarks? The first motivation for quarks came from observing the family patterns of the hadrons, the neutrons, protons, pions, and other things, which had been discovered up through the early sixties. As you know, in atomic spectra we observe degenerate multiplets in which energy levels with different magnetic quantum numbers, say, have exactly the same energy in the absence of magnetic fields. Only by applying perturbations (in the form of magnetic fields) do you break that degeneracy and learn about all the individual levels that are there.

That line of analysis of atomic spectra (which led to the introduction of group theory into physics), that way of thinking of degenerate multiplets, carries over to other situations and is used again and again in our attempts to understand the fundamental constituents. The first new setting is the observation that the proton and neutron seem very much alike. Both particles live in the atomic nucleus. They have almost exactly the same mass. One happens to be charged; the other isn't. The similarity led to the idea of a family partnership between them, to the idea of *isospin*.

In the same way, one could look at the particles which had been discovered in the early sixties and notice family partnerships among them. One of the great heroic enterprises of that period was to try to figure out what were the multiplets, which particles went together, and so on. Well, that's a long and fine story. The end of that long and fine story is that there's a symmetry group called $SU(3)$ (which you'll hear about in Chris Hill's lecture this afternoon), and that all of the particles known at that time could be classified as members of $SU(3)$ families.

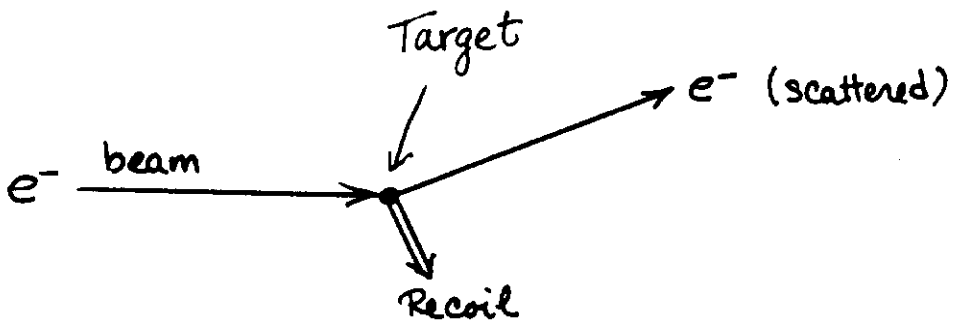
A puzzle to be explained was that whereas for angular momentum (or the rotation group) you can build up arbitrarily large multiplets, the $SU(3)$ clans seemed to be limited to families of a few small sizes. In the case of the particles like the pions, the so-called *mesons*, the families contained either one member or eight members. And in the case of particles like the proton, called *baryons*, all the families had one or eight or ten members.

So a challenge after the establishment of $SU(3)$ symmetry was to understand why only a few of these family sizes were special. The way you can do that is by saying that the hadrons, which we already know to be composite because of their finite size, are composite in a very special sense. There is a fundamental triplet of quarks (which we now call *up, down,* and *strange*) three flavors of quarks if you like, and there are simple rules for combining these three fundamental entities into the mesons and the baryons, the pion-like particles and the proton-like particles.

If you make the rule that a meson is one quark and one antiquark joined together by a force to be understood later, then the arithmetic of $SU(3)$ tells you that a family of three members times a family of anti-three members gives resulting families of either one or eight members. That's good; that's the result that you wanted to get. And if you say that particles like the proton are made of three quarks joined together, it turns out that, by the arithmetic of $SU(3)$, you can only make families of one or eight or ten members, again the desired result. Having found that this arithmetic works, you must then ask whether quarks are real, and what are the forces that allow these combinations to form and prevent more complicated combinations like six quarks or 27 quarks from joining together.

What we find is that we are led by the success of this picture to try to give it a deeper meaning, and to understand on a dynamical level why these things happen.

If you say that there are quarks inside the proton, then there ought to be some way of learning that they are there. One piece of evidence which makes that plausible is found by studying the scattering of an electron beam from a target. Here, in Fig. 5, is a standard experiment. You take an electron beam of known energy, and allow it to hit a target. The target might be a piece of carbon, a bottle of hydrogen, whatever you like. And then you observe the direction and energy of the scattered electron and, if you wish, you can observe something about the recoil particle or



$$Q^2 = (\text{Momentum transfer})^2$$

$$= (p_{e\text{-beam}} - p_{e\text{-scattered}})^2$$

Figure 5: Electron scattering kinematics

particles.

The point of this exercise is to see what happens as you vary the angle and energy of the scattered electron, and to understand what that reveals about the inner structure of the target material. Let's proceed by analogy, by looking at the historical precedent. Take as a target a carbon nucleus, really a carbon fiber, scatter electrons from it, and require that the carbon nucleus remains intact after the scattering, so we are studying the reaction

$$\text{electron} + \text{Carbon nucleus} \rightarrow \text{electron} + \text{Carbon nucleus.} \tag{1.1}$$

If you hit the carbon nucleus very hard, because it's a loosely bound collection of protons and neutrons or maybe of alpha particles, it is likely to fly apart. By requiring that it stay together, you are selecting a very rare occurrence. This is called the form factor effect. If you require that the carbon nucleus remain intact, you find that the rate at which this process occurs decreases rapidly as the amount of energy you deliver to the carbon nucleus increases. This is illustrated in Fig. 6(a).

On the other hand, if you relax the constraint that the carbon nucleus must come off intact and just say that you are going to observe the outgoing electron without regard to what came out with it, then you find that the cross section is almost independent of how hard a blow is delivered (dot-dashed line in Fig. 6(a)). The reason for this difference is that you're seeing the scattering of the electron from the individual protons inside the carbon nucleus, and at a certain resolution those protons behave as structureless particles.

So in the old days, in doing nuclear physics scattering experiments you could deduce the idea that there must be relatively structureless, electrically charged objects inside the nucleus by seeing the slow variation of this inelastic scattering rate. Of course, you could also knock the protons directly out of the carbon nucleus and verify your conclusion. If you pursue this, you can change to a situation in which your target is an individual proton, as shown in Fig. 6(b), where I've changed the scale of my abscissaby a couple orders of magnitude. Whereas I was hitting my carbon nucleus with 0.06 units of punch, I'm now hitting the proton 100 times harder.

On this scale, the proton itself doesn't like to remain intact. We see the structure of the proton reflected in the fact that the cross section or

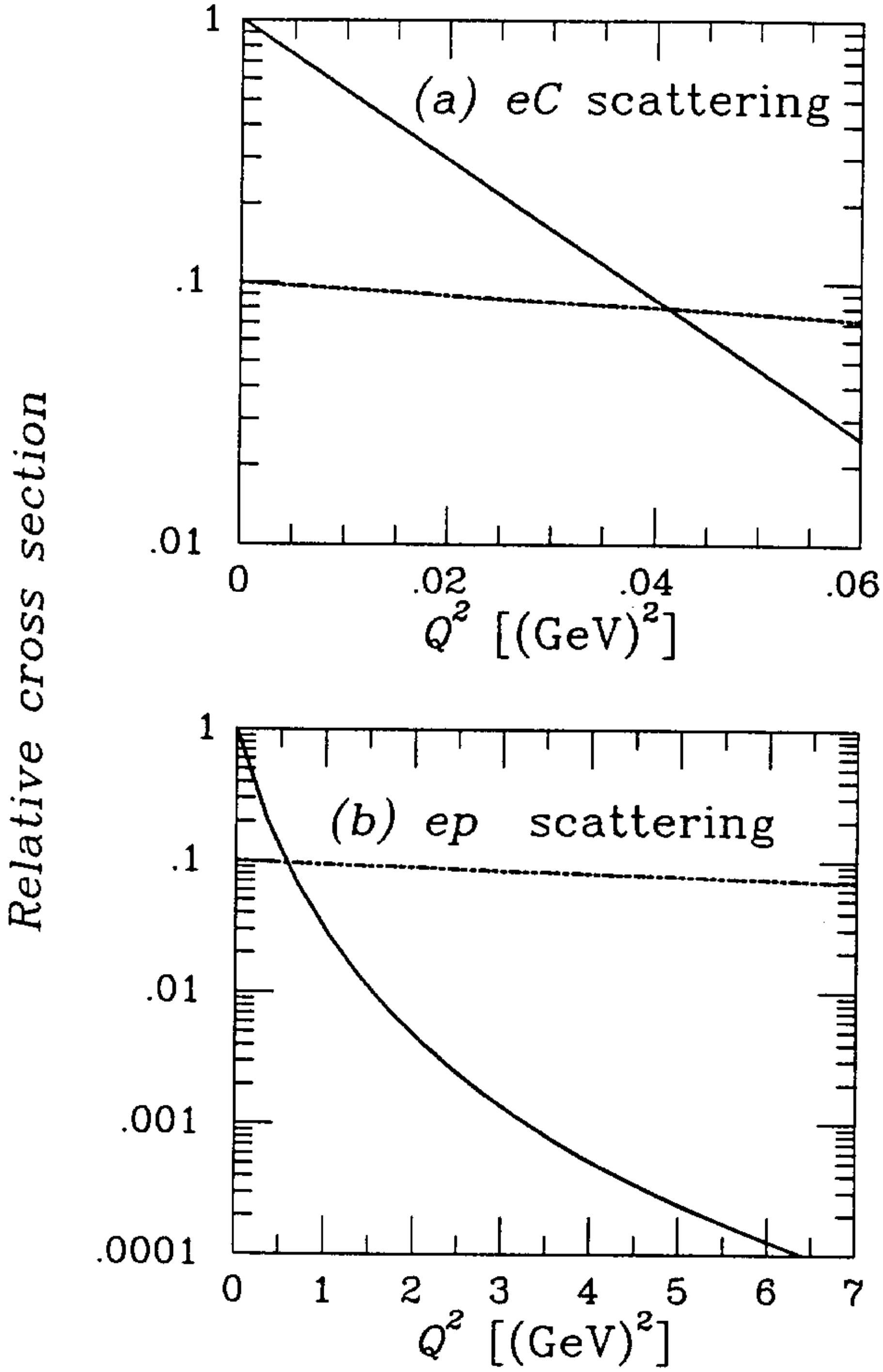


Figure 6: Elastic (solid lines) and inelastic (dot-dashed lines) cross sections for (a) eC scattering; (b) ep scattering.

rate for the reaction

$$\text{electron} + \text{proton} \rightarrow \text{electron} + \text{proton} \tag{1.2}$$

falls off rapidly. That's because the proton tends to become excited or to produce new particles when it is hit hard. On the other hand, if we relax the constraint that the proton come off intact, we find that there is once again a contribution to the cross section which is essentially independent of how hard you hit the proton.

Just as we interpreted the proton as being something hard and pointlike and electrically charged inside the carbon nucleus, it's tempting to conclude that there is something hard and pointlike and electrically charged inside the proton, and that is a role which could well be played by the quarks. Experiments which first showed this were done at Stanford Linear Accelerator Center in 1967 and 1968, and immediately led people not to accept, but to take seriously the idea that quarks really were inside the proton.

If quarks can't be knocked out of the proton, how do we know anything about the properties of quarks? Let me evoke just a few of the ways that we learn about quarks.The quark electric charges are unusual, compared to common experience: the up quark has charge $2/3$; the down and strange quarks have charge $-1/3$. These are measured in units in which the proton's charge is $+1$ and the electron's charge is -1. These assignments come in the first instance from the group theory of $SU(3)$, but you can seek more direct ways of determining them.

One of these more direct ways is to look at the decay rates for spin-one particles made out of a quark and an antiquark, the so-called *vector mesons,* particles that resemble heavy photons, and which decay into pairs of electron and positron (anti-electron). The way this happens in the quark model is that the quark and the antiquark which make up the vector meson can annihilate each other, if they find themselves at the same place, in a burst of electromagnetic energy we call a *virtual photon,* which later on will disintegrate according to the laws of quantum electrodynamics into the electron-positron pair.

