INTRODUCTION TO ELEMENTARY PARTICLES

David Griffiths
Reed College

WILEY

JOHN WILEY & SONS, INC.
New York • Chichester • Brisbane • Toronto • Singapore

INTRODUCTION TO ELEMENTARY PARTICLES

Library of Congress Cataloging-in-Publication Data

Griffiths, David J. (David Jeffrey), 1942–
Introduction to elementary particles.

Includes bibliographies and index.
1. Particles (Nuclear physics) I. Title.
QC793.2.G75 1987 539.7'21 86-25709
ISBN 0-471-60386-4

Printed and bound in the United States of America

20 19 18 17 16 15 14

CONTENTS

PREFACE

This introduction to the theory of elementary particles is intended primarily for advanced undergraduates who are majoring in physics. Most of my colleagues consider this subject inappropriate for such an audience—mathematically too sophisticated, phenomelogically too cluttered, insecure in its foundations, and uncertain in its future. Ten years ago I would have agreed. But in the last decade the dust has settled to an astonishing degree, and it is fair to say that elementary particle physics has come of age. Although we obviously have much more to learn, there now exists a coherent and unified theoretical structure that is simply too exciting and important to save for graduate school or to serve up in diluted qualitative form as a subunit of modern physics. I believe the time has come to integrate elementary particle physics into the standard undergraduate curriculum.

Unfortunately, the research literature in this field is clearly inaccessible to undergraduates, and although there are now several excellent graduate texts, these call for a strong preparation in advanced quantum mechanics, if not quantum field theory. At the other extreme, there are many fine popular books and a number of outstanding *Scientific American* articles. But very little has been written specifically for the undergraduate. This book is an effort to fill that need. It grew out of a one-semester elementary particles course I have taught from time to time at Reed College. The students typically had under their belts a semester of electromagnetism (at the level of Lorrain and Corson), a semester of quantum mechanics (at the level of Park), and a fairly strong background in special relativity.

In addition to its principal audience, I hope this book will be of use to beginning graduate students, either as a primary text, or as preparation for a more sophisticated treatment. With this in mind, and in the interest of greater completeness and flexibility, I have included more material here than one can comfortably cover in a single semester. (In my own courses I ask the students to read Chapters 1 and 2 on their own, and begin the lectures with Chapter 3. I skip Chapter 5 altogether, concentrate on Chapters 6 and 7, discuss the first two sections of Chapter 8, and then jump to Chapter 10). To assist the reader (and the teacher) I begin each chapter with a brief indication of its purpose and content, its prerequisites, and its role in what follows.

This book was written while I was on sabbatical at the Stanford Linear Accelerator Center, and I would like to thank Professor Sidney Drell and the other members of the Theory Group for their hospitality.

DAVID GRIFFITHS

Introduction

ELEMENTARY PARTICLE PHYSICS

Elementary particle physics addresses the question, "What is matter made of?" on the most fundamental level—which is to say, on the smallest scale of size. It's a remarkable fact that matter at the subatomic level consists of tiny chunks, with vast empty spaces in between. Even more remarkable, these tiny chunks come in a small number of different types (electrons, protons, neutrons, pi mesons, neutrinos, and so on), which are then replicated in astronomical quantities to make all the "stuff" around us. And these replicas are absolutely perfect copies—not just "pretty similar," like two Fords coming off the same assembly line, but utterly *indistinguishable.* You can't stamp an identification number on an electron, or paint a spot on it—if you've seen one, you've seen them all. This quality of absolute identicalness has no analog in the macroscopic world. (In quantum mechanics it is reflected in the Pauli exclusion principle.) It enormously simplifies the task of elementary particle physics: we don't have to worry about big electrons and little ones, or new electrons and old ones—an electron is an electron is an electron. It didn't have to be so easy.

My first job, then, is to introduce you to the various kinds of elementary particles, the actors, if you will, in the drama. I could simply *list* them, and tell you their properties (mass, electric charge, spin, etc.), but I think it is better in this case to adopt a historical perspective, and explain how each particle first came on the scene. This will serve to endow them with character and personality, making them easier to remember and more interesting to watch. Moreover, some of the stories are delightful in their own right.

Once the particles have been introduced, in Chapter 1, the issue becomes, "How do they interact with one another?" This question, directly or indirectly, will occupy us for the rest of the book. If you were dealing with two *macroscopic* objects, and you wanted to know how they interact, you would probably begin by suspending them at various separation distances and measuring the force between them. That's how Coulomb determined the law of electrical repulsion between two charged pith balls, and how Cavendish measured the gravitational attraction of two lead weights. But you can't pick up a proton with tweezers or tie an electron onto the end of a piece of string; they're just too small. For practical reasons, therefore, we have to resort to less direct means to probe the interactions of elementary particles. As it turns out, almost all our experimental information comes from three sources: (1) scattering events, in which we fire one particle at another and record (for instance) the angle of deflection; (2) decays, in which a particle spontaneously disintegrates and we examine the debris; and (3) bound states, in which two or more particles stick together, and we study the properties of the composite object. Needless to say, determining the interaction law from such indirect evidence is not a trivial task. Ordinarily, the procedure is to *guess* a form for the interaction and compare the resulting theoretical calculations with the experimental data.

The formulation of such a guess ("model" is a more respectable term for it) is guided by certain general principles, in particular, special relativity and quantum mechanics. In the diagram below I have indicated the four realms of mechanics:

	Small →	
Fast ↓	Classical mechanics	Quantum mechanics
	Relativistic mechanics	Quantum field theory

The world of everyday life, of course, is governed by classical mechanics. But for objects that travel very fast (at speeds comparable to c), the classical rules are modified by special relativity, and for objects that are very small (comparable to the size of atoms, roughly speaking), classical mechanics is superseded by quantum mechanics. Finally, for things that are both fast *and* small, we require a theory that incorporates relativity and quantum principles: quantum field theory. Now, elementary particles *are* extremely small, of course, and typically they are also very fast. So elementary particle physics naturally falls under the dominion of quantum field theory.

Please observe the distinction here between a *type of mechanics* and a *particular force law*. Newton's law of universal gravitation, for example, describes a specific interaction (gravity), whereas Newton's three laws of motion define a mechanical system (classical mechanics), which (within its jurisdiction) governs *all* interactions. The force law tells you what F *is*, in the case at hand; the mechanics tells you how to *use* F to determine the motion. The goal of elementary particle dynamics, then, is to guess a set of force laws which, within the context of quantum field theory, correctly describe particle behavior.

However, some general features of this behavior have nothing to do with the detailed form of the interactions. Instead they follow directly from relativity, from quantum mechanics, or from the combination of the two. For example, in relativity, energy and momentum are always conserved, but (rest) mass is not. Thus the decay $\Delta \rightarrow p + \pi$ is perfectly acceptable, even though the Δ weighs more than the sum of p plus π. Such a process would not be possible in classical mechanics, where mass is strictly conserved. Moreover, relativity allows for particles of zero (rest) mass—the very idea of a massless particle is nonsense in classical mechanics—and as we shall see, photons, neutrinos, and gluons are all (apparently) massless.

In quantum mechanics a physical system is described by its *state*, s (represented by the wave function ψ_s in Schrodinger's formulation, or by the *ket* $|s\rangle$ in Dirac's). A physical process, such as scattering or decay, consists of a *transition* from one state to another. But in quantum mechanics the outcome is not uniquely determined by the initial conditions; all we can hope to calculate, in general, is the *probability* for a given transition to occur. This indeterminacy is reflected in the observed behavior of particles. For example, the charged pi meson ordinarily disintegrates into a muon plus a neutrino, but occasionally one will decay into an *electron* plus a neutrino. There's no difference in the original pi mesons; they're all identical. It is simply a fact of nature that a given particle can go either way.

Finally, the union of relativity and quantum mechanics brings certain extra dividends that neither one by itself can offer: the existence of antiparticles, a proof of the Pauli exclusion principle (which in nonrelativistic quantum mechanics is simply an ad hoc hypothesis), and the so-called *TCP theorem*. I'll tell you more about these later on; my purpose in mentioning them here is to emphasize that these are features of the mechanical system itself, not of the particular model. Short of a catastrophic revolution, they are untouchable. By the way, quantum field theory in all its glory is difficult and deep, but don't be alarmed: Feynman invented a beautiful and intuitively satisfying formulation that is not hard to learn; we'll come to that in Chapter 6. (The *derivation* of Feynman's rules from the underlying quantum field theory is a different matter, which can easily consume the better part of an advanced graduate course, but this need not concern us here.)

In the last few years a theory has emerged that describes all of the known elementary particle interactions except gravity. (As far as we can tell, gravity is much too weak to play any significant role in ordinary particle processes.) This theory—or, more accurately, this collection of related theories, incorporating quantum electrodynamics, the Glashow–Weinberg–Salam theory of electroweak processes, and quantum chromodynamics—has come to be called the *Standard Model*. No one pretends that the Standard Model is the final word on the subject, but at least we now have (for the first time) a full deck of cards to play with. Since 1978, when the Standard Model achieved the status of "orthodoxy," it has met every experimental test. It has, moreover, an attractive aesthetic feature: in the Standard Model all of the fundamental interactions derive from a single general principle, the requirement of *local gauge invariance*. It seems likely that future developments will involve extensions of the Standard Model, not its repudiation. This book might be called an "Introduction to the Standard Model."

As that alternative title suggests, this is a book about elementary particle *theory,* with very little on experimental methods or instrumentation. These are important matters, and an argument can be made for integrating them into a text such as this, but they can also be distracting and interfere with the clarity and elegance of the theory itself. (I encourage you to read about experimental aspects of the subject, and from time to time I will refer you to particularly accessible accounts.) For now, I'll confine myself to scandalously brief answers to the two most obvious experimental questions.

HOW DO YOU *PRODUCE* ELEMENTARY PARTICLES?

Electrons and protons are no problem; these are the stable constituents of ordinary matter. To produce electrons one simply heats up a piece of metal, and they come boiling off. If one wants a *beam* of electrons, one then sets up a positively charged plate nearby, to attract them over, and cuts a small hole in it; the electrons that make it through the hole constitute the beam. Such an *electron gun* is the starting element in a television tube or an oscilloscope or an electron accelerator (Fig. I.1).

To obtain protons you ionize hydrogen (in other words, strip off the electron). In fact, if you're using the protons as a *target,* you don't even need to bother about the electrons; they're so light that an energetic particle coming in will knock them out of the way. Thus, a tank of hydrogen *is* essentially a tank of protons. For more exotic particles there are three main sources: cosmic rays, nuclear reactors, and particle accelerators.

Cosmic Rays The earth is constantly bombarded with high-energy particles (principally protons) coming from outer space. What the source of these particles might be remains something of a mystery; at any rate, when they hit atoms in the upper atmosphere they produce showers of secondary particles (mostly muons, by the time they reach ground level), which rain down on us all the time. As a source of elementary particles, cosmic rays have two virtues: they are free, and their energies can be enormous—far greater than we could possibly produce in the laboratory. But they have two major disadvantages: The rate at which they strike any detector of reasonable size is very low, and they are completely uncontrollable. So cosmic ray experiments call for patience and luck.

Nuclear Reactors When a radioactive nucleus disintegrates, it may emit a variety of particles—neutrons, neutrinos, and what used to be called alpha rays (actually, alpha *particles,* which are bound states of two neutrons plus two protons), beta rays (actually, electrons or positrons), and gamma rays (actually, photons).

Particle Accelerators You start with electrons or protons, accelerate them to high energy, and smash them into a target. By skillful arrangements of absorbers and magnets, you can separate out of the resulting debris the particle species you wish to study. Nowadays it is possible in this way to generate intense sec-

Figure I.1 The Stanford Linear Accelerator Center (SLAC). Electrons and positrons are accelerated down a straight tube 2 miles long, reaching energies as high as 45 GeV. (Photo courtesy of SLAC.)

ondary beams of positrons, muons, pions, kaons, and antiprotons, which in turn can be fired at another target. The stable particles—electrons, protons, positrons, and antiprotons—can even by fed into giant *storage rings* in which, guided by powerful magnets, they circulate at high speed for hours at a time, to be extracted and used at the required moment (Fig. I.2).

In general, the heavier the particle you want to produce, the higher must

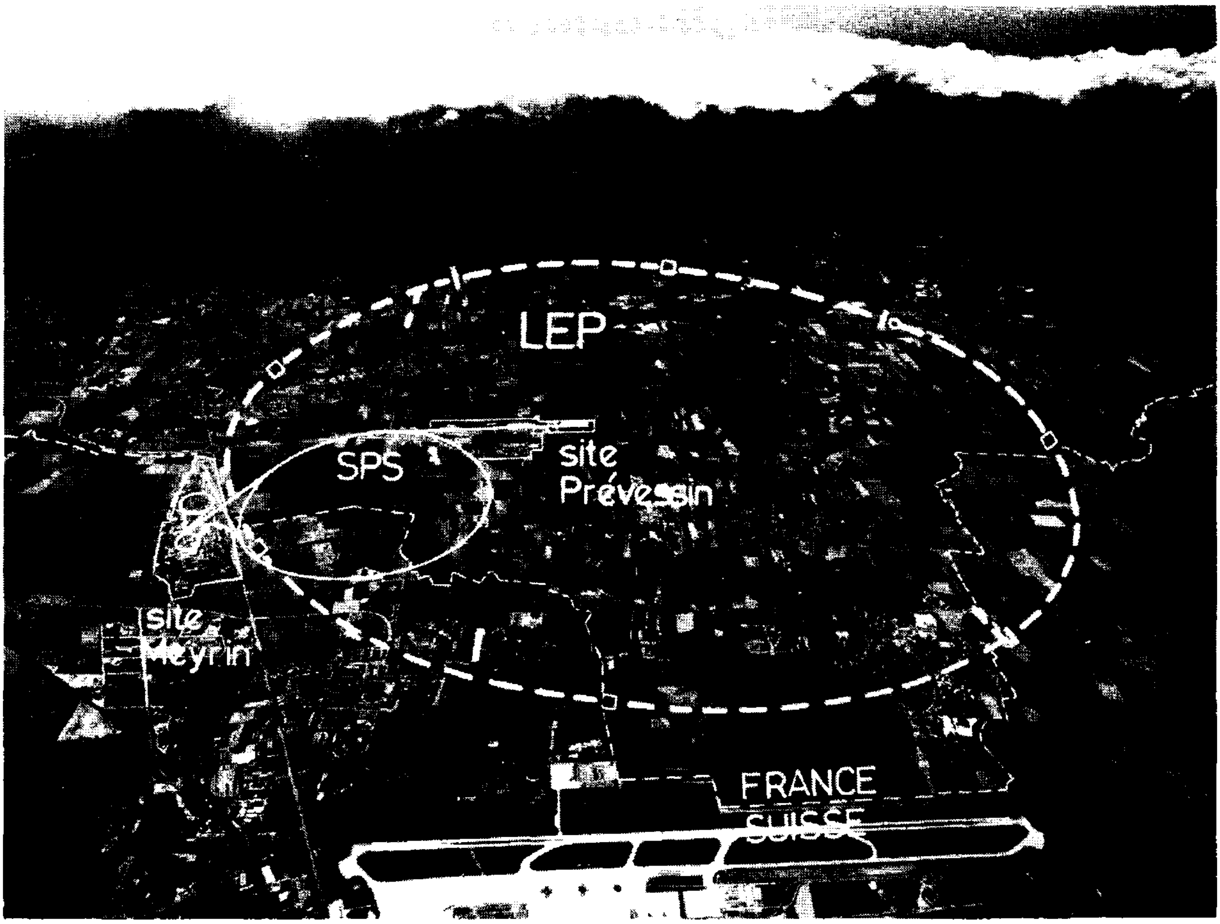


Figure I.2 CERN, outside Geneva, Switzerland. SPS is the 450 GeV Super Proton Synchrotron, later modified to make a proton–antiproton collider; LEP is a 50 GeV electron–positron storage ring now under construction. (Photo courtesy of CERN.)