Now, you can calculate this decay rate. In fact, Professor Weisskopf did it first. But we don't have to do that; we can normalize one rate to the other, as follows. The rate at which the decay occurs is determined by two basic things, as indicated in Fig. 7. One is the probability for

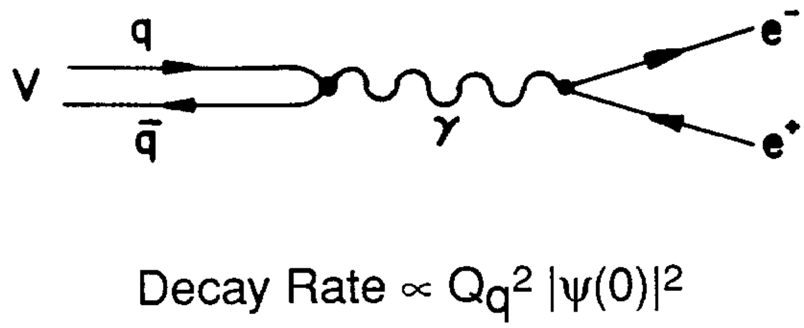


Figure 7: Decay of a vector meson into an electron-positron pair, in the quark model.

the quark and antiquark to get together and annihilate in the first place. In nonrelativistic language, this is related to the probability for them to meet at a point — so that's given by the quantum-mechanical wave function squared at the origin, *i.e.* for zero separation between the quark and antiquark. [That's this factor $|\psi(0)|^2$ in Fig. 7.] I'm going to make the gross assumption that for the vector mesons I want to talk about, that probability is the same, that they have more or less the same structure. So that's one factor which must be present, but which I'm going to pretend has no effect.

The other thing that enters is the strength of the electromagnetic coupling between the quark and antiquark, the rate at which they combine to make photons. The electromagnetic strength is just governed by the *charge* of those objects, and so the overall rate is proportional to the charge squared of these things. You can look at things called the *rho, omega,* and *phi* mesons and measure their decay rates into electron and positron pairs. You will find that the ratio of those rates is exactly in the proportion suggested by these funny charge assignments. There are numerous other ways of making that test, as well.

One of the most striking pieces of evidence that quarks are real came later with the discovery of families of particles made of two kinds of still heavier quarks called the *charm* quark and the *bottom* or *beauty* quark. And here, in Fig. 8(a) I show you the spectrum of paricles composed of the charm quark and anticharm quark. You see that there are various levels, with different values of angular momentum. They make atom-like transitions from one state to another, so that if this spectrum were unlabelled and you were asked to identify it, it would be natural to say

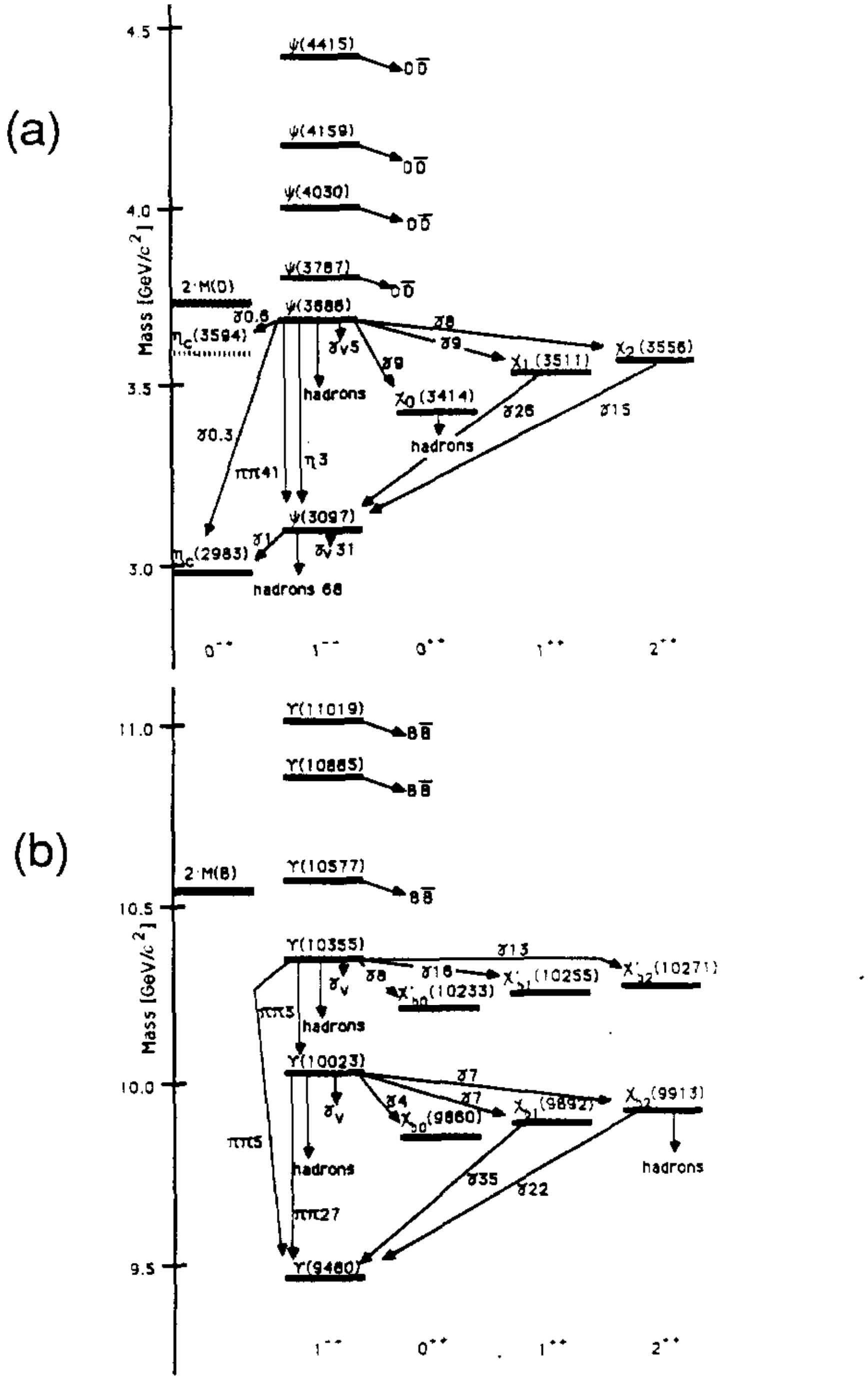


Figure 8: (a) The charmonium states (first members observed in 1974); (b) the upsilon family (first members observed in 1977).

this is an atomic spectrum of some kind. We find it both here, for the *charmonium* states, and also for the heavier *upsilon* particles made up of b-quarks, shown in Fig. 8(b). The number of states and the order of levels are exactly in agreement with the idea that fundamental spin-$\frac{1}{2}$ objects are put together — one particle plus one antiparticle — to make them up.

Still another piece of evidence for the reality of quarks — and again it's because we cannot see them directly that we have to keep making indirect arguments and asking, "Does the world behave as if there really were elementary quarks inside the hadrons?" — comes from looking at the reaction

$$\text{electron} + \text{positron} \rightarrow \text{hadrons (mostly pions)}. \qquad (1.3)$$

There are large facilities in which we make storage rings for elctrons and positrons and bring them into head-on collision.

In the quark model, we believe that the way this reaction happens is sort of to run backwards the decay reaction we just looked at: the electron and positron come together and make a virtual photon which then disintegrates into a quark and antiquark. We don't observe the quark and antiquark; by some process which is still a little mysterious to us (although we believe we understand it in principle), the quark and antiquark materialize into well-collimated sprays of pions and other hadrons.

Let's study these reactions at high energies. Here (in Fig. 9) is a projection onto a large detector about two meters in diameter at an accelerator laboratory in Hamburg, Germany. The beams were perpendicular to the plane of the page, and you see going out from the collision point one spray of pions here, one spray of pions there. It is difficult not to be led to the conclusion that one spray represents the direction of the outgoing quark and the other the direction of the outgoing antiquark. The routine events that we see at high energy do seem to display and "remember" the directions of the quark and antiquark.

Indeed, you can go further. Knowing that the quark and antiquark are spin-$\frac{1}{2}$ particles like the muon, you can say that the angular distribution of these sprays, the rate at which you see them, with respect to the beam direction, ought to be the same as the angular distribution of the reaction

$$\text{electron} + \text{positron} \rightarrow \text{muon} + \text{antimuon}. \qquad (1.4)$$

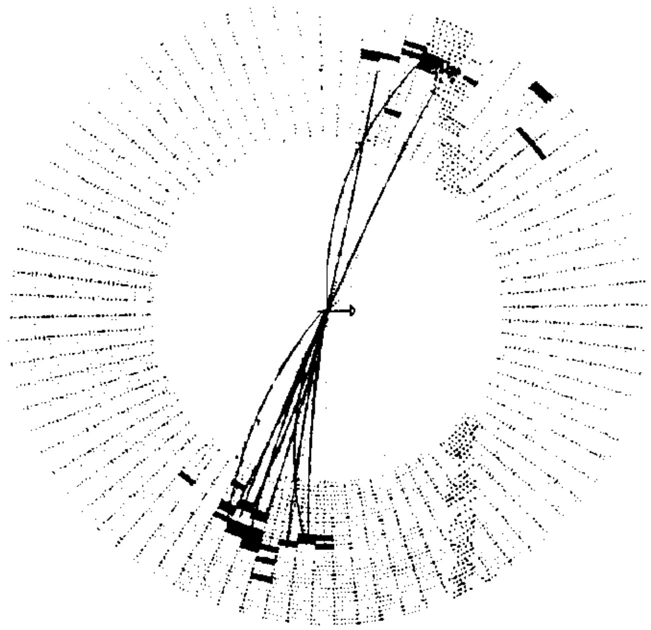


Figure 9: A two-jet event produced in 30 GeV electron-positron annihilations.

And that behavior is precisely what is observed.

Well, all this is part of the evidence for the reality of quarks, and as we say in France, it's a good story. It's a good story, but it's not completely consistent. It's not completely consistent because in building models of physical phenomena, it has paid off over the years to respect the grand principles that have great force and wide applicability. One such is the Pauli exclusion principle, which tells us how to build up the periodic table of the elements. The Pauli principle has served us well there. We can show in quantum theory that it must be true, and so it should serve us well for quarks, too.

The problem is that if you make the simplest quark model you can think of for the baryons, for particles like the Delta resonance, the Pauli principle seems not to be respected. Let me just remind you of how that goes. This first resonance, the Δ^{++}, which weighs 1232 MeV$/c^2$, has charge $+2$. In the quark model we make it out of three up quarks:

$$\Delta^{++} \sim uuu. \tag{1.5}$$

It's the lowest-mass particle of that kind, and so you expect on general grounds that each of the up-quarks is in an *s*-wave relative to any other: there's no orbital excitation between them. In order to get the total spin of the particle equal to 3/2, all three of the quarks must align their spins in the same direction. And similarly, the isospin, the *up-versus-down*-ness of the quarks, has to be aligned. That is similar to the statement that they are all up quarks. All this means that if a make an interchange of any two of the up quarks in this particle, the wave function is unchanged — symmetric. It's symmetric in space because the quarks are in relative *s*-waves, and in spin and isospin because we have completely symmetric configurations for both of those quantum numbers.

We are taught in quantum mechanics courses that bound-state fermion wave functions, wave functions of particles with half-integer spin, are supposed to be *anti*symmetric when we exchange everything in sight. So we are faced with two logical possibilities. One logical possibility is that the quark model is fundamentally flawed. We have come to a contradiction and either we have to give up the Pauli principle or abandon the quark model.

The other possibility, which seems like the easy way out, but turns out to be extremely profound, is that everything in sight isn't everything there is. There is some new degree of freedom that we haven't thought of yet, and in terms of that new degree of freedom, the three up quarks are *not* identical particles, but in fact can be distinguished. Then we can, if we like, make the wave function antisymmetric in terms of the dintinguishing characteristics. This new degree of freedom now is named *color.* We say that each quark flavor: up, down strange, and the others, comes in three distinct colors: red, green, and blue, if you like, and we require any hadron to be neutral in color. So a proton must be made of a red, a green, and a blue ("white") and a quark and antiquark must be of the same color and anticolor to form a meson.

This seems too easy, to invent something you've never seen before and couldn't see as an excuse for complying with the Pauli principle. Is there not some way to show that this additional attribute, color, is present? Let us return to the very simple reaction of electron-positron annihilation into hadrons to see if we can find evidence for the new degree of freedom. We used this reaction to argue that hadrons were emitted in jets, and that those reflected the production of quarks. Now I'm going to use that fact to make a model in which I can calculate the rate of hadron production, assuming that the things initially produced are quarks. Again, I know how to calculate these rates in all their glory, but I don't want to do that. As I told my students yesterday in the middle of a disastrous calculation on the blackboard, I only do arithmetic in public to make them feel more secure.