be the energy of the collision. That's why, historically, lightweight particles tend to be discovered first, and as time goes on, and accelerators become more powerful, heavier and heavier particles are found. At present, the heaviest known particle is the Z^0, with nearly 100 times the mass of the proton. It turns out that the particle gains enormously in energy if you collide two high-speed particles head-on, as opposed to firing one particle at a stationary target. (Of course, this calls for much better aim!) Therefore, most contemporary experiments involve colliding beams from intersecting storage rings; if the particles miss on the first pass, they can try again the next time around. Indeed, with electrons and positrons (or protons and antiprotons) the *same ring* can be used, with the plus charges circulating in one direction and the minus charges in the other.

There is another reason why particle physicists are always pushing for higher energies: In general, the higher the energy of the collision, the closer the two particles come to one another. So if you want to study the interaction at very short range, you need very energetic particles. In quantum-mechanical terms, a particle of momentum p has an associated wavelength λ given by the de Broglie formula $\lambda = h/p$, where h is Planck's constant. At large wavelengths (low momenta) you can only hope to resolve relatively large structures; in order to examine something extremely small, you need comparably short wavelengths, and hence high momenta. If you like, consider this a manifestation of the uncertainty principle ($\Delta x\, \Delta p \geqslant h/4\pi$)—to make Δx small, Δp must be large. However you look at it, the conclusion is the same: to probe *small distances* you need *high energies.*

HOW DO YOU *DETECT* ELEMENTARY PARTICLES?

There are many kinds of particle detectors—Geiger counters, cloud chambers, bubble chambers, spark chambers, photographic emulsions, Čerenkov counters, scintillators, photomultipliers, and so on (Fig. I.3). Actually, a typical modern detector has whole arrays of these devices, wired up to a computer that tracks the particles and displays their trajectories on a television screen (Fig. I.4). The details do not concern us, but there is one thing to be aware of: Most detection mechanisms rely on the fact that when high-energy charged particles pass through matter they ionize atoms along their path. The ions then act as "seeds" in the formation of droplets (cloud chamber) or bubbles (bubble chamber) or sparks (spark chamber), as the case may be. But electrically *neutral* particles do not cause ionization, and they leave no tracks. If you look at the bubble chamber photograph in Fig. 1.11, for instance, you will see that the five neutral particles are "invisible"; their paths have been reconstructed by analyzing the tracks of the *charged* particles in the picture and invoking conservation of energy and momentum at each vertex. Notice also that most of the tracks in the picture are *curved* (actually, *all* of them are, to some extent; try holding a ruler up to one you think is straight). Evidently the bubble chamber was placed between the poles of a giant magnet. In a magnetic field B, a particle of charge q and momentum p will move in a circle of radius R given by the famous *cyclotron formula:* $R = pc/qB$, where c is the speed of light. The curvature of the track in a known

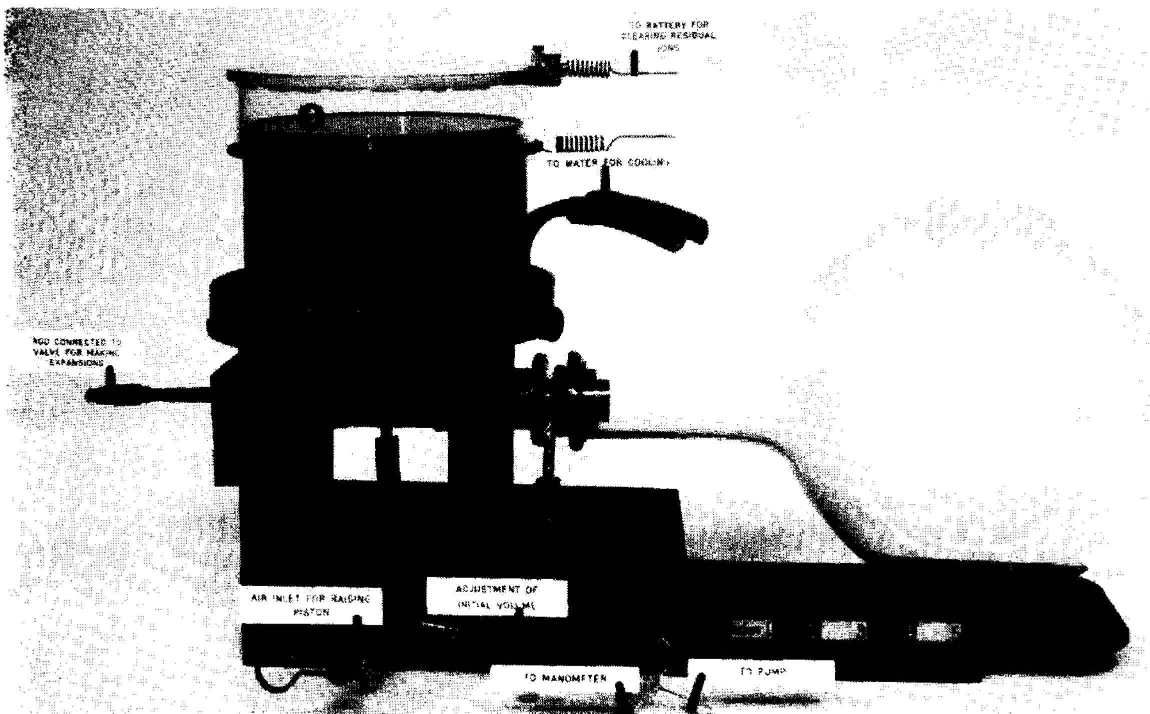


Figure I.3 An early particle detector: Wilson's cloud chamber (ca. 1900). (Photo courtesy Science Museum, London.)

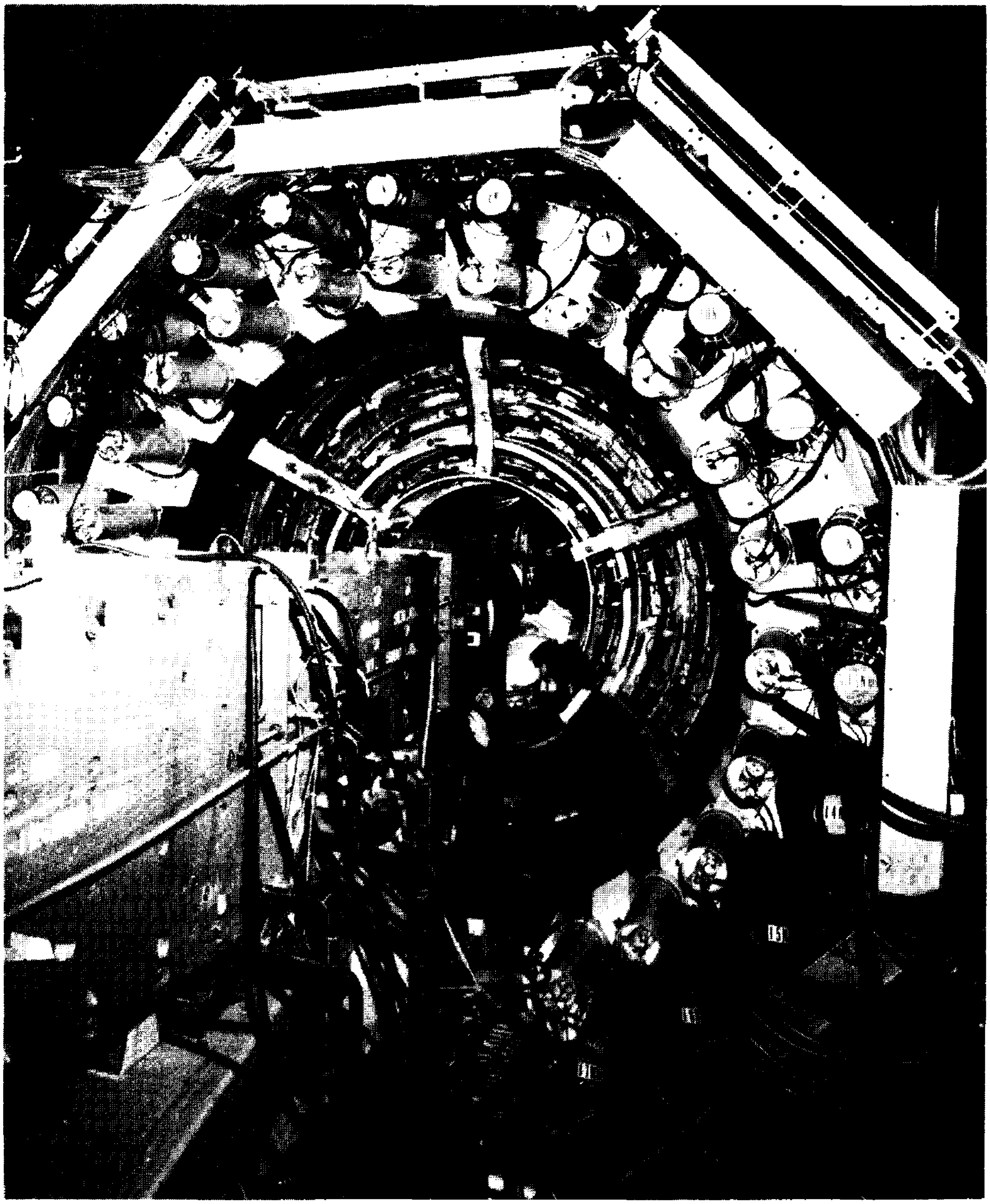


Figure I.4 A modern particle detector: The Mark I, at SLAC. (Photo courtesy SLAC.)

magnetic field thus affords a very simple measure of the particle's momentum. Moreover, we can immediately tell the *sign* of the charge from the *direction* of the curve.

UNITS

Elementary particles are small, so for our purposes the *normal* mechanical units—grams, ergs, joules, and so on—are inconveniently large. Atomic physicists introduced the *electron volt*—the energy acquired by an electron when accelerated through a potential difference of 1 volt: 1 eV = 1.6×10^{-19} joules. For us the eV is inconveniently *small,* but we're stuck with it. Nuclear physicists use keV (10^3 eV); typical energies in particle physics are MeV (10^6 eV), GeV (10^9 eV), or even TeV (10^{12} eV). Momenta are measured in MeV/c (or GeV/c, or whatever), and masses in MeV/c^2. Thus the proton weighs 938 MeV/c^2 = 1.67×10^{-24} g.

Actually, particle theorists are lazy (or clever, depending on your point of view)—they seldom include the c's and $\hbar$'s ($\hbar \equiv h/2\pi$) in their formulas. You're just supposed to fit them in for yourself at the end, to make the dimensions come out right. As they say in the business, "set $c = \hbar = 1$." This amounts to working in units such that time is measured in centimeters and mass and energy in inverse centimeters; the unit of time is the time it takes light to travel 1 centimeter, and the unit of energy is the energy of a photon whose wavelength is 2π centimeters. Only at the end of the problem do we revert to conventional units. This makes everything look very elegant, but I thought it would be wiser in this book to keep all the c's and $\hbar$'s where they belong, so that you can check for dimensional consistency as you go along. (If this offends you, remember that it is easier for you to ignore an $\hbar$ you don't like than for someone else to conjure one up in just the right place.)

Finally, there is the question of what units to use for electric charge. In introductory physics courses most instructors favor the *SI* system, in which charge is measured in *coulombs,* and Coulomb's law reads

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2} \qquad \text{(SI)}$$

Most advanced work is done in the *Gaussian* system, in which charge is measured in *electrostatic units* (esu), and Coulomb's law is written

$$F = \frac{q_1 q_2}{r^2} \qquad \text{(G)}$$

But elementary particle physicists prefer the *Heaviside–Lorentz* system, in which Coulomb's law takes the form

$$F = \frac{1}{4\pi}\frac{q_1 q_2}{r^2} \qquad \text{(HL)}$$

The three units of charge are related as shown:

$$q_{\text{HL}} = \sqrt{4\pi}\, q_{\text{G}} = \frac{1}{\sqrt{\epsilon_0}}\, q_{\text{SI}}$$

In this book I shall use Gaussian units exclusively, in order to avoid unnecessary confusion in an already difficult subject. Whenever possible I will express results in terms of the *fine structure constant*

$$\alpha = \frac{e^2}{\hbar c} = \frac{1}{137}$$

where e is the charge of the electron in Gaussian units. Most elementary particle texts write this as $e^2/4\pi$, because they are measuring charge in Heaviside–Lorentz units and setting $c = \hbar = 1$; but everyone agrees that the *number* is $\frac{1}{137}$.

REFERENCES AND NOTES

This book is a brief survey of an enormous and rapidly changing subject. My aim is to introduce you to some important ideas and methods, to give you a sense of what's out there to be learned, and perhaps to stimulate your appetite for more. If you want to read further in quantum field theory, I particularly recommend:

Bjorken, J. D., and S. D. Drell. *Relativistic Quantum Mechanics* and *Relativistic Quantum Fields.* New York: McGraw-Hill, 1964.

Sakurai, J. J. *Advanced Quantum Mechanics.* Reading, MA: Addison-Wesley, 1967.

Itzykson, C., and J.-B. Zuber. *Quantum Field Theory.* New York: McGraw-Hill, 1980.

I warn you, however, that these are all difficult and advanced books. For elementary particle physics itself, the following books (listed in order of increasing difficulty) are especially useful:

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

Frauenfelder, H., and E. M. Henley. *Subatomic Physics.* Englewood Cliffs, NJ: Prentice-Hall, 1974.

Perkins, D. H. *Introduction to High-Energy Physics,* 2d Ed. Reading, MA: Addison-Wesley, 1982.

Halzen, F., and A. D. Martin. *Quarks and Leptons.* New York: Wiley, 1984.

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

Close, F. E. *An Introduction to Quarks and Partons.* London: Academic, 1979.

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

Cheng, T.-P., and L.-F. Li. *Gauge Theories of Elementary Particle Physics.* New York: Oxford University Press, 1984.

Chapter 1

Historical Introduction to the Elementary Particles

This chapter is a kind of "folk history" of elementary particle physics. Its purpose is to provide a sense of how the various particles were first discovered, and how they fit into the overall scheme of things. Along the way some of the fundamental ideas that dominate elementary particle theory are explained. This material should be read quickly, as background to the rest of the book. (As history, the picture presented here is certainly misleading, for it sticks closely to the main track, ignoring the false starts and blind alleys that accompany the development of any science. That's why I call it "folk" history—it's the way particle physicists like to remember the subject—a succession of brilliant insights and heroic triumphs unmarred by foolish mistakes, confusion, and frustration. It wasn't really quite so easy.)

1.1 THE CLASSICAL ERA (1897–1932)

It is always a little artificial to pinpoint such things, but I'd say that elementary particle physics was born in 1897, with J. J. Thomson's discovery of the electron.[1] (It is fashionable to carry the story all the way back to Democritus and the Greek atomists, but apart from a few suggestive words their metaphysical speculations have nothing in common with modern science, and although they may be of modest antiquarian interest, their relevance is infinitesimal.) Thomson knew that *cathode rays* emitted by a hot filament could be deflected by a magnet. This suggested that they carried electric charge; in fact, the direction of the curvature required that the charge be negative. It seemed, therefore, that these were not rays at all, but rather streams of particles. By passing the beam through crossed electric and magnetic fields, and adjusting the field strength until the net deflection was zero, Thomson was able to determine the velocity of the particles (about a tenth the speed of light) as well as their charge-to-mass ratio. (See Fig. 1.1 and Problem 1.1). This ratio turned out to be enormously greater than for any known ion, indicating that either the charge was extremely large or the mass was very small. Indirect evidence pointed to the second conclusion. Thomson called the particles *corpuscles,* and their charge the *electron.* Later the word electron was applied to the particles themselves.

Thomson correctly surmised that these electrons were essential constituents of atoms; however, since atoms as a whole are electrically neutral and very much heavier than electrons, there immediately arose the problem of how the compensating plus charge—and the bulk of the mass—is distributed within an atom. Thomson himself imagined that the electrons were suspended in a heavy, positively charged paste, like (as he put it) the plums in a pudding. But Thomson's model was decisively repudiated by Rutherford's famous scattering experiment, which showed that the positive charge, and most of the mass, was concentrated in a tiny core, or *nucleus,* at the center of the atom. Rutherford demonstrated this by firing a beam of α-particles (ionized helium atoms) into a thin sheet of

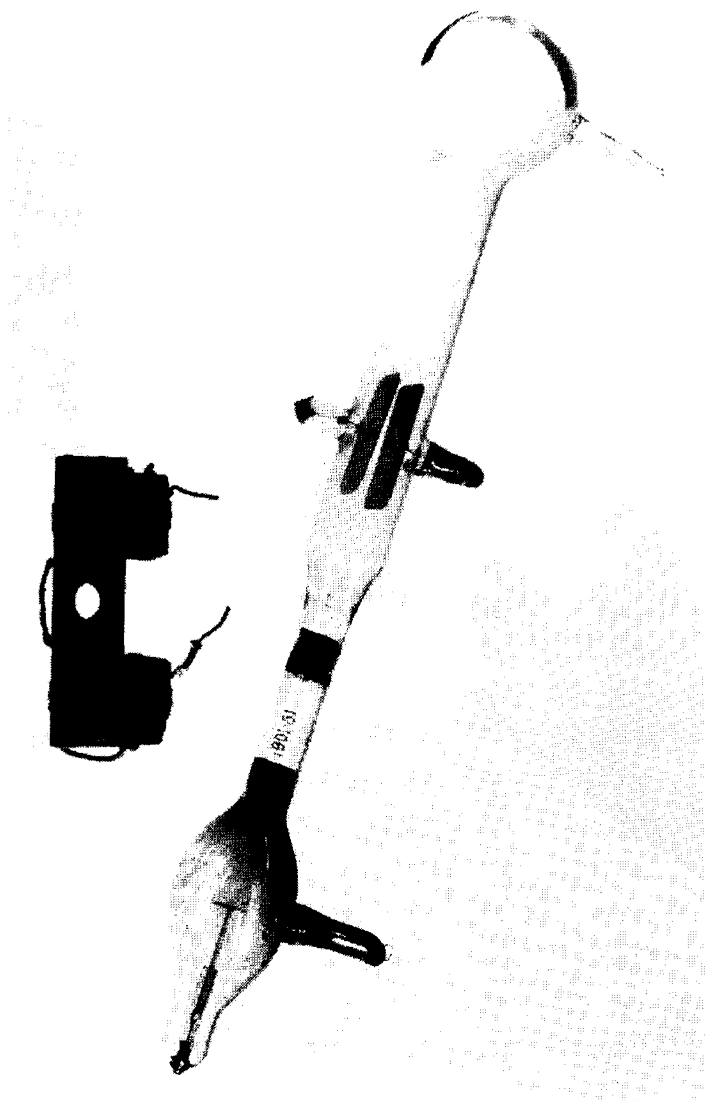


Figure 1.1 The apparatus with which J. J. Thomson discovered the electron. (Photo courtesy Science Museum, London.)

gold foil (see Fig. 1.2). Had the gold atoms consisted of rather diffuse spheres, as Thomson supposed, then all of the α-particles should have been deflected a bit, but none would have been deflected much—any more than a bullet is deflected much when it passes, say, through a bag of sawdust. What in *fact* occurred was that *most* of the α-particles passed through the gold completely undisturbed, but a few of them bounced off at wild angles. Rutherford's conclusion was that the α-particles had encountered something very small, very hard, and very heavy. Evidently the positive charge, and virtually all of the mass, was concentrated at the center, occupying only a tiny fraction of the volume of the atom (the electrons are too light to play any role in the sattering; they are knocked right out of the way by the much heavier α-particles).

The nucleus of the lightest atom (hydrogen) was given the name *proton* by Rutherford. In 1914 Niels Bohr proposed a model for hydrogen consisting of a single electron circling the proton, rather like a planet going around the sun, held in orbit by the mutual attraction of opposite charges. Using a primitive version of the quantum theory, Bohr was able to calculate the spectrum of hydrogen, and the agreement with experiment was nothing short of spectacular. It was natural then to suppose that the nuclei of heavier atoms were composed of two or more protons bound together, supporting a like number of orbiting electrons. Unfortunately, the next heavier atom (helium), although it does indeed carry two electrons, weighs *four* times as much as hydrogen, and lithium (three electrons) is *seven* times the weight of hydrogen, and so it goes. This dilemma

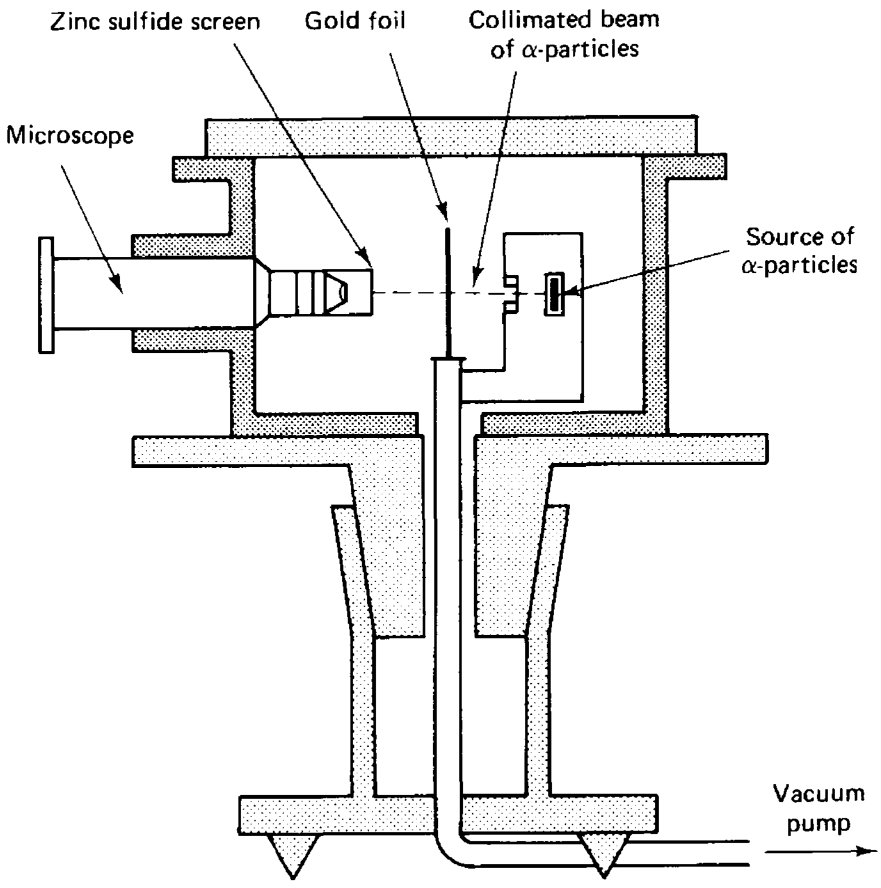


Figure 1.2 Schematic diagram of the apparatus used in the Rutherford scattering experiment. Alpha particles scattered by the gold foil strike a fluorescent screen, giving off a flash of light, which is observed visually through a microscope.

was finally resolved in 1932 with Chadwick's discovery of the neutron—an electrically neutral twin to the proton. The helium nucleus, it turns out, contains two neutrons in addition to the two protons; lithium evidently includes four; and in general the heavier nuclei carry very roughly the same number of neutrons as protons. (The number of neutrons is in fact somewhat flexible: the same atom, chemically speaking, may come in several different *isotopes*, all with the same number of protons, but with varying numbers of neutrons.)

The discovery of the neutron put the final touch on what we might call the *classical period* in elementary particle physics. Never before (and I'm sorry to say never since) has physics offered so simple and satisfying an answer to the question, "What is matter made of?" In 1932 it was all just protons, neutrons, and electrons. But already the seeds were planted for the three great ideas that were to dominate the *middle period* (1930–1960) in particle physics: Yukawa's meson, Dirac's positron, and Pauli's neutrino. Before we come to that, however, I must back up for a moment to introduce the photon.

1.2 THE PHOTON (1900–1924)

In some respects the photon is a very "modern" particle, having more in common with the W and Z (which were not discovered until 1983) than with the classical trio. Moreover, it's hard to say exactly when or by whom the photon was really "discovered," although the essential stages in the process are clear enough. The first contribution was made by Planck in 1900. Planck was attempting to explain the so-called *blackbody spectrum* for the electromagnetic radiation emitted by a hot object. Statistical mechanics, which had proved brilliantly successful in explaining other thermal processes, yielded nonsensical results when applied to electromagnetic fields. In particular, it led to the famous "ultraviolet catastrophe," predicting that the total power radiated should be *infinite*. Planck found that he could escape the ultraviolet catastrophe—and fit the experimental curve—if he assumed that electromagnetic radiation is *quantized*, coming in little "packages" of energy

$$E = h\nu \tag{1.1}$$

where ν is the frequency of the radiation and h is a constant, which Planck adjusted to fit the data. The modern value of Planck's constant is

$$h = 6.626 \times 10^{-27} \text{ erg s} \tag{1.2}$$

Planck did not profess to know *why* the radiation was quantized; he assumed that it was due to a peculiarity in the emission process: For some reason a hot surface only gives off light* in little squirts.

Einstein, in 1905, put forward a far more radical view. He argued that quantization was a feature of the electromagnetic field itself, having nothing to

* In this book the word *light* stands for *electromagnetic radiation*, whether or not it happens to fall in the visible region.

do with the emission mechanism. With this new twist, Einstein adapted Planck's idea, and his formula, to explain the *photoelectric effect:* When electromagnetic radiation strikes a metal surface, electrons come popping out. Einstein suggested that an incoming light quantum hits an electron in the metal, giving up its energy ($h\nu$); the excited electron then breaks through the metal surface, losing in the process an energy w (the so-called *work function* of the material—an empirical constant that depends on the particular metal involved). The electron thus emerges with an energy

$$E \leq h\nu - w \tag{1.3}$$

(It may lose some energy before reaching the surface. That's the reason for using $\leq$, instead of $=$.) Einstein's formula (1.3) is pretty trivial to *derive,* but it carries an extraordinary implication: The maximum electron energy is *independent of the intensity of the light* and depends only on its *color* (frequency). To be sure, a more intense beam will knock out *more* electrons, but their *energies* will be the same.