I've already commented [see page 16] on the similarity between muon pair production and quark pair production. I'm going to use the rate for muon pair production as the unit of cross section. At any energy, the rate at which muons are produced is the unit called *one*. That's a convenient name because this rate will be proportional to the charge squared of the muon — it's an electromagnetic interaction. The charge squared of the muon is 1, so the cross section is 1.

The quark model lets me make up quarks or down quarks or strange quarks, which then materialize as they choose into hadrons. But I can calculate the rate just by saying that the probability for the quarks to materialize into hadrons is unity. Once I've made the quarks, they will turn themselves into hadrons, and for the moment I don't have to know how that happens. The probability of making up quarks in our convenient units is the charge squared of the up quark, which is $(2/3)^2 = 4/9$. To make down quarks, it's $(-1/3)^2 = 1/9$. And to make strange quarks it's $(-1/3)^2 = 1/9$. So if I add up the three different ways I can make hadrons, I find that the cross section for making hadrons should be

$$\sigma(\text{hadrons}) = 2/3, \tag{1.6}$$

in units in which the muon pair cross section is one.

That's assuming that there is only one kind of up quark, one kind of down quark, and one kind of strange quark. If I now accept the color hypothesis and say that there are red, green, and blue up quarks or down quarks or strange quarks, then I have not three diagrams of the kind shown in Fig. 10, but in fact nine diagrams, all leading to distinct final states. And so the prediction that I make for the cross section will be, not 2/3 but three times that, or

$$\sigma(\text{hadrons})\,|_{\text{color}} = 2. \tag{1.7}$$

Now, we may go off and do an experiment (or in fact a whole series of experiments) to see which of these predictions, if any, is true. Here in

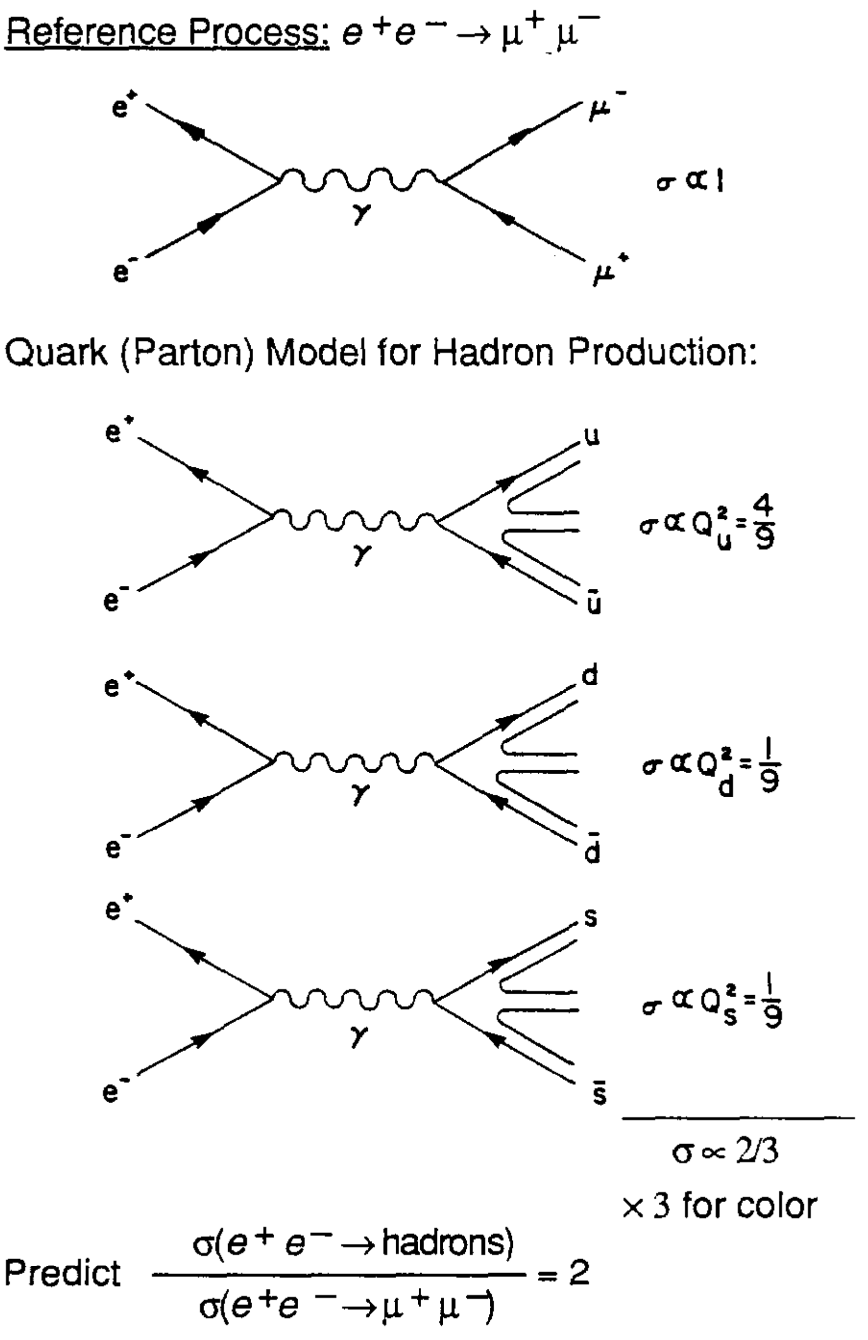


Figure 10: Calculating cross sections for electron-positron annihilations into hadrons.

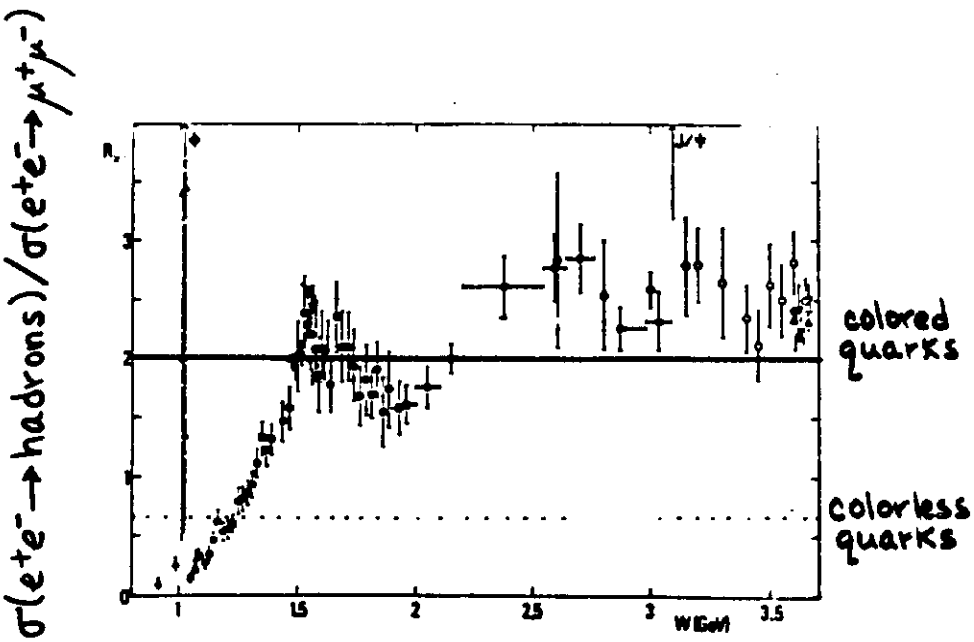


Figure 11: The ratio of hadron production to muon pair production.

Fig. 11 is the ratio of the rate of hadron production compared to the rate of muon pair production, as measured in electron-positron annihilations. At low energies there are individual resonances, the rho and omega (which are not shown on this plot), phi (which is shown here at about 1 GeV), some wiggles, and then after a while the ratio settles down to some approximately constant number. The quark model said the ratio should be a constant, so that's good. And the measured constant is within shouting distance of two, our prediction with colored quarks. It is humiliatingly far from the prediction of 2/3 in the case of colorless quarks. So this is a piece of evidence that the color degree of freedom is present.

As we move up to higher energies, we can make other flavors of quarks like charm quarks and beauty quarks. What happens there is shown in Fig. 12, where the energy scale ranges all the way up to 40 GeV. You can see that from about eleven billion electron volts up to 40 billion volts the cross section is constant and equal to a number close to 11/3. A prediction of 11/3 is precisely what you get by taking three times the

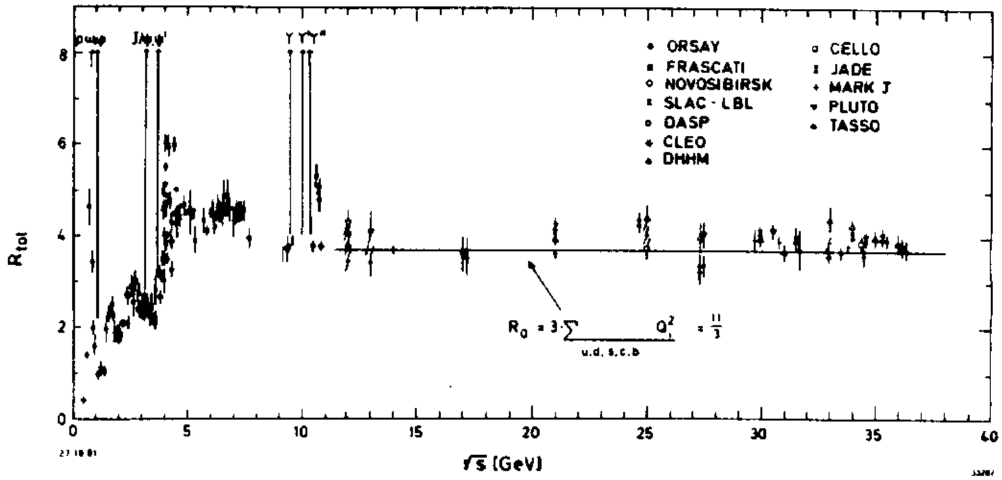


Figure 12: The ratio of hadron production to muon pair production at higher energies.

charge squared of up and down, strange and charm, and beauty. [Charm has charge $+2/3$, beauty has $-1/3$.] And so you see that there is very good agreement between the colored quark prediction and experiment, and there would be terrible disagreement, in the absence of color.

There are a number of other ways of getting at the color quantum number and convincing yourself that it is there, but they are all in this same spirit of counting up degrees of freedom in a more or less direct way. Our knowledge of the quarks is summarized in Fig. 13.

This brings us to a rough knowledge of the fundamental constituents. We have discovered particles which, at the current limit of resolution, are structureless and indivisible. For the quarks there are two and a-half families known, pending the observation of the top quark. [The indirect evidence for its existence is overwhelming.] And for the leptons, there are the three families we have discussed earlier. As we near the end of this lecture, then, our world view has advanced to the state of knowledge represented in Fig. 14. The quarks experience the strong, electromagnetic, weak, and gravitational interactions, and the leptons

QUARKS (COLOR TRIPLETS)			
Particle name	Symbol	Mass (MeV/c^2)	Electric Charge
up	u	310	2/3
down	d	310	−1/3
charm	c	1500	2/3
strange	s	505	−1/3
top/truth	t	$\gtrsim 22,500$	2/3
bottom/beauty	b	5000	−1/3

Figure 13: Some characteristics of the quarks.

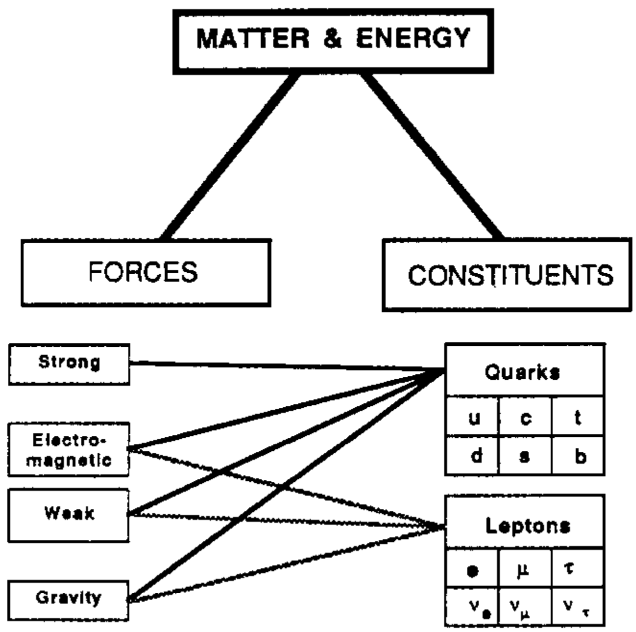


Figure 14: Progress toward the Standard Model.

the last three, but not the strong interaction.