Unlike Planck's theory, Einstein's theory met a hostile reception, and over the next 20 years he was to wage a lonely battle for the *light quantum.*[2] In saying that electromagnetic radiation is *by its nature* quantized, regardless of the emission mechanism, Einstein came dangerously close to resurrecting the discredited particle theory of light. Newton, of course, had introduced such a *corpuscular* model, but a major achievement of nineteenth-century physics was the decisive repudiation of Newton's idea in favor of the rival wave theory. No one was prepared to see that accomplishment called into question, even when the experiments came down on Einstein's side. In 1916 Millikan completed an exhaustive study of the photoelectric effect and was obliged to report that "Einstein's photoelectric equation . . . appears in every case to predict exactly the observed results. . . . Yet the semicorpuscular theory by which Einstein arrived at his equation seems at present wholly untenable."[3]

What finally settled the issue was an experiment conducted by A. H. Compton in 1923. Compton found that the light scattered from a particle at rest is shifted in wavelength, according the equation

$$\lambda' = \lambda + \lambda_c(1 - \cos\theta) \tag{1.4}$$

where λ is the incident wavelength, λ' is the scattered wavelength, θ is the scattering angle, and

$$\lambda_c = h/mc \tag{1.5}$$

is the so-called *Compton wavelength* of the target particle (mass m). Now, this is *precisely* the formula you get (Problem 3.24) if you treat light as a particle of zero rest mass with energy given by Planck's equation, and apply the laws of conservation of (relativistic) energy and momentum—just as you would for an ordinary elastic collision (Fig. 1.3). That clinched it; here was direct and incontrovertible experimental evidence that light behaves as a particle, on the subatomic scale. We call this particle the *photon* (a name suggested by the chemist Gilbert Lewis, in 1926); the symbol for a photon is γ (from *gamma ray*). How

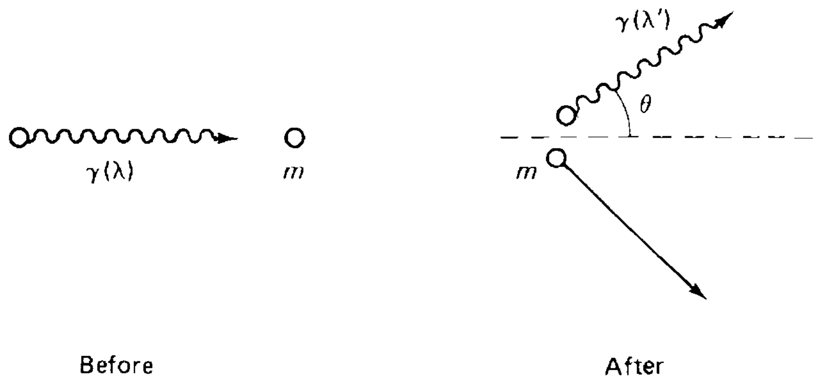


Figure 1.3 Compton scattering. A photon of wavelength λ scatters off a particle, initially at rest, of mass m. The scattered photon carries wavelength λ′ given by equation (1.4).

the particle nature of light on this level is to be reconciled with its well-established wave behavior on the macroscopic scale (exhibited in the phenomena of interference and diffraction) is a story I'll leave to the quantum texts.

Although the photon initially *forced* itself on an unreceptive community of physicists, it eventually found a natural place in quantum field theory, and was to offer a whole new perspective on electromagnetic interactions. In classical electrodynamics, we attribute the electrical repulsion of two electrons, say, to the electric field surrounding them; each electron contributes to the field, and each one responds to the field. But in quantum field theory, the electric field is *quantized* (in the form of photons), and we may picture the interaction as consisting of a stream of photons passing back and forth between the two charges, each electron continually emitting them and continually absorbing them. And the same goes for *any* noncontact force: where classically we interpret "action at a distance" as "mediated" by a *field,* we now say that it is mediated by an *exchange of particles* (the *quanta* of the field). In the case of electrodynamics, the mediator is the photon; for gravity, it is called the *graviton* (though a fully successful quantum theory of gravity has yet to be developed and it may well be centuries before anyone detects a graviton experimentally).

You will see later on how these ideas are implemented in practice, but for now I want to dispel one common misapprehension. When I say that every force is mediated by the exchange of particles, I am *not* speaking of a merely *kinematic* phenomenon. Two ice skaters throwing snowballs back and forth will of course move apart with the succession of recoils; they "repel one another by exchange of snowballs," if you like. But that's *not* what is involved here. For one thing, this mechanism would have a hard time accounting for an *attractive* force. You might think of the mediating particles, rather, as "messengers," and the message can just as well be "come a little closer" as "go away."

I said earlier that in the "classical" picture ordinary matter is made of atoms, in which electrons are held in orbit around a nucleus of protons and neutrons by the electrical attraction of opposite charges. We can now give this model a more sophisticated formulation by attributing the binding force to the exchange of photons between the electrons and the protons in the nucleus. However, for the purposes of atomic physics this is overkill, for in this context quantization of the electromagnetic field produces only minute effects (notably the

Lamb shift and the anomalous magnetic moment of the electron). To excellent approximation we can pretend that the forces are given by Coulomb's law (together with various magnetic dipole couplings). The point is that in a bound state enormous numbers of photons are continually streaming back and forth, so that the "lumpiness" of the field is effectively smoothed out, and classical electrodynamics is a suitable approximation to the truth. But in *most* elementary particle processes, such as the photoelectric effect or Compton scattering, *individual* photons are involved, and quantization can no longer be ignored.

1.3 MESONS (1934–1947)

Now there is one conspicuous problem to which the "classical" model does not address itself at all: What holds the *nucleus* together? After all, the positively charged protons should repel one another violently, packed together as they are in such close proximity. Evidently there must be some other force, more powerful than the force of electrical repulsion, that binds the protons (and neutrons) together; physicists of that less imaginative age called it, simply, the *strong force.* But if there exists such a potent force in nature, why don't we notice it in everyday life? The *fact* is that virtually every force we experience directly, from the contraction of a muscle to the explosion of dynamite is electromagnetic in origin; the only exception, outside a nuclear reactor or an atomic bomb, is gravity. The answer must be that, powerful though it is, the strong force is of very short *range.* (The range of a force is like the arm's reach of a boxer—beyond that distance its influence falls off rapidly to zero. Gravitational and electromagnetic forces have *infinite* range, but the range of the strong force is about the size of the nucleus itself.)*

The first significant theory of the strong force was proposed by Yukawa in 1934. Yukawa assumed that the proton and neutron are attracted to one another by some sort of *field,* just as the electron is attracted to the nucleus by an electric field and the moon to the earth by a gravitational field. This field should properly be quantized, and Yukawa asked the question: What must be the properties of its *quantum*—the particle (analogous to the photon) whose exchange would account for the known features of the strong force? For example, the short range of the force indicated that the mediator would be rather heavy; Yukawa calculated that its mass should be nearly 300 times that of the electron, or about a sixth the mass of a proton. (See Problem 1.2.) Because it fell between the electron and the proton, Yukawa's particle came to be known as the *meson* (meaning "middle-weight"). [In the same spirit the electron is called a *lepton* ("light-weight"), whereas the proton and neutron are *baryons* ("heavy-weight").] Now, Yukawa knew that no such particle had ever been observed in the laboratory, and he therefore assumed his theory was wrong. But at the time a number of systematic studies

* This is a bit of an oversimplification. Typically, the forces go like $e^{-(r/a)}/r^2$, where a is the "range." For Coulomb's law and Newton's law of universal gravitation, $a = \infty$; for the strong force a is about 10^{-13} cm (one fermi).

of cosmic rays were in progress, and by 1937 two separate groups (Anderson and Neddermeyer on the West Coast, and Street and Stevenson on the East) had identified particles matching Yukawa's description. Indeed, the cosmic rays with which you are being bombarded every few seconds as you read this consist primarily of just such middle-weight particles.

For a while everything seemed to be in order. But as more detailed studies of the cosmic ray particles were undertaken, disturbing discrepancies began to appear. They had the wrong lifetime and they seemed to be significantly lighter than Yukawa had predicted; worse still, different mass measurements were not consistent with one another. In 1946 (after a period in which physicists were engaged in a less savory business) decisive experiments were carried out in Rome demonstrating that the cosmic ray particles interacted very weakly with atomic nuclei.[4] If this was really Yukawa's meson, the transmitter of the strong force, the interaction should have been dramatic. The puzzle was finally resolved in 1947, when Powell and his co-workers at Bristol[5] discovered that there are actually *two* middle-weight particles in cosmic rays, which they called π (or "pion") and μ (or "muon"). (Marshak reached the same conclusion simultaneously, on theoretical grounds.[6]) The true Yukawa meson is the π; it is produced copiously in the upper atmosphere, but ordinarily disintegrates long before reaching the ground. (See Problem 3.4.) Powell's group exposed their photographic emulsions on mountain tops (see Fig. 1.4). One of the decay products is the lighter (and longer-lived) μ, and it is primarily muons that one observes at sea level. In the search for Yukawa's meson, then, the muon was simply an imposter, having nothing whatever to do with the strong interactions. In fact, it behaves in every way like a heavier version of the electron and properly belongs in the *lepton* family (though some people to this day call it the "mu-meson" by force of habit).

1.4 ANTIPARTICLES (1930–1956)

Nonrelativistic quantum mechanics was completed in the astonishingly brief period 1923–1926, but the relativistic version proved to be a much thornier problem. The first major achievement was Dirac's discovery, in 1927, of the equation that bears his name. The Dirac equation was supposed to describe free electrons with energy given by the relativistic formula $E^2 - \mathbf{p}^2c^2 = m^2c^4$. But it had a very troubling feature: For every positive-energy solution ($E = +\sqrt{\mathbf{p}^2c^2 + m^2c^4}$) it admitted a corresponding solution with *negative* energy ($E = -\sqrt{\mathbf{p}^2c^2 + m^2c^4}$). This meant, given the natural tendency of every system to evolve in the direction of lower energy, that the electron should "runaway" to increasingly negative states, radiating off an infinite amount of energy in the process. To rescue his equation, Dirac proposed a resolution that made up in brilliance for what it lacked in plausibility: He postulated that the negative energy states are all filled by an infinite "sea" of electrons. Because this sea is always there, and perfectly uniform, it exerts no net force on anything, and we are not normally aware of it. Dirac then invoked the Pauli exclusion principle (which says that no two electrons can occupy the same state), to "explain" why the

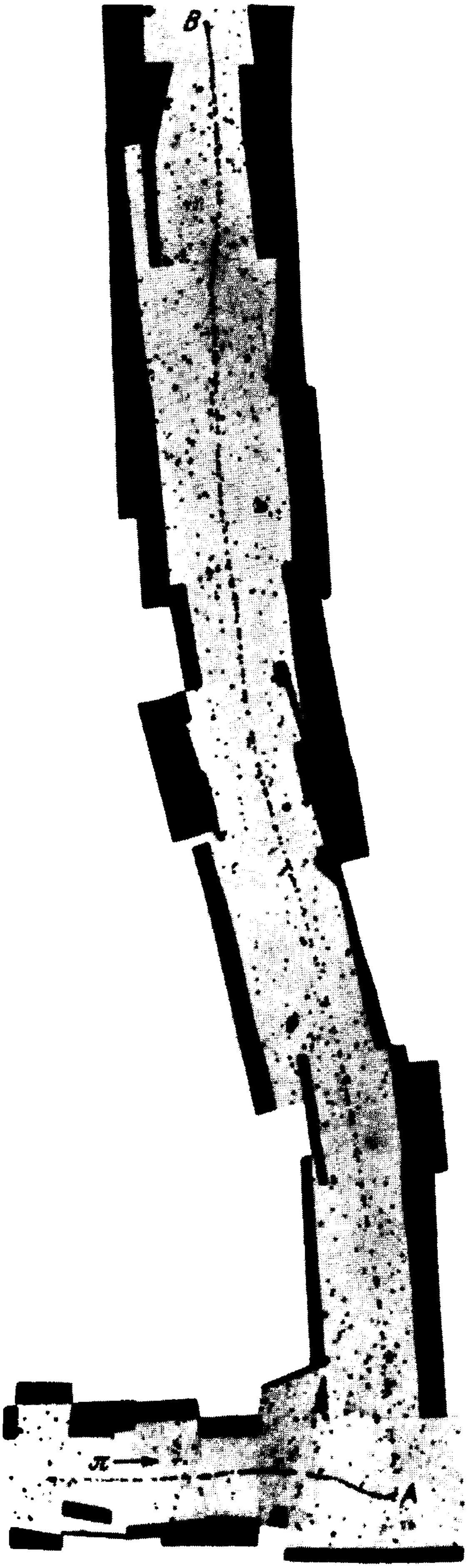


Figure 1.4 One of Powell's earliest pictures showing the track of a pion in a photographic emulsion exposed to cosmic rays at high altitude. The pion (entering from the left) decays into a muon and a neutrino (the latter is electrically neutral, and leaves no track). Reprinted by permission from C. F. Powell, P. H. Fowler, and D. H. Perkins, *The Study of Elementary Particles by the Photographic Method* (New York: Pergamon, 1959). First published in *Nature* **159**, 694 (1947).

electrons we *do* observe are confined to the positive energy states. But if this is true, then what happens when we impart to one of the electrons in the "sea" an energy sufficient to knock it into a positive energy state? The *absence* of the "expected" electron in the sea would be interpreted as a net positive charge in that location, and the absence of its expected negative energy would be seen as a net positive energy. Thus a "hole in the sea" would function as an ordinary particle with *positive* energy and *positive* charge. Dirac at first hoped that these holes might be *protons,* but it was soon apparent that they had to carry the same mass as the electron itself—2000 times too light to be a proton. No such particle was known at the time, and Dirac's theory appeared to be in trouble. What may have seemed a fatal defect in 1930, however, turned into a spectacular triumph in late 1931, with Anderson's discovery of the *positron* (Fig. 1.5), a positively-charged twin for the electron, with precisely the attributes Dirac required.[7]

Figure 1.5 The positron. In 1932, Anderson took this photograph of the track left in a cloud chamber by a cosmic ray particle. The chamber was placed in a magnetic field (pointing into the page) which caused the particle to travel in a curve. But was it a negative charge traveling downward, or a positive charge traveling upward? In order to tell, Anderson had placed a lead plate across the center of the chamber (the thick horizontal line in the photograph). A particle passing through the plate slows down, and subsequently moves in a tighter circle. By inspection of the curves, it is clear that this particle traveled upward, and hence must have been positively charged. From the curvature of the track, and from its texture, Anderson was able to show that the mass of the particle was close to that of the electron. (Photo courtesy California Institute of Technology)

Still, many physicists were uncomfortable with the notion that we are awash in an infinite sea of invisible electrons, and in the forties Stuckelberg and Feynman provided a much simpler and more compelling interpretation of the negative-energy states. In the Feynman–Stuckelberg formulation the negative-energy solutions are reexpressed as *positive*-energy states of a *different particle* (the positron); the electron and positron appear on an equal footing, and there is no need for Dirac's "electron sea" or for its mysterious "holes." We'll see in Chapter 7 how this—the modern interpretation—works. Meantime, it turned out that the dualism in Dirac's equation is a profound and universal feature of quantum field theory: For *every* kind of particle there must exist a corresponding *antiparticle,* with the same mass but opposite electric charge. The positron, then, is the *antielectron.* (Actually, it is in principle completely arbitrary which one you call the "particle" and which the "antiparticle"—I could just as well have said that the electron is the antipositron. But since there are a lot of electrons around, and not so many positrons, we tend to think of electrons as "matter" and positrons as "antimatter"). The (negatively charged) antiproton was first observed experimentally at the Berkeley Bevatron in 1955, and the (neutral) antineutron was discovered at the same facility the following year.[8]

The standard notation for antiparticles is an overbar. For example, p denotes the proton and $\bar{p}$ the antiproton; n the neutron and $\bar{n}$ the antineutron. However, in some cases it is more customary simply to specify the charge. Thus most people write e^+ for the positron (not $\bar{e}$) and μ^+ for the antimuon (not $\bar{\mu}$). [But you must not *mix* conventions: $\bar{e}^+$ is ambiguous, like a double negative—the reader doesn't know if you mean the positron or the *anti*positron, (which is to say, the electron).] Some neutral particles are their *own* antiparticles. For example, the photon: $\bar{\gamma} = \gamma$. In fact, you may have been wondering how the antineutron differs physically from the neutron, since both are uncharged. The answer is that neutrons carry other "quantum numbers" besides charge (in particular, baryon number), which change sign for the antiparticle. Moreover, although its *net* charge is zero, the neutron *does* have a charge *structure* (positive at the center and at the edges, negative in between) and a magnetic dipole moment. These, too, have the opposite sign for $\bar{n}$.

There is a general principle in particle physics that goes under the name of *crossing symmetry.* Suppose that a reaction of the form

$$A + B \rightarrow C + D$$

is known to occur. Any of these particles can be "crossed" over to the other side of the equation, provided it is turned into its antiparticle, and the resulting interaction will also be allowed. For example,

$$\begin{aligned} A &\rightarrow \bar{B} + C + D \\ A + \bar{C} &\rightarrow \bar{B} + D \\ \bar{C} + \bar{D} &\rightarrow \bar{A} + \bar{B} \end{aligned}$$

In addition, the *reverse* reaction occurs $C + D \rightarrow A + B$, but technically this derives from the principle of *detailed balance,* rather than from crossing symmetry. Indeed, as we shall see, the *calculations* involved in these various reactions are practically identical. We might almost regard them as different manifestations of the same fundamental process. Now, there is one important *caveat* in this: Conservation of energy may veto a reaction that is otherwise permissible. For example, if A weighs less than the sum of B, C, and D, then the decay $A \rightarrow \bar{B} + C + D$ cannot occur; similarly, if A and C are light, whereas B and D are heavy, then the reaction $A + \bar{C} \rightarrow \bar{B} + D$ will not take place unless the initial kinetic energy exceeds a certain "threshold" value. So perhaps I should say that the crossed (or reversed) reaction is *dynamically* permissible, but it may or may not be *kinematically* allowed. The power and beauty of crossing symmetry can scarcely be exaggerated. It tells us, for instance, that Compton scattering

$$\gamma + e^- \rightarrow \gamma + e^-$$

is "really" the same process as pair annihilation

$$e^- + e^+ \rightarrow \gamma + \gamma$$

although in the laboratory they are completely different phenomena.

The union of special relativity and quantum mechanics, then, leads to a pleasing matter/antimatter symmetry. But this raises a disturbing question: How come *our* world is populated with protons, neutrons, and electrons, instead of antiprotons, antineutrons, and positrons? Matter and antimatter cannot coexist for long—if a particle meets its antiparticle, they annihilate. So maybe it's just a historical accident that in our corner of the universe there happened to be more matter than antimatter, and pair annihilation has eliminated all but a leftover residue of matter. If this is so, then presumably there are other regions of space in which antimatter predominates. Unfortunately, the astronomical evidence is pretty compelling that all of the observable universe is made of ordinary matter. Recently, Wilczek and others have put forward a possible explanation for this cosmic asymmetry. I shall not go into it here, but if you are interested, I recommend Wilczek's article in *Scientific American* (December 1980).

1.5 NEUTRINOS (1930–1962)

For the third strand in the story we return again to the year 1930.[9] A problem had arisen in the study of nuclear beta decay. In beta decay a radioactive nucleus A is transformed into a slightly lighter nucleus B, with the emission of an electron:

$$A \rightarrow B + e^- \tag{1.6}$$

Conservation of charge requires that B carry one more unit of positive charge than A. [We now realize that the underlying process here is the conversion of a neutron (in A) into a proton (in B), but remember that in 1930 the neutron had not yet been discovered.] Thus the "daughter" nucleus (B) lies one position farther along on the Periodic Table. There are many examples of beta decay: Potassium goes to calcium (${}^{40}_{19}K \rightarrow {}^{40}_{20}Ca$), copper goes to zinc (${}^{64}_{29}Cu \rightarrow {}^{64}_{30}Zn$), tritium goes to helium (${}^{3}_{1}H \rightarrow {}^{3}_{2}He$), and so on. [The upper number is the *atomic*

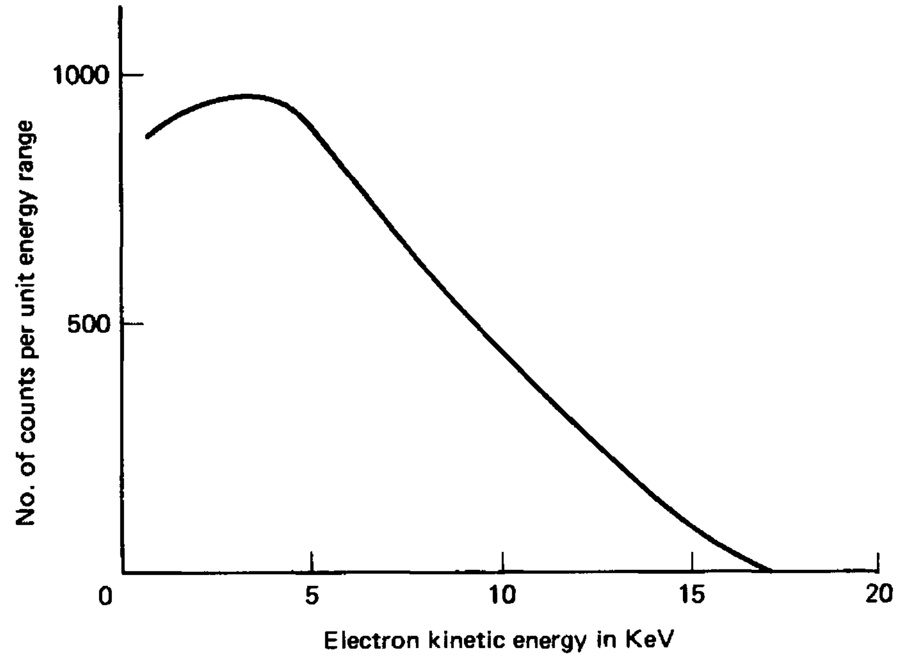


Figure 1.6 The beta decay spectrum of tritium ($^3_1H \rightarrow {}^3_2He$). (*Source:* G. M. Lewis, *Neutrinos* (London: Wykeham, 1970), p. 30.)

weight (the number of neutrons plus protons) and the lower number is the *atomic number* (the number of protons).]

Now, it is a characteristic of two-body decays such as expression (1.6) that the outgoing energies are kinematically determined, in the center-of-mass frame. Specifically, if the "parent" nucleus (A) is at rest, so that B and e come out back-to-back with equal and opposite momenta, then conservation of energy dictates that the electron energy is

$$E = \left(\frac{m_A^2 - m_B^2 + m_e^2}{2m_A}\right)c^2 \tag{1.7}$$

The derivation of this result will be explained in Chapter 3; for now, the point to notice is that E is *fixed,* once the three masses are specified. But when the experiments are done it is found that the emitted electrons vary considerably in energy. Equation (1.7) only determines the *maximum* electron energy, for a particular beta-decay process (see Fig. 1.6).

This was a most disturbing result. Niels Bohr (not for the first time) was ready to abandon the law of conservation of energy.* Fortunately, Pauli took a more sober view, suggesting that another particle was emitted along with the electron, a silent accomplice that carries off the "missing" energy. It had to be electrically neutral, to conserve charge (and also, of course, to explain why it left no track); Pauli proposed to call it the *neutron.* The whole idea was greeted with some skepticism, and in 1932 Chadwick preempted the name. But in the following year Fermi presented a theory of beta decay that incorporated Pauli's

* It is interesting to note that Bohr was an outspoken critic of Einstein's light quantum (prior to 1924), that he discouraged Dirac's work on the relativistic electron theory (telling him, incorrectly, that Klein and Gordon had already succeeded), that he opposed Pauli's introduction of the neutrino, that he ridiculed Yukawa's theory of the meson, and that he disparaged Feynman's approach to quantum electrodynamics.

particle and proved so brilliantly successful that Pauli's suggestion had to be taken seriously. From the fact that the observed electron energies range up to the value given in equation (1.7) it follows that the new particle is extremely light; as far as we know, its mass is in fact *zero*. Fermi called it the *neutrino*. (For reasons you'll see in a moment, we now call it the *anti*neutrino.) In modern terminology, then, the fundamental beta-decay process is

$$n \longrightarrow p^+ + e^- + \bar{\nu} \tag{1.8}$$

(neutron goes to proton plus electron plus antineutrino).

Now, you may have noticed something peculiar about Powell's picture of the disintegrating pion (Fig. 1.4): The muon emerges at about 90° with respect to the original pion direction. (That's not the result of a *collision*, by the way; collisions with atoms in the emulsion account for the dither in the tracks, but they cannot produce an abrupt left turn.) What that kink indicates is that some *other* particle was produced in the decay of the pion, a particle that left no footprints in the emulsion, and hence must have been electrically neutral. It was natural (or at any rate *economical*) to suppose that this was again Pauli's neutrino:

$$\pi \longrightarrow \mu + \nu \tag{1.9}$$

A few months after their first paper, Powell's group published an even more striking picture, in which the subsequent decay of the muon is also visible (Fig. 1.7). Now, muon decays had been studied for many years, and it was well established that the charged secondary is an electron. From the figure there is clearly a neutral product as well, and you might guess that it is again a neutrino. However, this time it is *two* neutrinos:

$$\mu \longrightarrow e + 2\nu \tag{1.10}$$

How do we know there are *two* of them? Same way as before: We repeat the experiment over and over, each time measuring the energy of the electron. If it always comes out the same, we know there are just two particles in the final state. But if it *varies*, then there must be (at least) three. By 1949* it was clear that the electron energy in muon decay is *not* fixed, and the emission of two neutrinos was the accepted explanation. By contrast, the *muon* energy in *pion* decay is perfectly constant, within experimental uncertainties, confirming that this is a genuine two-body decay.

By 1950, then, there was compelling *theoretical* evidence for the existence of neutrinos, but there was still no direct *experimental* verification. A skeptic might have argued that the neutrino was nothing but a bookkeeping device—a purely hypothetical particle whose only function was to rescue the conservation laws. It left no tracks, it didn't decay; in fact, no one had ever seen a neutrino *do anything*. The reason for this is that neutrinos interact extraordinarily weakly

* Here, and in the original beta-decay problem, conservation of angular momentum also requires a third outgoing particle, quite independently of energy conservation. But the spin assignments were not so clear in the early days, and for most people energy conservation was the compelling argument. In the interests of simplicity, I will keep angular momentum out of the story until Chapter 4.

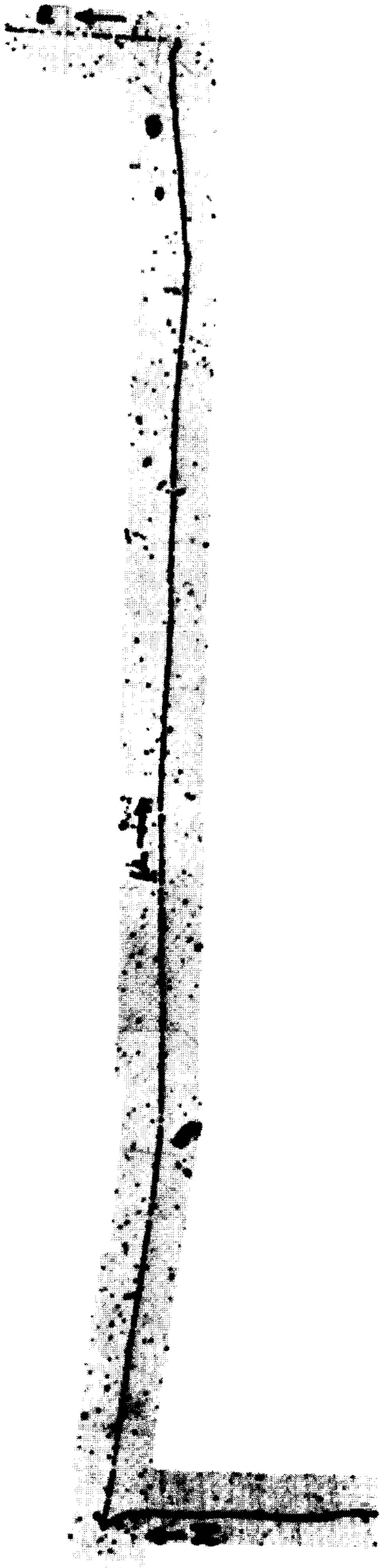


Figure 1.7 Here, a pion decays into a muon (plus a neutrino); the muon subsequently decays into an electron (and two neutrinos). Reprinted by permission from C. F. Powell, P. H. Fowler, and D. H. Perkins, *The Study of Elementary Particles by the Photographic Method* (New York: Pergamon, 1959). First published in *Nature* **163**, 82 (1949).

with matter; a neutrino of moderate energy could easily penetrate a thousand light-years(!) of lead.* To have a chance of detecting one you need an extremely intense source. The decisive experiments were conducted at the Savannah River nuclear reactor in South Carolina, in the mid-fifties. Here Cowan and Reines set up a large tank of water and watched for the "inverse" beta-decay reaction

$$\bar{\nu} + p^+ \rightarrow n + e^+ \tag{1.11}$$

At their detector the antineutrino flux was calculated to be 5×10^{13} particles per square centimeter per second, but even at this fantastic intensity they could only hope for two or three events every hour. On the other hand, they developed an ingenious method for identifying the outgoing positron. Their results provided unambiguous confirmation of the neutrino's existence.[10]

As I mentioned earlier, the particle produced in ordinary beta decay is actually an antineutrino, not a neutrino. Of course, since they're electrically neutral, you might ask—and many people *did*—whether there is any distinction between a neutrino and an antineutrino. The neutral pion, as we shall see, is its *own* antiparticle; so too is the photon. On the hand, the antineutron is definitely *not* the same as a neutron. So we're left in a bit of a quandary: *Is* the neutrino the same as the antineutrino, and if not, what property distinguishes them? In the late fifties, Davis and Harmer put this question to an experimental test.[11] From the positive results of Cowan and Reines, we know that the crossed reaction

$$\nu + n \rightarrow p^+ + e^- \tag{1.12}$$

must also occur, and at about the same rate. Davis looked for the analogous reaction using *anti*neutrinos:

$$\bar{\nu} + n \rightarrow p^+ + e^- \tag{1.13}$$

He found that this reaction does *not* occur, and thus established that the neutrino and antineutrino are distinct particles.