In the second lecture I want to concentrate on the left-hand side of our diagram, but in the few minutes that remain before you rush out to drink coffee, I want to say a few words about experiment. This is offered as a stimulus to thought, and obviously not as the definitive treatment of the subject.

To a good approximation there is a single experiment done in high energy physics. It is shown in Fig. 15. A beam enters from the left and interacts with a target, from which a product emerges. If you are Lord Rutherford, the beam is alpha particles and the target is a gold foil. The product may be the same as the incident beam, or something different. The detector is often depicted in textbooks as a tin cup into which little things fall and collect.

Now, the point of doing these experiments is to try to study what is going on inside the target, and to see what are the manifestations of the interactions between the beam and the target. We have the possibility

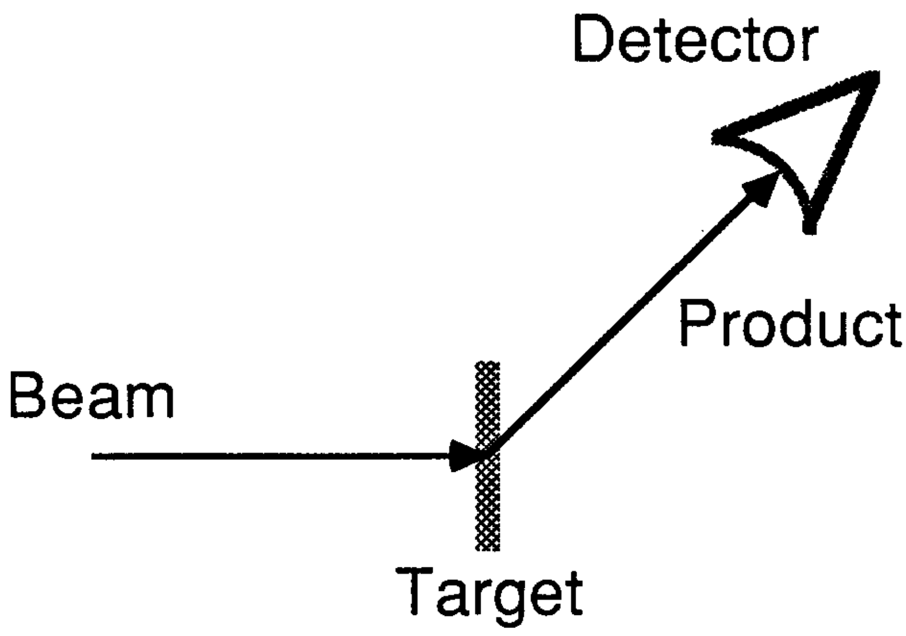


Figure 15: The Experiment.

of changing targets, of varying the properties of the beams by changing species or energies, and of observing different products. In the reactions that we study at Fermilab, it's often the case that the number of products is on the order of a hundred or so, and we want to learn as much as we can about all of those.

What is the goal of a detector system? The goal of a detector system is to measure all you can about everything that happens in the event; that is to say, to measure all the characteristics of all the particles produced in an event. This places the following requirements on a detector: you have to cover as much space as you can, as much of the angular range as possible, in all three dimensions, so you don't miss anything. But for reasons that we'll discuss immediately, you want to have high spatial resolution. If you had just one large tin cup that registered everything coming out of a collision and didn't distinguish where it was coming out, that would be less interesting — because it gives less information — than having a lot of little tin cups and counting who went here, who went there, and so on.

You would also like to be able to identify particle characteristics. And finally, one of the great challenges, particularly now, is to try to select events of interest, the special things that you want to study, from the routine background. There is a saying in particle physics that yesterday's sensation is today's calibration and tomorrow's background.

You want to do all of this keeping the cost of construction, operation, and data reconstruction within reasonable bounds. What are reasonable bounds? There is a detector you can see in a big orange building down the road, which does all these things. The price of that detector (the Collider Detector, CDF) is about $50 million. With respect to data reduction, the coin of the realm here is a VAX-11/780 computer, and the data analysis for that experiment is estimated to require 50 such computers running full time.

Let me now say a few words about the principles that underlie detection, the ways that we can think of learning things about these produced particles.

Charged particles lose energy by ionization as they pass through matter. Of course, there is a whole science and technology built up of how they ionize, which has to do with electrodynamics, the properties of materials, *etc.*, which is itself very interesting and good physics. What you want to do is to measure the position and magnitude of the ionization trails to learn something about where the particles went and what they were.

In some cases, if you measure how long it took to go from here to there, as you do in elementary physics labs, you can measure the velocity of particles and therefore infer something about their identity.

Magnetic fields deflect charged particles into curved orbits. By measuring the curvature of the orbit you can, knowing the properties of the magnetic field, determine the particle's momentum.

Beyond that, different kinds of radiation can be emitted by particles under different conditions. One of the most useful so far in particle physics is the Cherenkov radiation emitted by particles which pass through a medium faster than the speed of light in that medium. A shock front builds up radiation with characteristic opening angle and intensity patterns, and by measuring the intensity of the radiation and the angle with respect to the particle direction, you can make inferences about the energy of the particle, its mass, and other characteristics. Coherent radiation is also emitted by particles crossing the interface between materials (transition radiation), and by particles passing through magnetic fields (synchrotron radiation).

Neutrinos are wonderful particles to detect. They interact so feebly that they are almost not there,[3] and so you infer their presence by the fact that you didn't see something. Pauli's original reason for inventing the neutrino was that there seemed to be missing energy in radioactive beta-decay. In the same way, we can try to sum up all the momentum carried by particles produced in a high-energy collision, and if there is a big lump missing off in that direction, then you say, "Ah, a neutrino or something like a neutrino went off in that direction," because you believe in momentum conservation. So nonobservation can be a good way of observing, provided you can be sure that you would have observed something else, had it been there.

Among the particles that do something interesting when they pass through matter, electrons and photons are special because they produce characteristic electromagnetic showers, converting all the original energy

[3]See "Cosmic Gall," in John Updike, *Telephone Poles* (Alfred A. Knopf, New York, 1969), p. 5.

of the particle to ionization and then the relaxation of excited atoms. By recording the deposited energy, you can do electromagnetic calorimetry. You can observe the development of the shower and, by adding up all the energy, learn what the energy of the electron of photon was.

Hadrons passing through matter lose some energy to ionization, but also have strong interactions with the nuclei and so they will start nuclear cascades, nuclear showers. If you make a large enough block of instrumented steel, you can again collect all the energy from the incident hadron.

Finally, muons are an exception to these sorts of patterns because they can go through huge thicknesses of material without losing much energy. They radiate much less than electrons do because they are so much heavier, and they do not induce nuclear showers. As a result you can identify muons by making a big block of material and watching what charged particles come out the other side. In the case of the Fermilab neutrino beam, we have about a kilometer or so of steel and earth in the way just to absorb all the muons which otherwise would contaminate the beam.

Using all these principles we can arrive at the idea of *layered detectors*. What you try to do is to exploit different characteristics of the various physical principles of detection to do different things. Close in you need a detector which has very good spatial resolution and can sustain high rates because lots of particles are emerging from a small volume. There is a special class of detectors called *vertex detectors* used close in. Next there are charged particle tracking chambers which trace the progress (often through a magnetic field) of particles coming out from the collision point.

Combined with this, or sometimes in addition to this, there is often an attempt made to identify particle types by using some of the coherent radiation schemes. After all that nondestructive tracking has been done with only a little material in the way of the outgoing particles, you then put lots of material of various sorts in the way to do the calorimetry, contrived so that anything that penetrates the entire detector must be a muon, which you may wish to measure again.

Here (Fig. 16) is a picture of the Collider Detector, which I hope you will take the time to see while you are here at Fermilab. Note from the sketch that a typical person is one-fourth to one-fifth the size of the detector. An exploded view of half of the detector is shown in Fig. 17. There is

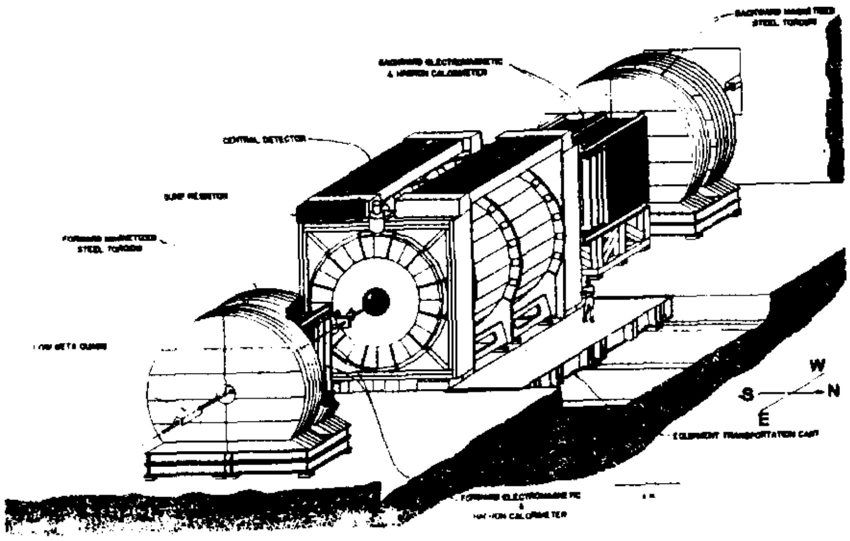


Figure 16: CDF, the Collider Detector at Fermilab.

a highly sophisticated vertex tracking device around the interaction point. Then you find, all immersed in a superconducting solenoid, the central tracking, an electromagnetic shower calorimeter, hadron calorimetry, the magnet yoke (which is iron), so that the particles which penetrate to the outside should be muons. That's all in the central region. The same sorts of pieces are found as you go toward the forward direction. At each location and for each task, you try to choose the best detector in terms of performance, reliability, cost, and so on.

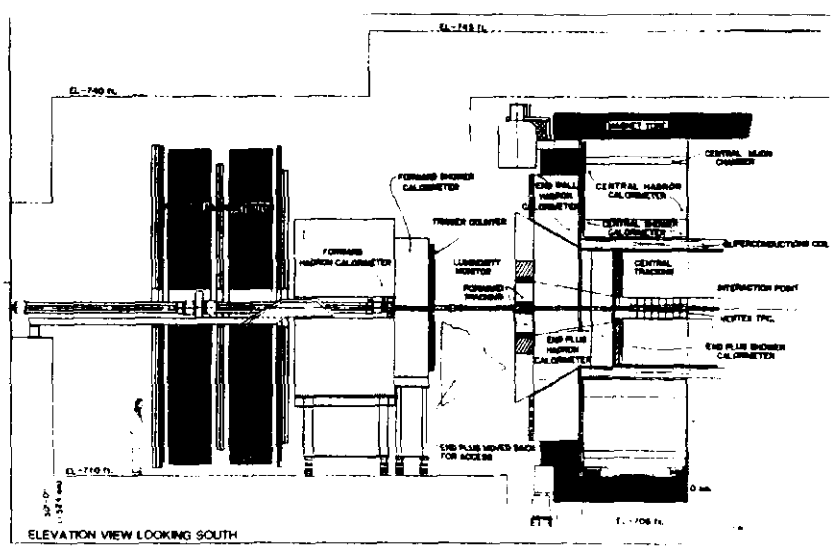


Figure 17: Exploded view of the Fermilab Collider Detector.

I do hope that while you are here you will spend some time looking at these detection devices and trying to understand a little bit about them. After the break, we will move on to the strategy of gauge theories.

LECTURE 2: THE IDEA OF GAUGE THEORIES

In this second talk, I want to focus on the interactions and to explain a bit of the motivation for gauge theories, and the basic elements of the gauge theory strategy. What we shall see is that symmetries in Nature, when we recognize and use them properly, can be used not only as restrictions that guide the formulation of theories, but also as tools that help us construct the theories directly.

Let us now recall the theories currently in use to describe the strong, weak, and electromagnetic interactions. The first of these, and in many ways the prototype for quantum theories, is *Quantum Electrodynamics,* or *QED.* This is the most successful of physical theories: it works, essentially without modification, from distances on the subatomic scale down to nearly 10^{-16} cm out to enormous distances on the interplanetary scale. When you consider that the theory is built upon experiments first done by Cavendish and others on the scale of half a meter or so, the success of the extrapolation is really quite striking.