Davis's result was not unexpected. In fact, back in 1953 Konopinski and Mahmoud[12] had introduced a beautifully simple rule for determining which reactions [such as (1.12)] will work, and which [like (1.13)] will not. In effect,† they assigned a *lepton number* $L = +1$ to the electron, the muon, and the neutrino, and $L = -1$ to the positron, the positive muon, and the antineutrino (all other particles are given a lepton number of zero). They then proposed the law of conservation of lepton number (analogous to the law of conservation of charge): In any physical process, the sum of the lepton numbers before must equal the sum of the lepton numbers after. Thus the Cowan–Reines reaction (1.11) is allowed ($L = -1$ before and after), but the Davis reaction (1.13) is forbidden (on the left $L = -1$, on the right $L = +1$). [It was in anticipation of this rule that I called the beta-decay particle, in expression (1.8), an *anti*neutrino.] In

* That's a comforting realization when you learn that hundreds of billions of neutrinos pass through every square inch of your body per second, night and day, coming from the sun (they hit you from below, at night, having passed right through the earth).

† Konopinski and Mahmoud (ref. 12) did not use this terminology, and they got the muon assignments wrong. But never mind, the essential idea was there.

view of the conservation of lepton number, the charged pion decays (1.9) should actually be written

$$\begin{aligned} \pi^- &\rightarrow \mu^- + \bar{\nu} \\ \pi^+ &\rightarrow \mu^+ + \nu \end{aligned} \tag{1.14}$$

and the muon decays (1.10) are really

$$\begin{aligned} \mu^- &\rightarrow e^- + \nu + \bar{\nu} \\ \mu^+ &\rightarrow e^+ + \nu + \bar{\nu} \end{aligned} \tag{1.15}$$

What property distinguishes the neutrino from the antineutrino, then? The cleanest answer is: *lepton number*—it's $+1$ for the neutrino and -1 for the antineutrino. These numbers are experimentally determinable, just as electric charge is, by watching how the particle in question interacts with others. (As we shall see, they also differ in their *helicity:* the neutrino is "left-handed" whereas the antineutrino is "right-handed." But this is a technical matter best saved for later.)

There is a final twist to the neutrino story. Experimentally, the decay of a muon into an electron plus a *photon* is never observed:

$$\mu^- \not\rightarrow e^- + \gamma \tag{1.16}$$

and yet this process is consistent with conservation of charge and conservation of the lepton number. Now, there's a very reliable rule of thumb in particle physics (generally attributed to Richard Feynman) which says that whatever is not expressly *forbidden* is *mandatory.* The absence of $\mu \rightarrow e + \gamma$ suggests a law of conservation of "mu-ness"; but then how are we to explain the observed decays $\mu \rightarrow e + \nu + \bar{\nu}$? The answer occurred to a number of people in the late fifties and early sixties:[13] Suppose there are two different *kinds* of neutrino—one associated with the electron (ν_e) and one with the muon (ν_μ). If we assign a *muon number* $L_\mu = +1$ to μ^- and ν_μ, and $L_\mu = -1$ to μ^+ and $\bar{\nu}_\mu$, and at the same time an *electron number* $L_e = +1$ to e^- and ν_e, and $L_e = -1$ to e^+ and $\bar{\nu}_e$, and refine the conservation of lepton number into two separate laws—conservation of electron number and conservation of muon number—we can then account for *all* the allowed and forbidden processes. Neutron beta decay becomes

$$n \rightarrow p^+ + e^- + \bar{\nu}_e \tag{1.17}$$

the pion decays are

$$\begin{aligned} \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu \\ \pi^+ &\rightarrow \mu^+ + \nu_\mu \end{aligned} \tag{1.18}$$

and the muon decays take the form

$$\begin{aligned} \mu^- &\rightarrow e^- + \bar{\nu}_e + \nu_\mu \\ \mu^+ &\rightarrow e^+ + \nu_e + \bar{\nu}_\mu \end{aligned} \tag{1.19}$$

I said earlier that when pion decay was first analyzed it was "natural" and "economical" to assume that the outgoing neutral particle was the same as in beta decay, and that's quite true: It *was* natural, and it *was* economical, but it was *wrong*.

The first experimental test of the two-neutrino hypothesis (and the separate conservation of electron and muon number) was conducted at Brookhaven in 1962.[14] Using about 10^{14} antineutrinos from π^- decay, Lederman, Schwartz, Steinberger, and their collaborators identified 29 instances of the expected reaction

$$\bar{\nu}_\mu + p^+ \rightarrow \mu^+ + n \tag{1.20}$$

and no cases of the forbidden process

$$\bar{\nu}_\mu + p^+ \rightarrow e^+ + n \tag{1.21}$$

With only one kind of neutrino the second reaction would be just as common as the first. (Incidentally, this experiment presented truly monumental shielding problems. Steel from a dismantled warship was stacked up 44 feet thick, to make sure that nothing except neutrinos got through to the target.)

By 1962, then, the lepton family had grown to eight: the electron, the muon, their respective neutrinos, and the corresponding antiparticles (Table 1.1). The leptons are characterized by the fact that they do not participate in strong interactions. For the next 14 years things were pretty quiet, as far as the leptons go, so this is a good place to pause and let the strongly interacting particles—the mesons and baryons, known collectively as the *hadrons*—catch up.

1.6 STRANGE PARTICLES (1947–1960)

For a brief period in 1947 it was possible to believe that the major problems of elementary particle physics were solved. After a lengthy detour in pursuit of the muon, Yukawa's meson (the π) had finally been apprehended. Dirac's positron had been found, and Pauli's neutrino, although still at large (and, as we have

TABLE 1.1 THE LEPTON FAMILY, 1962–1976

	Lepton number	Electron number	Muon number
Leptons			
e^-	1	1	0
ν_e	1	1	0
μ^-	1	0	1
ν_μ	1	0	1
Antileptons			
e^+	−1	−1	0
$\bar{\nu}_e$	−1	−1	0
μ^+	−1	0	−1
$\bar{\nu}_\mu$	−1	0	−1

seen, still capable of making mischief), was basically under control. The role of the muon was something of a puzzle ("Who ordered *that?*" Rabi asked); it seemed quite unnecessary in the overall scheme of things. On the whole, however, it looked in 1947 as though the job of elementary particle physics was essentially done.

But this comfortable state did not last long. In December of that year Rochester and Butler[15] published the cloud chamber photograph shown in Figure 1.8. Cosmic ray particles enter from the upper left and strike a lead plate, producing a neutral particle, whose presence is revealed when it decays into two charged secondaries, forming the upside-down "V" in the lower right. Detailed analysis shows that these charged particles are in fact a π^+ and a π^-. Here, then, was a new neutral particle with at least twice the mass of the pion; we call it the K^0 ("kaon"):

$$K^0 \rightarrow \pi^+ + \pi^- \tag{1.22}$$

In 1949, Powell published the photograph reproduced in Figure 1.9, showing the decay of a charged kaon:

$$K^+ \rightarrow \pi^+ + \pi^+ + \pi^- \tag{1.23}$$

(The K^0 was first known as the V^0 and later as the θ^0; the K^+ was originally called the τ^+. Their identification as neutral and charged versions of the same basic particle was not completely settled until 1956—but that's another story, to which we shall return in Chapter 4.) The kaons behave in some respects like heavy pions, and so the *meson* family was extended to include them. In due course, many more mesons were discovered—the η, the ϕ, the ω, the ρ's, and so on.

Meanwhile, in 1950 another neutral "V" particle was found, this time by Anderson's group at Cal Tech. The photographs were similar to Rochester's (Fig. 1.8), but this time the products were a p^+ and a π^-. Evidently this particle is substantially heavier than the proton; we call it the Λ:

$$\Lambda \rightarrow p^+ + \pi^- \tag{1.24}$$

The lambda belongs with the proton and the neutron in the *baryon* family. To appreciate this, we must go back for a moment to 1938. The question had arisen, "Why is the proton stable?" Why, for example, doesn't it decay into a positron and a photon:

$$p^+ \rightarrow e^+ + \gamma \tag{1.25}$$

Needless to say, it would be unpleasant for us if this reaction were common (all atoms would disintegrate), and yet it does not violate any law known in 1938. (Actually, this particular process does violate conservation of lepton number, but that law was not recognized, remember, until 1953.) Stückelberg[16] proposed to account for the stability of the proton by asserting a law of conservation of baryon number: Assign to all baryons (which in 1938 meant the proton and the neutron) a "baryon number" $A = +1$, and to the antibaryons ($\bar{p}$ and $\bar{n}$)

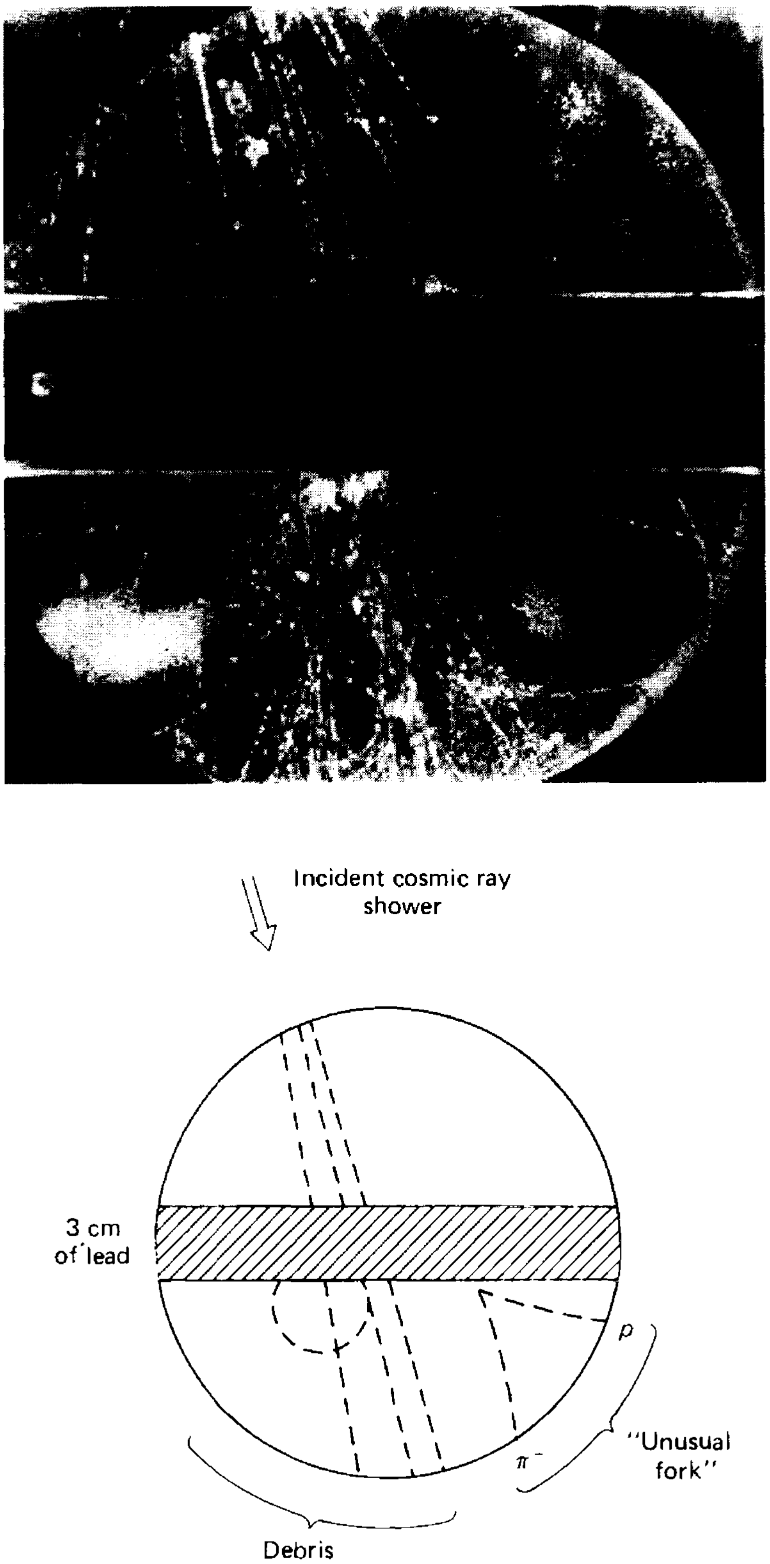


Figure 1.8 The first strange particle. Cosmic rays strike a lead plate, producing a K^0, which subsequently decays into a pair of charged pions. (Photo courtesy of Prof. G. D. Rochester. Reprinted by permission from *Nature* **160**, 855. Copyright © 1947, Macmillan Journals Limited.)