Quantum electrodynamics is in part the model for, and is incorporated in, the theory of weak and electromagnetic interactions brought to its final form by Weinberg and Salam 20 years ago. The resulting theory describes at the same time the weak and electromagnetic interactions. Although for the moment it is not nearly as well tested as QED itself, the electroweak theory has many very precise experimental successes. It anticipated a new kind of radioactivity called neutral weak currents, required the existence of the charmed quark, predicted the recently discovered carriers of the weak interactions, W^+, W^-, and Z^0, and (to the level at which we have been able to do experiments) gives a precise and quantitative description of everything we see in the electroweak realm.

A theory that we'll discuss at somewhat greater length is *Quantum Chromodynamics*, a theory of the strong interactions. It is called "chromo" because it is based on the idea that the *color* property of quarks which distinguishes them from leptons and enabled the quark model to survive the Pauli principle functions in some sense as a strong charge. And so the theory is called *QCD* in imitation of QED.

QCD is based on the color symmetry of the quarks in a way we'll review a bit later. For a variety of reasons, not least of which is that the strong interactions are strong and theoretical physicists are only good at calculating the consequences of feeble interactions, QCD has not yet been tested as precisely as the other interactions. It does give us lots of insight into the systematics of high energy collisions and the spectrum of hadrons. It predicts force-carrying particles called *gluons*, and in some restricted realms there are some quantitative successes which are rather impressive.

I'm now going to explain where gauge theories come from, and the strategy involved in deriving them. So far as we can tell, gauge theories provide the basis for correct, useful descriptions of all the fundamental interactions. They have a number of properties which we'll talk about later on. The reason for talking in general terms about how we construct gauge theories is that it's very easy to make up theories, and it's particularly easy to make up wrong theories. If you can find some guiding principles, they may restrict your search for different classes of theories. Now, you have to be careful not to restrict yourself too much, but if you pick a guiding principle like energy conservation or Lorentz invariance or some such, which is supported in great detail by lots of experimental data, and say provisionally that you will only look at theories which satisfy that principle, then you've saved yourself the trouble of looking at a lot of theories which have no chance of being correct. In the same spirit, if you can find and attach yourself to a principle which will lead you only to make theories from the class of those that might possibly be right, that's a good thing, at least in terms of economy of effort.

The strategy of gauge theories goes roughly like this. We recognize a symmetry in Nature. This afternoon you will be reminded that for many sorts of symmetries (continuous symmetries like rotation invariance, translation invariance, and so on), there is a deep connection with conservation laws. Rotation invariance is intimately related with the conservation of angular momentum, for example. By recognizing conservation laws, by seeing symmetries in Nature, we are led to build equations of physics that respect the symmetries in question. Having done that, we then try to impose the symmetry in a stricter form. I'll show you immediately by means of an example what I'm trying to say here, but for purposes of giving an outline let me proceed without explaining. When the new requirement is imposed, it will happen that the equations of physics from which we began must be modified in order to accommodate the stricter form of the symmetry. This can be done in a mathematically consistent way only by introducing new sorts of interactions, and new particles to carry those interactions.

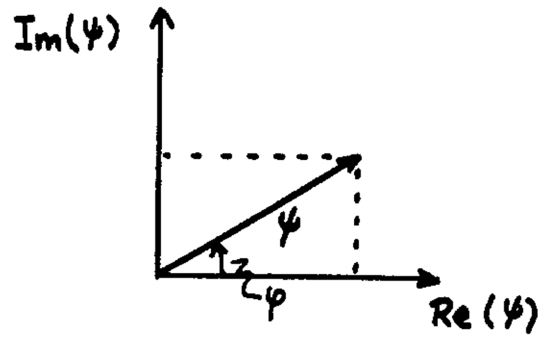
There is an opportunity for blunder here. If I pick a symmetry that I think I see in Nature and I go through this program, I may well arrive at a theory which is mathematically self-consistent but which, because I was inept in my choice of the symmetry, doesn't describe the world we live in. The literature is littered with the corpses of such theories, and I will spare you examples of them.

Now, you may ask, "What is he trying to say? What does all that mean?"

To give an example, I have to beg your indulgence. The indulgence is to suppose that we know quantum mechanics but not electromagnetism. Now, from the times I've taught graduate courses in electricity and magnetism, I know that half of that statement (at least) is likely to be true. And from the times I've taught quantum mechanics, I have my doubts about the other half of the supposition.

I'm going to begin with quantum mechanics and lead us to electromagnetism.[4] What I need to know about quantum mechanics is that the

[4]This may seem to be a fake, because it's not the way electromagnetism was invented. But it should have been!



$$\psi = Re(\psi) + i\, Im(\psi) \qquad \psi^* = Re(\psi) - i\, Im(\psi)$$

$$i^2 = -1$$

OR $$\psi = |\psi| e^{i\varphi} \qquad |\psi| = \sqrt{\psi^* \psi}$$

$$e^{i\varphi} = \cos\varphi + i \sin\varphi$$

$$\tan\varphi = Im(\psi)/Re(\psi)$$

Figure 18: Argand diagram representation of the quantum mechanical wave function $\psi(x)$.

quantum mechanical state of a system is described by some complex wave function called $\psi(x)$. This is a complex function with a real part and an imaginary part and, if I like, I can describe it as a vector in an Argand plot as in Fig. 18. As you know, you can characterize that in various ways, as shown in the sketch. Corresponding to the wave function ψ there is the complex conjugate ψ^* of the wave function, its reflection about the real axis. Of the various representations for the wave function given in Fig. 18, the one most convenient for our purposes will be to write the wave function as

$$\psi(x) = |\psi(x)| \exp i\phi. \tag{2.1}$$

Now, everyone knows that in quantum mechanics observable quantities, things you can measure in the laboratory, are expressed as *expectation values* or *scalar products* which are integrals over some appropriate region of space, of a volume element times the complex conjugate of the wave function times a Hermitian operator O times the wave function,

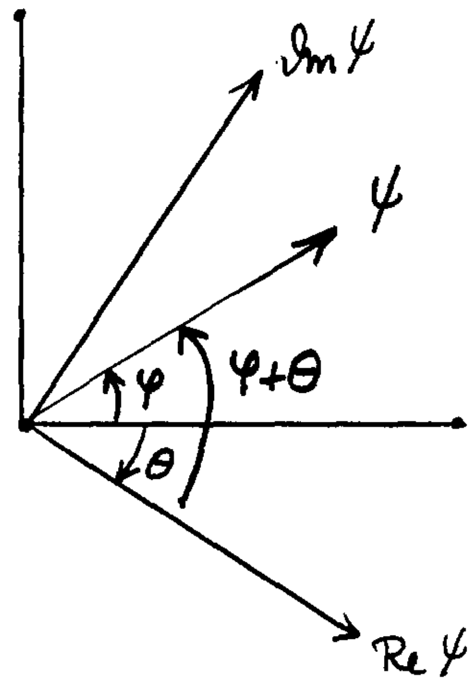


Figure 19: New definition of the real and imaginary axes.

symbolically

$$\langle \mathcal{O} \rangle = \int_{(V)} dV \psi^*(x) \mathcal{O} \psi(x). \tag{2.2}$$

We can verify by heavy-handed means that the quantity we're going to measure is unchanged if we redefine the phase. I come along and say that the real and imaginary axes in Fig. 18 are very fine for you, but I don't like them. I'm going to change to the new coordinate system shown in Fig. 19. In terms of those coordinates, I can of course measure the real part and the imaginary part of ψ, or express ψ in terms of a new set of polar coordinates. Rotating the definition of the real axis down by an angle theta is equivalent to multiplying ψ by $e^{i\theta}$:

$$\begin{array}{lcl} \psi & \rightarrow & e^{i\theta}\psi \\ \psi^* & \rightarrow & e^{-i\theta}\psi \end{array} ; \tag{2.3}$$

ψ^* gets rotated in the opposite sense.

In terms of the new ψ and ψ^*, our observable becomes

$$\begin{aligned} \langle O \rangle &= \int_{(V)} dV \psi^*(x) e^{-i\theta} O e^{i\theta} \psi(x) \\ &= \int_{(V)} dV \psi^*(x) O \psi(x). \end{aligned} \tag{2.4}$$

The factors $e^{-i\theta}$ and $e^{i\theta}$ eat each other up, giving back *one*, so the quantity we are calculating is unchanged by the operation of changing coordinates. It is the same before and after I've made the change of phase indicated in Fig. 19. That is to say that the absolute phase of the quantum mechanical wave function is *arbitrary*. It is not something to which measurements can be sensitive.

Now, in fact this sort of phase symmetry has a deep connection, if you formulate it properly in detail, with the conservation of electric charge. From phase symmetry of precisely this kind you can derive the fact that the electric charge must be conserved.

For the moment, I'm not going to focus on that, but only to admire the fact that I could make this change of convention. Just to put it in symmetry language, I can say that ordinary quantum mechanics is invariant under *global phase rotations*, phase rotations in which the convention is changed by the same amount at every seat in the Fermilab auditorium. Here, in Fig. 20(a), is where we started out. Each of you agreed with me that this would be our direction for the positive real axis, the original convention for zero phase ($\phi = 0$). Later on, we all agreed together that the direction shown in Fig. 20(b) would define the direction of zero phase. We found that physics didn't change when we made that rotation.

Now, some of you might object and say, "Why should *you* be able to tell *me* what *my* phase is? Couldn't we be more democratic and choose a different phase convention independently at every point in space?" Not in a haphazard fashion: you might want to have some common harmony with your neighbors, but might it be possible to have a position-dependent definition of the zero of phase? Would that be all right? Do the laws of quantum mechanics admit that sort of symmetry, a more general symmetry than the phase invariance we have just investigated?

This freedom to choose a position-dependent phase convention would mean that instead of multiplying my wave function by a fixed rotation, $e^{i\theta}$, I would multiply it by a position dependent phase, $e^{i\alpha(x)}$. Does that

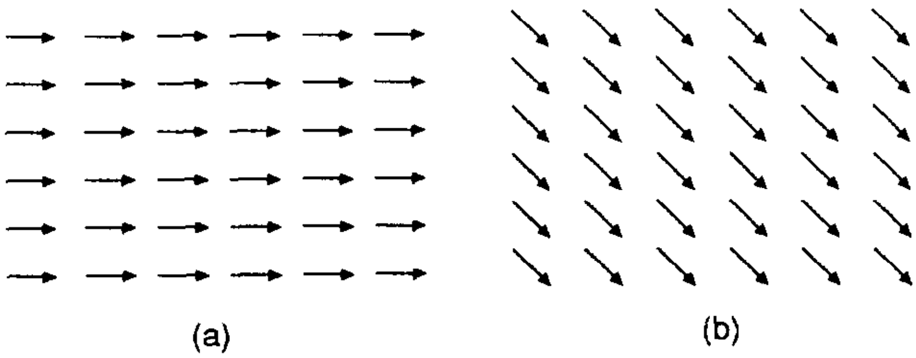


Figure 20: (a) Original convention for zero phase of the quantum mechanical wave function; (b) new convention for zero phase. Each arrow represents a seat in the Fermilab auditorium.

work? Well, let's suppose we are talking about the quantum mechanics of a free particle. We can check whether the new symmetry is respected either by talking about observable quantities, as we have just done, or by plugging the transformation law into the Schrödinger equation and asking whether the same equation holds before and after the phase change. Let me do the exercise in terms of observables; what we'll find is that the quantum mechanics of a free particle is not invariant under local phase rotations.

How can we see that? There are observables in the world like momentum, and there are pieces of the Schrödinger equation itself, which involve derivatives or gradients. What we can do is to calculate how a gradient changes when you make a position-dependent phase change on the wave function. Here is what happens: as

$$\psi(x) \rightarrow \exp i\alpha(x) \cdot \psi(x), \tag{2.5}$$

the gradient becomes

$$\nabla\psi(x) \rightarrow \exp i\alpha(x) \cdot [\nabla\psi(x) + i\psi(x)\nabla\alpha(x)]. \tag{2.6}$$

Unlike the wave function, the gradient is not simply multiplied by a phase. The fact that $\alpha(x)$ has a position dependence means that its gradient is nonvanishing, and that gives rise to the second term on the right-hand side of Eqn. (2.6).

That extra term means that if I try to calculate an expectation value like

$$\psi^* \nabla \psi \rightarrow \psi^* e^{-i\alpha(x)} e^{i\alpha(x)} \left[\nabla \psi(x) + i\psi(x) \nabla \alpha(x)\right], \tag{2.7}$$

I do not recover the original value. So the answer to the question, "Does ordinary quantum mechanics admit a local variation of phase of the wave function?" is that it does not.

Now, especially after Chris Hill's lectures this afternoon in which symmetry will be made a part of your being, you might ask yourself, "Couldn't I change the equations of physics a little bit because that symmetry seems so nice and appealing?" In other words, is it possible, if only as a little mathematical homework problem, to make some changes in the equations so that everything will work out and the modified equations admit this more general phase symmetry?

The answer to that question is yes, but only if you introduce an *interaction*, a specific kind of interaction called a *gauge field*. What we need to introduce will be the electromagnetic field, or something like it. Let me show you the arithmetic rather schematically. I'll then remind you that the answer we get to is something you already know, and you will be prepared to take the gauge theory leap of faith.

The solution is that I'm going to introduce (in three vector notation) an electromagnetic vector potential $\mathbf{A}(x)$, and I'm going to make the following rule: when I rotate the phase of the wave function by an amount[5]

$$\psi(x) \rightarrow \exp iq\phi(x) \cdot \psi(x), \tag{2.8}$$

I'm going to change my newly introduced electromagnetic vector potential by shifting it my an amount

$$\mathbf{A} \rightarrow \mathbf{A} - \nabla\phi. \tag{2.9}$$

How do I know to do that? I know to do that because I went through the equations and asked, "What do I have to do so they come out right after phase rotation?" And this is the answer.

Now, in addition to all that, I make an agreement with myself that everywhere in the laws of physics — in the definition of observables, in

[5] In terms of our earlier notation, I have renamed α to be $q\phi$, where q is supposed to be the electric charge. That's to suggest that the theory we will derive is electromagnetism. If I like, I can choose another charge and derive another theory.

the Schrödinger equation, and so forth — everywhere I see a gradient I will replace it with something named *the gauge covariant derivative*,

$$\mathcal{D} \equiv \nabla + iq\mathbf{A}. \tag{2.10}$$

Those of you who are good at juggling factors of i and $\hbar$ will notice a resemblance between this expression, to which I've been led by fiddling the equations of physics, and the familiar replacement of classical electrodynamics in which the momentum $\mathbf{p}$ becomes $\mathbf{p} - q\mathbf{A}$. This is a source of comfort and reassurance.

What remains is to verify that this new object, the generalized gradient, when acting on the wave function goes into simply a phase factor times itself under the combined transformations (2.8) and (2.9). As a little homework exercise, you can check that $\psi^* \mathcal{D} \psi$ is invariant under local phase transformations.

And so, I've invented a theory. I've had to change the gradient, redefine the momentum operator, *etc.* But I've invented a theory in which the appearance of the equations is identical before and after local phase rotations, and all the observables we can imagine will be the same before and after the local phase rotations.

The fact that you've seen the final results before invites you to believe — and it's even true — that the theory we've derived in this way is exactly the theory of electromagnetism. If we do the same steps in a covariant, relativistic way, the theory we derive is precisely quantum electrodynamics.

So that's the arithmetic of it, and that's the general strategy of it. We can carry out the same kind of analysis for other theories or for more complicated theories. The arithmetic becomes more involved, but the strategy is always the same. The encouragement for trying the strategy in other settings comes from noticing that we can recover the idea of QED by starting with a symmetry and proceeding along these simple lines.

The phase symmetry is just the gauge invariance of quantum electrodynamics; the shift we made in the vector potential is the freedom textbooks normally call gauge invariance. Something to be emphasized to students is that gauge invariance means more than just the ability to choose arbitrarily the zero of a potential. It has, as we have just seen, a deep connection with symmetry through quantum mechanics.

Now that we've looked at one example, we can ask what are the general consequences of this strategy. *Global symmetry,* in which we make a continuous transformation (like a phase rotation) everywhere by the same amount, leads to a conserved current, a conserved charge. In the example we have considered, this is the electric charge. The *local symmetry* implies in addition that there must be an interaction. It had to be mediated by a spin-one vector field. It turns out that it had to be a massless field. And furthermore, at least if you follow your nose, the interaction between that new force and matter turns out to be a form traditionally known as "minimal coupling."

In this light we can think of electrodynamics as the gauge theory [the theory built upon this phase invariance or gauge invariance[6]] built upon the group of phase transformations, the group of rotations in a plane called the unitary group $U(1)$. Can we do the same for other continuous groups? Do they have to be commuting ("Abelian") groups, or not? The answer is that you can always construct a theory, for any continuous gauge group. Some of them will have more complicated properties, but you can always make the construction.

That completes the first topic for this lecture, how to construct a gauge theory. I told you a moment ago that electrodynamics, grossly speaking, in the form of Maxwell's equations, is valid not only down

[6] Why do we use the term gauge invariance? The original argument in the style we have just explored was given in papers written around 1921 by Hermann Weyl. At that time, the known forces were electromagnetism and gravity, and so it was a natural impulse to try to give a unified basis for the two. Gravity had something to do with geometry, and so it was natural, I suppose, to try to think of a geometrical basis for electromagnetism. Weyl's contribution, for which the general strategy is exactly the one I used today, was to say, suppose that there is a scale invariance of the world, so that the laws of physics have to be the same as the scale, or measure of length, changes from point to point. Requiring the equations of physics to have this invariance, he found the necessity of inventing an interaction and hoped to identify this interaction with electromagnetism. It turns out, as we've just seen, that you need a phase change rather than a scale change to recover electromagnetism, but the idea — the strategy — has persisted. Since quantum mechanics wasn't invented until a few years after he'd made this proposal, we can hardly blame Weyl for not understanding the importance of phases at the time.

In Weyl's original papers he used a German term, *Eich*, meaning calibration or gauge. Following correspondence with Fock, London, and others, after the invention of quantum mechanics, Weyl changed his program to one of phase invariance, but retained the old term. Gauge invariance had caught on, and it is the term we still use today.

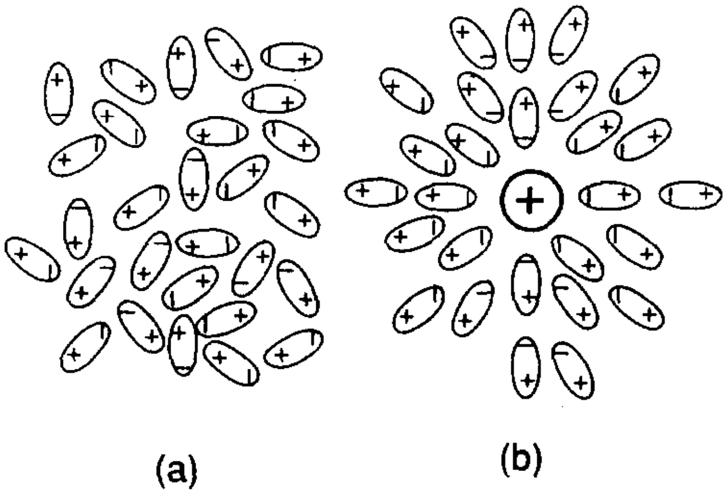


Figure 21: Bipolar molecules in a dielectric medium: (a) disordered state; (b) ordered (polarized) state in the presence of a test charge.

to very short distances (around 10^{-16} cm), but also applies out to very large distances. The best measurement that I know about comes from measuring the rate of falloff of the magnetic fields of the large planets, Jupiter and Saturn. It indicates that Maxwell's equations hold out to a distance of about $4 \cdot 10^{10}$ cm. There is indirect evidence that pushes the range of validity out another twelve orders of magnitude.

What is interesting for electrodynamics, and later for the theory of the strong interactions, is that in spite of the theory's enormous range of validity, there are well-understood modifications to the simple behavior in a polarizable medium. This is the phenomenon of charge screening. Let me make a model for such a medium. Here, in Fig. 21(a), is a polarizable medium in which molecules behave as little boats with a positively charged end and a negatively charged end. In the absence of some external source of charge, the boats are distributed without macroscopic order. Of course, the oppositely charged ends attract each other in such a way that, grossly speaking, in any lump of this medium you will find a net charge zero: the charges cancel out, on balance.

What happens if I stick some positive charge right in the middle of the medium? As long as the little boats are free to move about, they will orient themselves so that the negatively charged ends of the boats are attracted to the positive test charge. Schematically, the pattern will be as shown in Fig. 21(b).

The effect of this is that if I imagine probing the test charge, measuring the charge in the medium by inserting a hypothetically nonperturbing probe, the charge my probe feels will be less than the charge carried by the test charge. This is because the total charge enclosed in a circle centered on the test charge, with radius given by the distance from the test charge to the probe, is less than the test charge itself.

In order to see the full strength of the test charge in a molecular substance, I must approach the test charge closely — so closely that my probe is within the molecular scale. Once my probe is there it sees the whole charge, unscreened by the molecules in the medium. That's a gross way of indicating that the effective charge, the charge I measure, increases at short distances.

Because of quantum mechanical effects and the possibility that the vacuum can fluctuate into pairs of electrons and positrons for very short times, the same thing occurs in the vacuum. The vacuum that we live in is not an empty thing, but something in which pairs of electrons and positrons are coming and going all the time. While they are here, they can be polarized by a local test charge. The effect of that *vacuum polarization* is precisely the same as in a dielectric medium, to screen a charge and to make the effective charge larger at short distances than at large distances.

I now want to move on, building on the idea of gauge theories, to the force between quarks as another example of how we take a symmetry and build a theory from it. We noticed that every flavor of quark (*e.g.* up, down, strange, and charm) came in three distinct varieties called colors. And in order to make the mesons, we had to have antiquarks which came in anticolors. We named these colors red, green, and blue, but we could have chosen A, B, and C.

Physics isn't supposed to change if we change the names. The color symmetry means that when I interchange the names red, green, and blue, or reassign them in some continuous fashion, nothing should change. So I should build my laws of physics to have that property. That is to say that the interactions of a red quark and a green quark and a blue quark should all be the same. That suggests that I might be able to build a theory in which that freedom to name red, green, and blue is respected locally: a gauge theory of the color force, QCD.

Now, in this case, because we have three kinds of charge instead of only one, as in the case of electrodynamics, the arithmetic is more complicated,

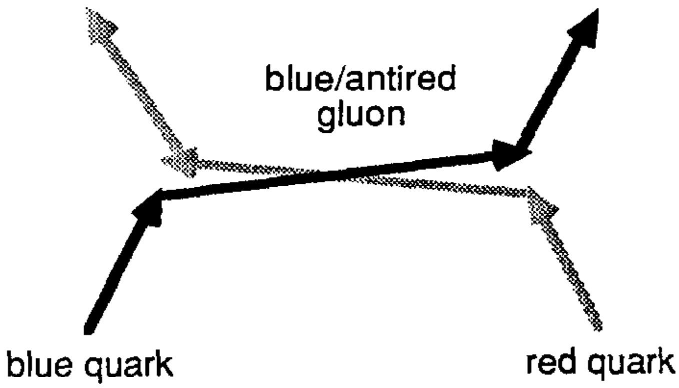


Figure 22: Quark-quark scattering in QCD.

and so I won't work it out in public. But you can make such a theory, a theory in which there are interactions between the quarks mediated by massless spin-one particles. We call these force particles gluons because they glue the quarks together. The strategy we have followed is to think of color, the attribute that differentiates the quarks from the leptons, as the charge of the strong interactions, and to build a theory based on local color symmetry.

What do these interactions look like? Fig. 22 shows the scattering of a blue quark and a red quark. They interact by exchanging a blue–antired gluon, and emerge as a red quark and a blue quark. Notice that in this example the gluons themselves carry color charge, in fact one color charge and one anticolor charge. Since the gluons are colored, they will have strong interactions mediated by gluons. You can construct these interactions just by drawing colored pictures. Here, in Fig. 23, is a green–antiblue gluon scattering from a green–antired by exhanging a blue–antired gluon.

Now, you've never seen a quark but you all believe that they exist. How can I convince you that the gluons exist? By doing a variation on one of the ways that I conviced you that quarks exist. Back on page 16 we talked about electron-positron annihilation into hadrons proceeding through the the formation of a quark–antiquark pair. The quarks materialized into hadrons, mostly pions, which remembered the direction of their parent quarks. That gave us the two-jet events of the kind illustrated in Fig. 9.