$A = -1$; then the total baryon number is conserved in any physical process. Thus, neutron beta decay ($n \rightarrow p^+ + e^- + \bar{\nu}_e$) is allowed ($A = 1$ before and after), and so also is the reaction in which the antiproton was first observed:

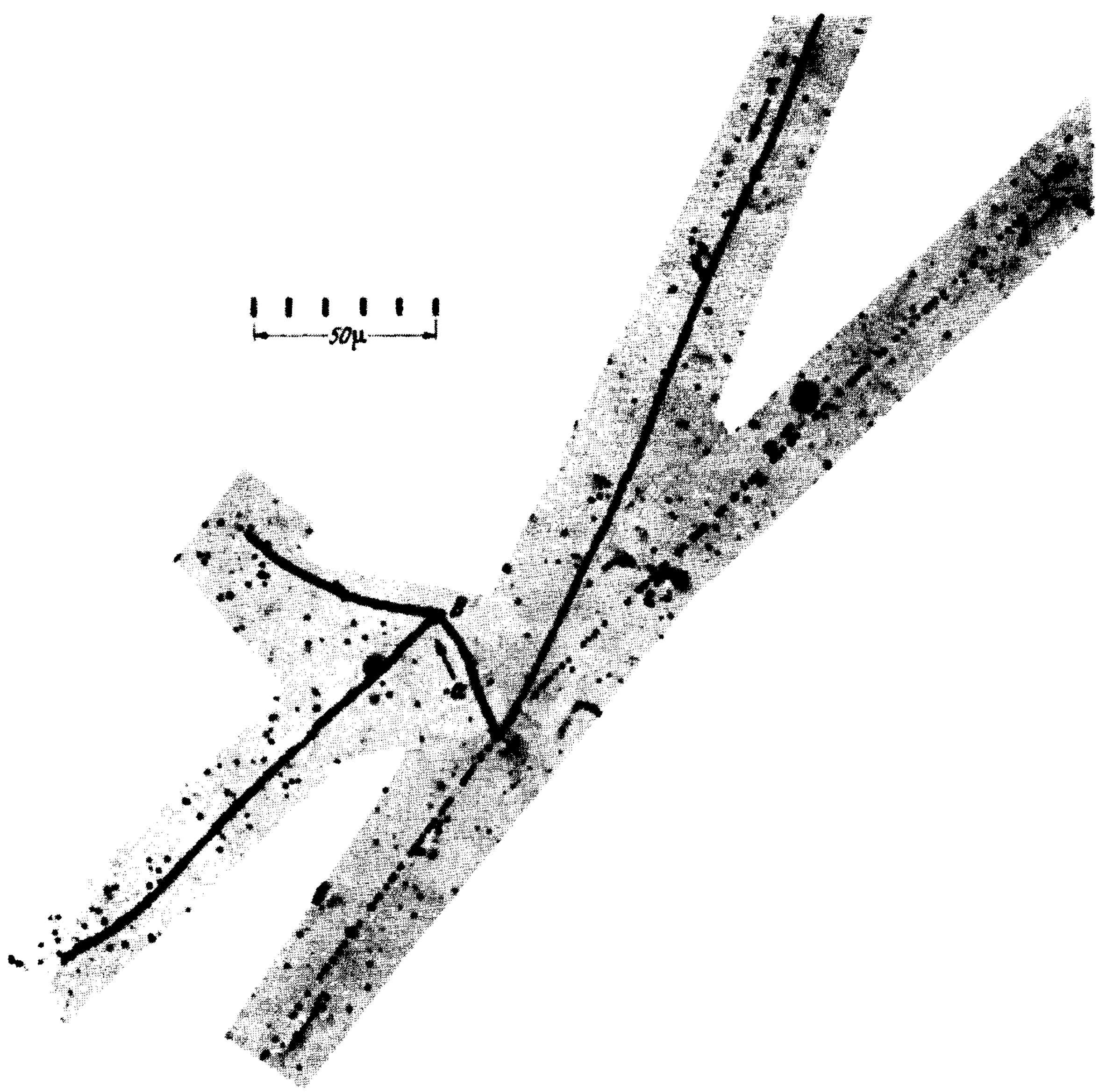


Figure 1.9 K^+, entering from above, decays at A: $K^+ \to \pi^+ + \pi^+ + \pi^-$. (The π^- subsequently causes a nuclear disintegration at B). [Reprinted by permission from C. F. Powell, P. H. Fowler, and D. H. Perkins, *The Study of Elementary Particles by the Photographic Method* (New York: Pergamon, 1959). First published in *Rep. Prog. Phys.* **13,** 384 (1950).]

$$p + p \to p + p + p + \bar{p} \tag{1.26}$$

($A = 2$ on both sides). But the proton, as the lightest baryon, has nowhere to go; conservation of the baryon number guarantees its absolute stability.* If we are to retain the conservation of baryon number in the light of reaction (1.24), the lambda must be assigned to the baryon family. Over the next few years many more heavy baryons were discovered—the Σ's, the Ξ's, and the Δ's, and so on. [By the way: unlike leptons and baryons, there is *no* conservation of mesons. In pion decay ($\pi^- \to \mu^- + \bar{\nu}_\mu$) a meson disappears, and in lambda decay ($\Lambda \to p^+ + \pi^-$) a meson is created.]

* Recent "grand unified" theories allow for a minute violation of baryon number conservation, and in these theories the proton is *not* absolutely stable. See the article by S. Weinberg in *Scientific American*, June 1981. The experimental situation is discussed by J. M. LoSecco *et al.*, *Scientific American*, June 1985.

It is some measure of the surprise with which these new heavy baryons and mesons were greeted that they came to be known collectively as "strange" particles. In 1952 the first of the modern particle accelerators (the Brookhaven Cosmotron) began operating, and soon it was possible to produce strange particles in the laboratory (before this the only source had been cosmic rays) . . . and with this, the rate of proliferation increased. Willis Lamb began his Nobel Prize acceptance speech in 1955 with the words

> When the Nobel Prizes were first awarded in 1901, physicists knew something of just two objects which are now called "elementary particles": the electron and the proton. A deluge of other "elementary" particles appeared after 1930; neutron, neutrino, μ meson, π meson, heavier mesons, and various hyperons. I have heard it said that "the finder of a new elementary particle used to be rewarded by a Nobel Prize, but such a discovery now ought to be punished by a $10,000 fine". [Source: Les Prix Nobel 1955, The Nobel Foundation, Stockholm.]

Not only were the new particles unexpected; there is a more technical sense in which they seemed "strange": They are *produced* copiously (on a time scale of about 10^{-23} sec), but they *decay* relatively slowly (typically about 10^{-10} sec). This suggested to Pais[17] and others that the mechanism involved in their production is entirely different from that which governs their disintegration. In modern language, the strange particles are *produced* by the *strong* force (the same one that holds the nucleus together), but they *decay* by the *weak* force (the one that accounts for beta decay and all other neutrino processes). The details of Pais's scheme required that the strange particles be produced in *pairs.* The experimental evidence for this was far from clear at that time, but in 1953 Gell-Mann[18] and Nishijima[19] found a beautifully simple, and, as it developed stunningly successful, way to implement and improve Pais's idea. They assigned to each particle a new property (Gell-Mann called it "strangeness") that (like charge, lepton number, and baryon number) is conserved in any strong interaction, but (*un*like those others) is *not* conserved in a weak interaction. In a pion-proton collision, for example, we might produce *two* strange particles:

$$\begin{aligned} \pi^- + p^+ &\rightarrow K^+ + \Sigma^- \\ &\rightarrow K^0 + \Sigma^0 \\ &\rightarrow K^0 + \Lambda \end{aligned} \tag{1.27}$$

Here the K's carry strangeness $S = +1$, the Σ's and the Λ have $S = -1$, and the "ordinary" particles—π, p, and n—have $S = 0$. But we never produce just *one* strange particle:

$$\begin{aligned} \pi^- + p^+ &\not\rightarrow \pi^+ + \Sigma^- \\ &\not\rightarrow \pi^0 + \Lambda \\ &\not\rightarrow K^0 + n \end{aligned} \tag{1.28}$$

On the other hand, when these particles *decay,* strangeness is *not* conserved:

$$\begin{aligned} \Lambda &\rightarrow p^+ + \pi^- \\ \Sigma^+ &\rightarrow p^+ + \pi^0 \\ &\rightarrow n + \pi^+ \end{aligned} \tag{1.29}$$

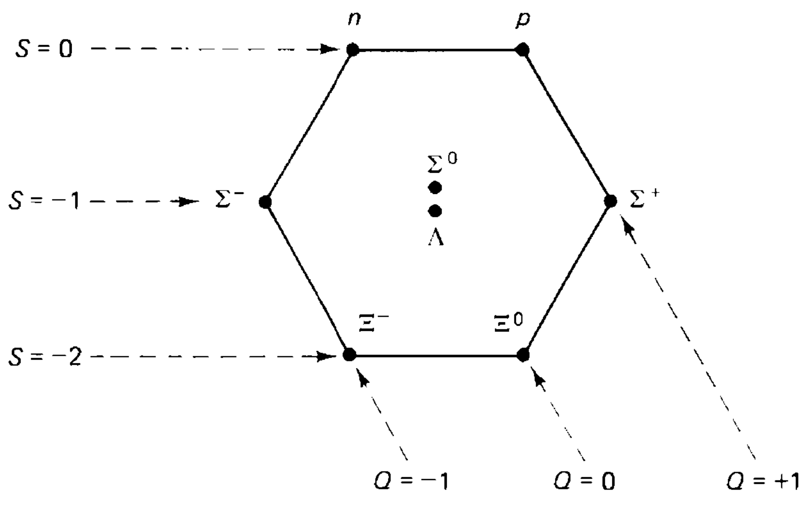
for these are *weak* processes, which do not respect conservation of strangeness.

There is some arbitrariness in the assignment of strangeness numbers, obviously. We could just as well have given $S = +1$ to the Σ's and the Λ, and $S = -1$ to K^+ and K^0; in fact, in retrospect it would have been a little nicer that way. [In exactly the same sense, Benjamin Franklin's original convention for plus and minus charge was perfectly arbitrary at the time, and unfortunate in retrospect since it made the current-carrying particle (the electron) negative.] The significant point is that there exists a consistent assignment of strangeness numbers to all the hadrons (baryons and mesons) that accounts for the observed strong processes and "explains" why the others do not occur. (The leptons and the photon don't experience strong forces at all, so strangeness does not apply to them.)

The garden which seemed so tidy in 1947 had grown into a jungle by 1960, and hadron physics could only be described as chaos. The plethora of strongly interacting particles was divided into two great families—the baryons and the mesons—and the members of each family were distinguished by charge, strangeness, and mass; but beyond that there was no rhyme or reason to it all. This predicament reminded many physicists of the situation in chemistry a century earlier, in the days before the Periodic Table, when scores of elements had been identified, but there was no underlying order or system. In 1960 the elementary particles awaited their own "Periodic Table."[20]

1.7 THE EIGHTFOLD WAY (1961–1964)

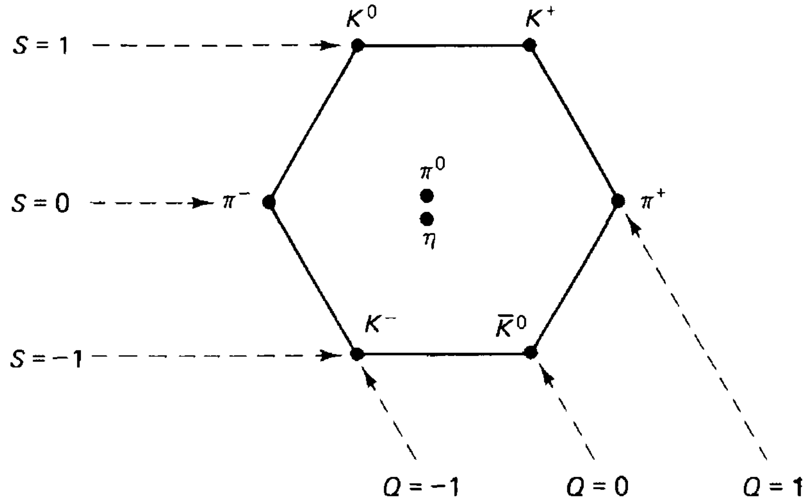
The Mendeleev of elementary particle physics was Murray Gell-Mann, who introduced the so-called *Eightfold Way* in 1961.[21] (Essentially the same scheme was proposed independently by Ne'eman.) The Eightfold Way arranged the baryons and mesons into weird geometrical patterns, according to their charge and strangeness. The eight lightest baryons fit into a hexagonal array, with two particles at the center:



The Baryon Octet

This group is known as the *baryon octet.* Notice that particles of like charge lie along the downward-sloping *diagonal* lines: $Q = +1$ (in units of the proton charge) for the proton and the Σ^+; $Q = 0$ for the neutron, the lambda, the Σ^0, and the Ξ^0; $Q = -1$ for the Σ^- and the Ξ^-. *Horizontal* lines associate particles of like *strangeness:* $S = 0$ for the proton and neutron, $S = -1$ for the middle line and $S = -2$ for the two Ξ's.

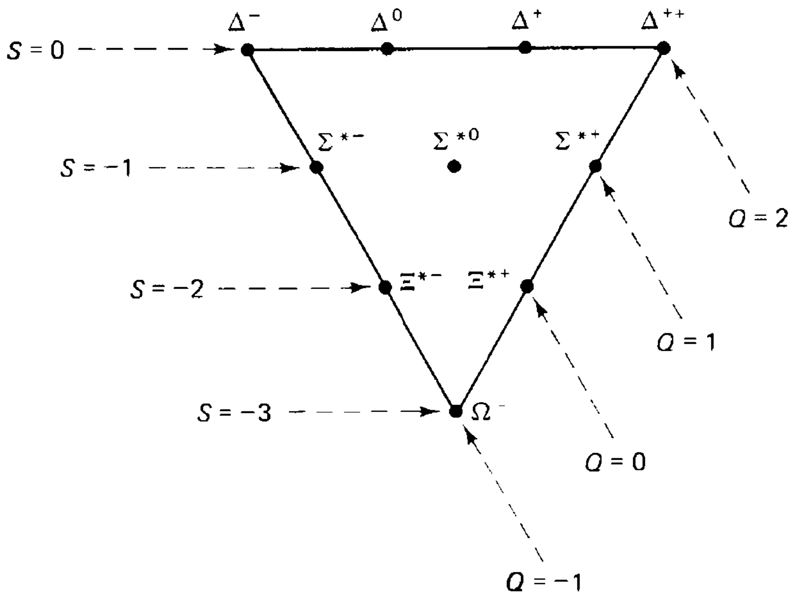
The eight lightest mesons fill a similar hexagonal pattern, forming the (*pseudo-scalar*) *meson octet:*



The Meson Octet

Once again, diagonal lines determine charge, and horizontals determine strangeness; but this time the top line has $S = 1$, the middle line $S = 0$, and the bottom line $S = -1$. (This discrepancy is a historical accident; Gell-Mann could just as well have assigned $S = 1$ to the proton and neutron, $S = 0$ to the Σ's and the Λ, and $S = -1$ to the Ξ's. In 1953 he had no reason to prefer that choice, and it seemed most natural to give the familiar particles—proton, neutron, and pion—a strangeness of zero. After 1961 a new term—*hypercharge*—was introduced, which was equal to S for the mesons and to $S + 1$ for the baryons. But later developments showed that strangeness was the better quantity after all, and the word "hypercharge" has now been taken over for a quite different purpose.)