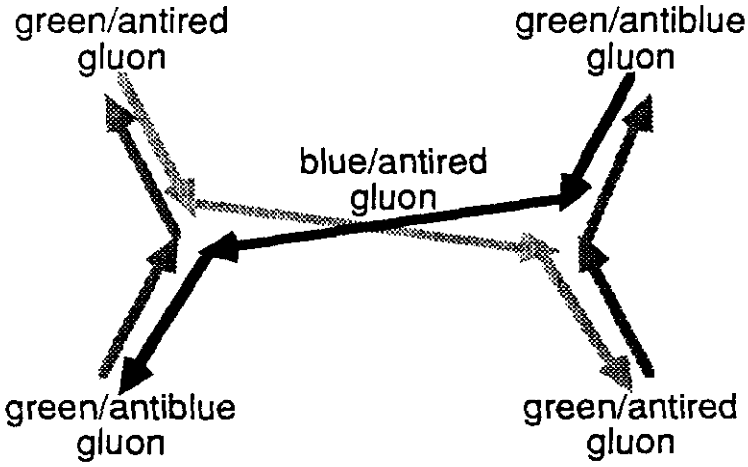


Figure 23: Gluon-gluon scattering in QCD.

In much the same way, now that we have invented gluons interacting with the quarks we may imagine that sometimes one of the outgoing quarks radiates a gluon,

$$\begin{aligned}\text{electron} + \text{positron} \;\rightarrow\; & \text{quark} + \text{antiquark} \\ & \hookrightarrow \text{quark} + \text{gluon},\end{aligned} \tag{2.11}$$

just as an outgoing muon may radiate a photon in

$$\begin{aligned}\text{electron} + \text{positron} \;\rightarrow\; & \text{muon} + \text{antimuon} \\ & \hookrightarrow \text{muon} + \text{photon}.\end{aligned} \tag{2.12}$$

When that happens, I expect my two-jet event to change into a three-jet event. One quark jet splits into a quark jet and a gluon jet.

At high energies in electron-positron annihilations, three-jet events are quite common. Fig. 24 shows a picture of one in the same detector in which we saw the two-jet event in Lecture 1. You can see one fully developed jet, and two smaller jets. The fully developed jet may represent the debris from the quark. Then the smaller jets are the offspring of the antiquark and the gluon. The frequency at which these events are seen and the detailed properties of the events are all consistent with the idea that the mechanism for generating them really is a quark, an antiquark, and a gluon in the semifinal state before the hadrons materialize.

Now I want to talk about polarization effects and the effective charge of the quarks. There will be similar screening effects to the one we discussed for electrodynamics. In this case, since we have three kinds of

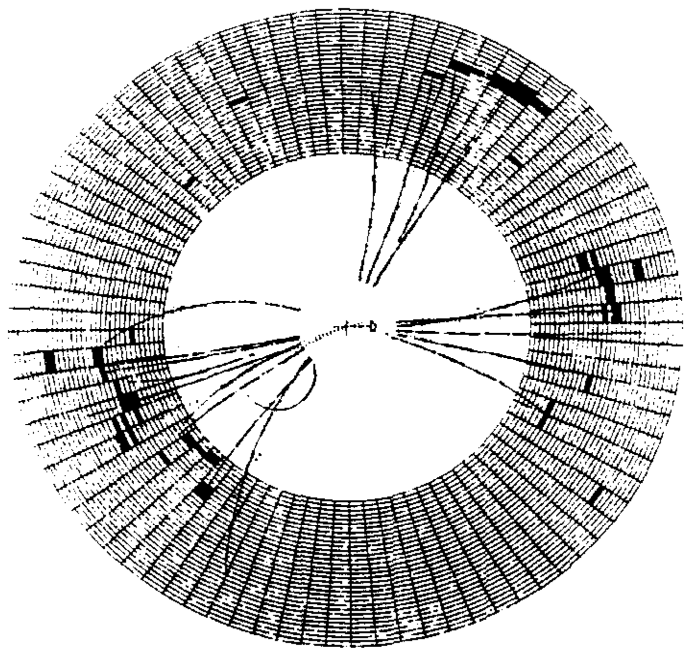


Figure 24: A three-jet event produced in 31 GeV electron-positron annihilations.

charge, I can imagine that the molecules in my analogy are little triangular objects which have a red corner, a green corner, and a blue corner. In the absence of a test charge, they will be oriented in some neutral, disordered way. If I insert a test charge, say a red quark, that will attract the blue and green corners of the surrounding "molecules," and repel the red corners. The resulting arrangement is shown schematically in Fig. 25.

Just as before, if I ask at a certain radius how much redness lies within a circle, the result will be less than the redness of the test charge because some of it is screened out or cancelled by the antiredness from the blue and green corners of the triangles. There is a color charge screening in this case, which tells you that the effective charge tends to become larger as you probe on shorter distance scales. This is entirely analogous to what we saw in QED.

The difference in this case is that there is something else that can happen, because the gluons carry color. Because the gluons carry color, quarks can *camouflage* themselves and hide their color. Fig. 26(a) shows our test charge, the red quark. We now send some emissary in to say,

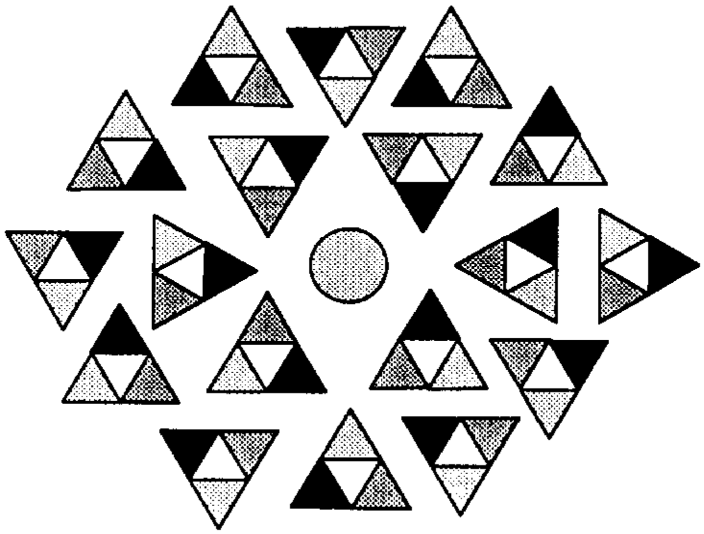


Figure 25: Colored molecules polarized by a test charge.

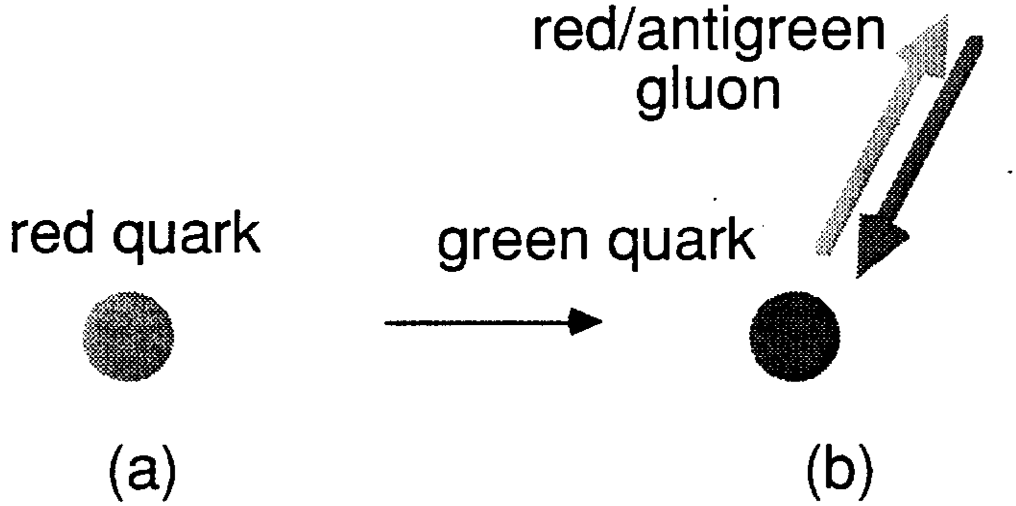


Figure 26: Camouflage.

"Hello, are you red?" While our emissary is on the way in to ask whether this is a red quark, the quark can fluctuate quantum mechanically into a quark and a gluon. And if it chooses to, it can fluctuate into a green quark and a red–antigreen gluon. The red–antigreen gluon goes out and takes a walk in quantum mechanics space, as indicated in Fig. 26(b).

Our probe arrives and says, "Hello, are you red?" And the quark says, "No, I'm green. Go away." So because of the fluctuations made possible by the fact that gluons can carry color you find, if you look too closely, *less* red charge than you thought was there. In order to see the full red charge, you've got to look on a bigger scale, the scale of the promenade of the gluons.

We have two effects going on: one, the normal screening effect as in electrodynamics; the second, the camouflage effect made possible because unlike the photons, which don't have an electric charge, the gluons do have a color charge.

There's a competition between these two effects, and in the theory we believe to be true, QCD, camouflage wins. The consequence of that is that the strong force, as measured by the effective color charge, becomes weaker and weaker at short distances. If you look closer and closer you find that the strong charge is getting tinier and tinier. What this means is that for practical purposes if you find quarks close together in a small space inside a bubble or a little balloon or a bag, they behave almost like independent particles. Because of the camouflage effect, as long as the quarks remain close together, each one hardly feels the color charge of the others.

On the other hand, if you try to separate two quarks by a large distance, then each is able to see more clearly the full charge of the neighboring (but no longer very close) quark. And so the strong force becomes more formidable as you go to large distances. We believe that this effect, properly implemented, is responsible for the fact that we can talk about the quarks as being quasi-free particles within protons, but we can't extract the quarks from the protons. The net *antiscreening* of color charge gives us the possibility of understanding that apparent paradox.

Let me now say just a few words about the theory of weak and electromagnetic interactions. The symmetry that we recognize here is the family symmetry between, say, the electron and its neutrino or the muon and its neutrino. This is a family pattern which seems to be perfectly respected for the leptons and very well respected for the quarks. What we do is to take that family symmetry and combine it with the phase invariance that we saw was a good thing in electromagnetism.

When you do that cleverly, you find that the resulting theory is a rather agreeable one in which the force carriers are the photon, two carriers of the charge-changing weak interactions, W^+ and W^-, plus a fourth force carrier called Z^0. The first three were expected on the basis of previous observations, but there was no evidence for charge-preserving weak interactions at the time the theory was formulated.

The rest, as they say, is history. The new kind of weak interaction which would have been mediated by the Z^0 was in fact discovered in experiments first at CERN, then here and at Brookhaven in 1973. The properties of the new interaction were refined by experiments over the next five years, and had precisely the character outlined by the electroweak gauge theory. Now, those of you alert to the newspapers may recognize that I asserted to you a few moments ago that the carriers of these forces had to be *massless* particles, and yet you have read that the particles carrying the weak interaction, the W and the Z, weigh 100 times as much as a proton.

And so there is something which had to be understood. The great contribution of Mr. Weinberg and Mr. Salam was to understand how to use a phenomenon called *spontaneous symmetry breaking* to change the force carriers to massive particles. Unfortunately, if I'm going to get to the University of Chicago in time for my afternoon class, I won't have time to tell you about that. Ask someone in the discussion section to explain how it works.

This then is the Standard Model (Fig. 27). Let us put aside gravity, since gravitation is generally a weak perturbation on particle physics. We have a few elementary forces, all of the same mathematical character. They are all mediated by spin-one particles whose properties we understand rather well, and which are given to us in large measure by the symmetries that generated the theories. We have a few (although perhaps not few enough) elementary particles. Putting together these elements we should be able to understand everything!

Now, the mathematical similarity of these theories and the observational similarity of quarks and leptons — the fact that apart from the color quantum number they seem to be so similar — invites us to ask,

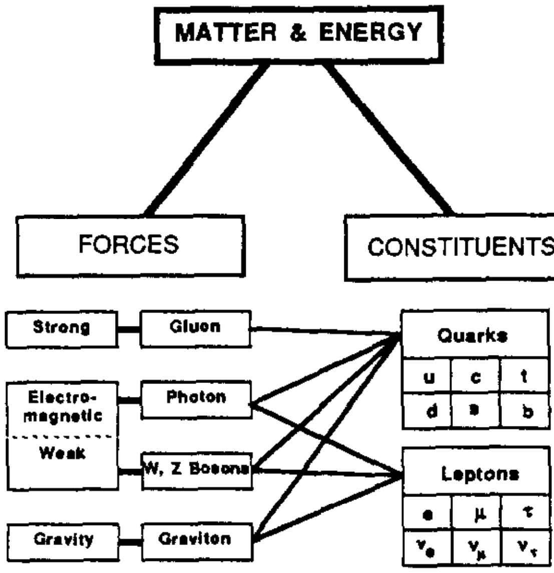


Figure 27: The Standard Model of Particle Physics.

"Is is possible to put the quarks together with the leptons? Is it possible to give a common basis to all these theories?"

The answer to that, at least in principle, is that it seems like a good idea to try, and that we know how to do it by constructing examples of *unified theories,* whether or not these theories turn out to be true. Howard Georgi will tell you more about this opportunity tomorrow.

Where do we stand, then, at the end of all this? We stand in a pretty good place. We have arrived at a fairly simple scheme. It has broad applicability, and if we ask how well are gauge theories tested, it is fair to say that there are no experimental embarrassments. There is no single piece of data that is both true and says the whole idea of this theory or that must be wrong. That is very important.

There are lots of predictions which we have to make sharper by doing the theoretical calculations better, and many others for which the experiments have been very difficult or outside the range of our instruments, and those must be tested better. QED, of course, is the standard by which we judge other theories; it is really very good. For the electroweak theory, the tests are becoming very quantitative at the level of tenths of a percent, but we still need to test it further. And for quantum chromodynamics, a few tests are becoming quantitative and we are learning how to do better.

For unified theories in which we try to put everything together, as you'll see in Howard's lectures, the questions that we're asking still are at the level of yes and no: Is this a good idea? Do its essential consequences actually follow in Nature? We're not really at the level of comparing prediction with observation for numbers.

Let me now take just a couple of minutes to talk about problems. I've sketched for you today this edifice (Fig. 27) of elementary particles and forces, and tried to convey to you a certain enthusiasm for the style of arguments that led us here, and a certain respect for the success of the theories in making predictions about the world around us. When confronted with that success we may briefly celebrate with our colleagues who invented the ideas and made the observations from which they sprang. But then you have to ask yourself, if this theory works so well, why is it working so well? Is it really internally consistent? What are the reasons that it goes together so well? That kind of questioning represents a whole range of activities now going on.

In addition to that, once you've solved every problem in sight, you can look at the few problems you didn't solve which are now made approachable by virtue of the last problem you solved. Let me mention a couple of those.

The theory of the weak and electromagnetic interactions helps us understand why quarks and leptons have masses. If you work out the formalism in more detail than I did this morning, you find that going through the arguments about recognizing a symmetry and then hiding, or spontaneously breaking, the symmetry, that you're given little spaces in the equations, little empty boxes where the mass of the electron belongs, and the mass of the muon, and the mass of the up quark, and so on. That's very good because until the invention of that theory, you didn't have those little boxes to write the numbers in — you didn't know how the masses could come about.

So that's progress. It's incomplete progress because nobody tells you from first principles what numbers to write in the boxes. At the moment, it is still information you have to take from experiment. It would be nice if we had a more complete theory in which we were told not merely that here's a box to put a number in, but here's the number to write in it or here's how to compute the number that goes in there.

Another annoyance: we have several sets of quarks and leptons, but for ordinary experience all we need is the electron and the up and down quarks that make up protons and neutrons. Crudely speaking, that's enough to account for us, so why should we need these other things? There are indications from the internal consistency of the electroweak theory that quarks and leptons do go together in some way. But why they come together in the way we observe, why there are three sets of them, whether there are more, all is outside the scope of present theories.

A lot of the complaints we have about the standard model have to do with *arbitrariness,* the general problem of having boxes that need numbers written in them. So, to engage in a little self-flagellation, let's count the parameters of the standard model. It doesn't matter that I haven't told you set what some of them are; you'll get the picture. There are three coupling strengths for the strong, weak, and electromagnetic interactions, six quark masses, three numbers that describe how the weak interactions of the quarks cross family lines, something called the CP-violating phase, two parameters of something called the Higgs potential, three masses for the electron, muon, and tau, and one number called a vacuum phase (you don't care what that is). But if you add them all up, it's a big number, namely 19.

If I go further and make a unified theory of the strong, weak, and electromagnetic interactions, I get some interrelations between parameters, but in order to build such a theory I need to introduce some new parameters. So the number is still around twenty. That seems not completely satisfying.

The other thing you can do is to count up the number of fundamental fields. Leaving aside the discovery of the top quark (which will come sometime soon), there are 15 quarks if you count all the colors, six leptons, one photon, three intermediate bosons W^+, W^-, and Z^0, eight colored gluons, a Higgs boson, and a graviton. [That last one is to show I'm not hopelessly reactionary.] This too is a total which exceeds the number of fingers and toes of a single theoretical physicist.

Well, there are lots of speculations about how to make our present theories more complete, and how to go beyond them, and all of us are hard at work on that. In addition to theoretical work, the other thing we need is clearly to get more experimental information, and to do this at the highest possible energies — the shortest possible distances. Part of the beauty of the current framework is that it is good enough, it needs to be taken seriously enough, that we can trust it to tell us when it doesn't work any more. In the case of the electroweak theory, the frontier is particularly well defined. From general arguments about the structure of the theory as it now stands, from every invention we've made to go beyond the standard electroweak theory, there is an indication that new and important clues have to be found in collisions of the fundamental particles at energies around 1 TeV, 10^{12} electron volts.

Because of this, when I'm not standing here in the Fermilab auditorium, one of the ways I occupy my time is in trying to convince the taxpayers of the United States that they should build for us an instrument to explore the 1 TeV scale. The device we have in mind is a large superconducting proton-proton collider. We want to have energies of 20 TeV per beam so that quarks and gluons and other things inside the proton will themselves carry several TeV into the elementary collisions. We use superconducting magnets to make a strong magnetic field to confine the protons in a relatively small circle as we're accelerating them, and also to lower the power consumption. The present design calls for magnets of about 6.5 Tesla.

How big is this device? Well, you all know the formula for the radius of curvature of a charged particle moving in a magnetic field. You may not know it in the appropriate engineering units, which are[7]

$$\text{Radius} = \frac{10}{3}\text{km} \cdot \frac{\text{Beam Momentum}}{\text{TeV}/c} \div \frac{\text{Magnetic Field}}{\text{Tesla}}. \qquad (2.13)$$

And so for a 20 TeV beam in 5 Tesla magnets, the radius of curvature would be about 13 km. If you make allowances for straight sections in which to do the experiments and the acceleration, this is a device which is about twenty miles in diameters. It is a large undertaking, and we are taking care to propose it in a sensible and responsible way.

I show you in Fig. 28 that it is not completely out of scale with human experience and human structures. At the left of the picture you can see the size of the Fermilab ring, a four-mile circle you can jog around during your visit. The largest circle shows the size of the supercollider we would like to build. The irregular loop is the Washington Beltway. You can see

[7] In Congressional Units, (10/3) km = 2 miles.

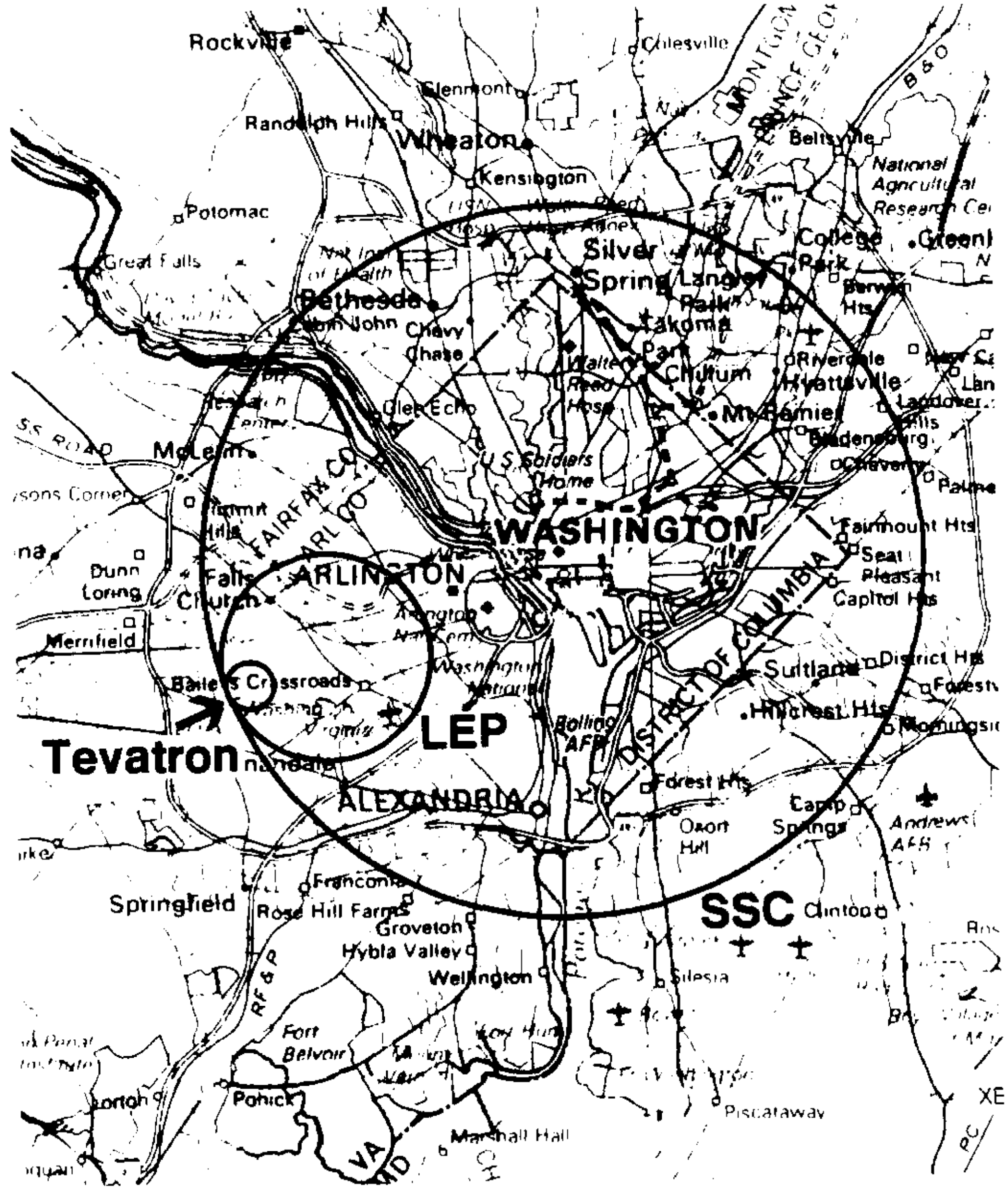


Figure 28: The Superconducting Super Collider and two smaller colliders, LEP at CERN and the Tevatron at Fermilab, superimposed to scale on the environs of Washington, D.C.

that they are about the same size. If only they had built that highway in the right shape, we would already have a site for our next accelerator!

Suggested Reading

C. Quigg, "Elementary Particles and Forces," *Scientific American* **252**, (4) 84 (April, 1985).

S. Weinberg, *The Discovery of Subatomic Particles,* W. H. Freeman, San Francisco, 1983.

J. Mulvey (editor), *The Nature of Matter,* Oxford University Press, Oxford, 1981.

F. E. Close, *The Cosmic Onion: quarks and the nature of the universe,* Heinemann Educational Books, London and Exeter, N. H., 1983.

C. Sutton, *The Particle Connection,* Simon and Schuster, New York, 1984.

C. T. Hill, "Quarks and Leptons: it's Elementary," *The Science Teacher,* September and October, 1982.

Physics through the 90s: Elementary Particle Physics, National Academy Press, Washington, 1986.

E. W. Kolb and C. Quigg, "Exploring the Universe from Quarks to Cosmology," *The Physics Teacher,* December, 1986.

C. Quigg and R. F. Schwitters, "Elementary Particle Physics and the Superconducting Super Collider," *Science* **231**, 1522 (March 28, 1986).

K. Gottfried and V. F. Weisskopf, *Concepts of Particle Physics, vol. 1,* Oxford University Press, Oxford, 1984.

D. H. Perkins, *Introduction to High-Energy Physics, second edition,* Addison-Wesley, Reading, Massachusetts, 1982.

I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics,* Adam Hilger, Bristol, 1982.

C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions,* Benjamin/Cummings, Reading, Massachusetts, 1983.