Hexagons were not the only figures allowed by the Eightfold Way; there was also, for example, a triangular array, incorporating 10 heavier baryons—the *baryon decuplet:*



The Baryon Decuplet

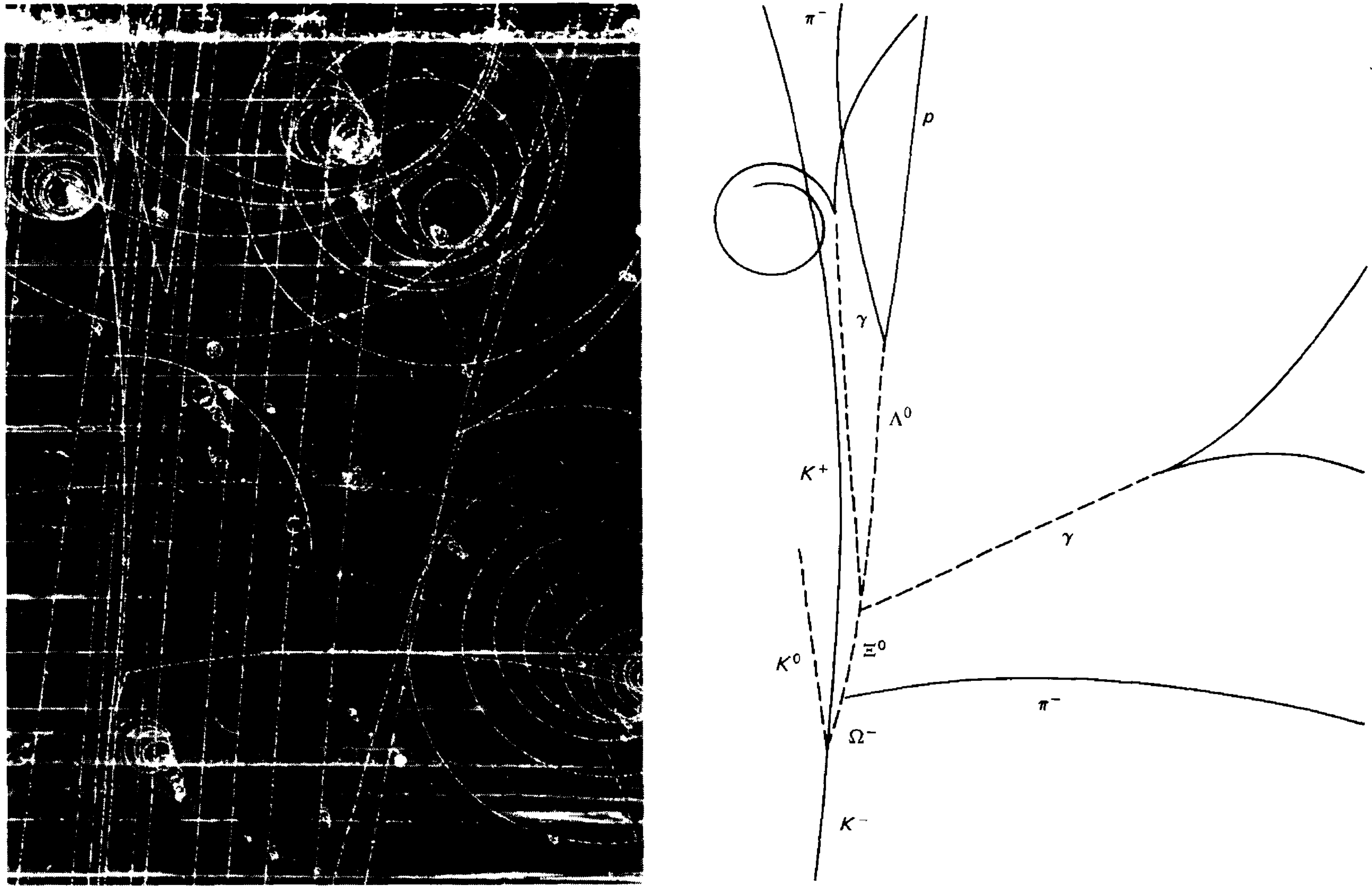


Figure 1.10 The discovery of the Ω^-. The actual bubble chamber photograph is shown on the left; a line diagram of the relevant tracks on the right. (Photo courtesy Brookhaven National Laboratory.)

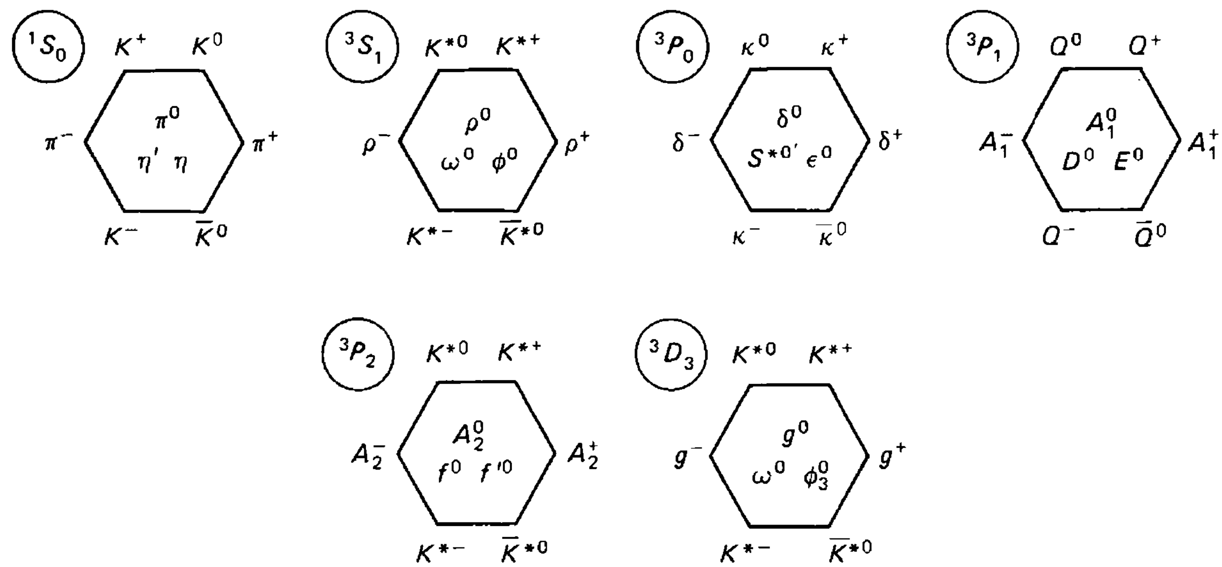


Figure 1.11 Established meson nonets. Obviously, we are running out of letters. It is customary to distinguish different particles represented by the same letter by indicating the mass parenthetically (in MeV/c^2), thus $K^*(892)$, $K^*(1430)$, $K^*(1650)$, and so on. In this figure the supermultiplets are labeled in spectroscopic notation (see Chap. 5). At present, there are no complete baryon supermultiplets beyond the octet and decuplet, although there are many partially filled diagrams.

Now, as Gell-Mann was fitting these particles into the decuplet, an absolutely lovely thing happened. Nine of the particles were known experimentally, but at that time the tenth particle—the one at the very bottom, with a charge of -1 and strangeness -3—was missing: No particle with these properties had ever been detected in the laboratory.[22] Gell-Mann boldly predicted that such a particle would be found, and told the experimentalists exactly how to produce it. Moreover, he calculated its mass—as you can for yourself, in Problem 1.6—and its lifetime, Problem 1.8—and sure enough, in 1964 the famous *omega-minus* particle was discovered,[23] precisely as Gell-Mann had predicted (see Fig. 1.10).

Since the discovery of the omega-minus (Ω^-), no one has seriously doubted that the Eightfold Way is correct.* Over the next 10 years, every new hadron found a place in one of the Eightfold Way *supermultiplets*. Some of these are shown in Figure 1.11. (This is not to say there were no false alarms; particles have a way of appearing and then *dis*appearing. Of the 26 mesons listed on a standard table in 1963, 19 were later found to be spurious!) In addition to the baryon octet, decuplet, and so on, there exist of course an *anti*baryon octet, decuplet, etc., with opposite charge and opposite strangeness. However, in the case of the mesons, the antiparticles lie in the *same supermultiplet* as the corresponding particles, in the diametrically opposite positions. Thus the antiparticle

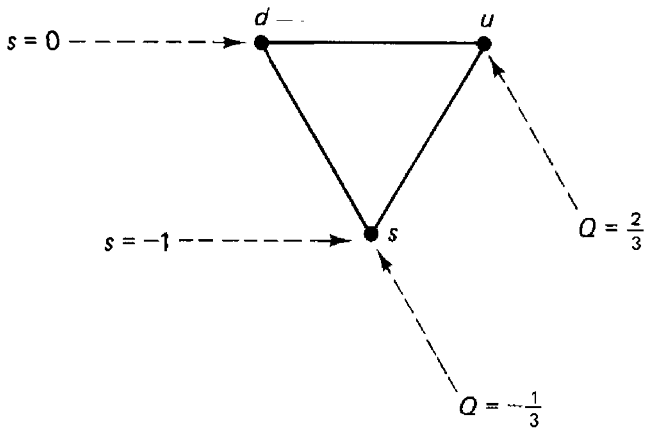
* A similar thing happened in the case of the Periodic Table. There were three famous "holes" (missing elements) on Mendeleev's chart, and he predicted that new elements would be discovered to fill in the gaps. Like Gell-Mann, he confidently described their properties, and within 20 years all three—gallium, scandium, and germanium—were found.

of the pi-plus is the pi-minus, the anti-*K*-minus is the *K*-plus, and so on (the pi-zero and the eta are their *own* antiparticles).

Classification is the first stage in the development of any science. The Eightfold Way did more than merely classify the hadrons, but its real importance lies in the organizational structure it provided. I think it's fair to say that the Eightfold Way initiated the modern era in particle physics.

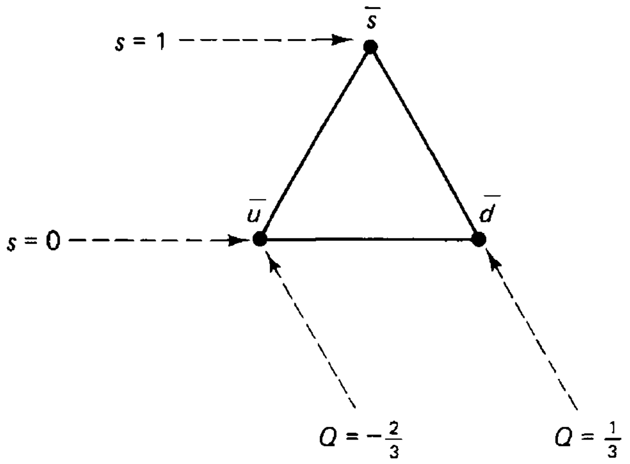
1.8 THE QUARK MODEL (1964)

But the very success of the Eightfold Way begs the question: *Why* do the hadrons fit into these curious patterns? The Periodic Table had to wait many years for quantum mechanics and the Pauli exclusion principle to provide its explanation. An understanding of the Eightfold Way, however, came already in 1964, when Gell-Mann and Zweig independently proposed that all hadrons are in fact composed of even more elementary constituents, which Gell-Mann called *quarks*.[24] The quarks come in three types (or "flavors"), forming a triangular "Eightfold-Way" pattern:



The Quarks

The u (for "up") quark carries a charge of $\frac{2}{3}$ and a strangeness of zero; the d ("down") quark carries a charge of $-\frac{1}{3}$ and $S = 0$; the s (originally "sideways", but now more commonly "strange") quark has $Q = -\frac{1}{3}$ and $S = -1$. To each quark (q) there corresponds an *anti*quark ($\bar{q}$), with the opposite charge and strangeness:



The Antiquarks

The quark model asserts that

1. Every baryon is composed of three quarks (and every *anti*baryon is composed of three *anti*quarks).
2. Every meson is composed of a quark and an antiquark.

With these two rules it is a matter of elementary arithmetic to construct the baryon decuplet and the meson octet. All we need to do is list the combinations of three quarks (or quark–antiquark pairs), and add up their charge and strangeness:

THE BARYON DECUPLET

qqq	Q	S	Baryon
uuu	2	0	Δ^{++}
uud	1	0	Δ^{+}
udd	0	0	Δ^{0}
ddd	-1	0	Δ^{-}
uus	1	-1	Σ^{*+}
uds	0	-1	Σ^{*0}
dds	-1	-1	Σ^{*-}
uss	0	-2	Ξ^{*0}
dss	-1	-2	Ξ^{*-}
sss	-1	-3	Ω^{-}

Notice that there are 10 combinations of three quarks. Three u's, for instance, at $Q = \frac{2}{3}$ each, yield a total charge of $+2$, and a strangeness of zero. This is the Δ^{++} particle. Continuing down the table, we find all the members of the decuplet ending with the Ω^{-}, which is evidently made of three s quarks.

A similar enumeration of the quark–antiquark combinations yields the meson table:

THE MESON NONET

$q\bar{q}$	Q	S	Meson
$u\bar{u}$	0	0	π^{0}
$u\bar{d}$	1	0	π^{+}
$d\bar{u}$	-1	0	π^{-}
$d\bar{d}$	0	0	η
$u\bar{s}$	1	1	K^{+}
$d\bar{s}$	0	1	K^{0}
$s\bar{u}$	-1	-1	K^{-}
$s\bar{d}$	0	-1	$\bar{K}^{0}$
$s\bar{s}$	0	0	??

But wait! There are *nine* combinations here, and only eight particles in the meson octet. The quark model requires that there be a third meson (in addition to the π^0 and the η) with $Q = 0$ and $S = 0$. As it turns out, just such a particle had already been found experimentally—the η'. In the Eightfold Way the η' had been classified as a *singlet*, all by itself. According to the quark model it properly belongs with the other eight mesons to form a *meson nonet*. (Actually, since $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ all have $Q = 0$ and $S = 0$, it is not possible to say, on the basis of anything we have done so far, which is the π^0, which the η, and which the η'. But never mind, the point is that there are *three* mesons with $Q = S = 0$.) By the way, the *anti*mesons automatically fall in the same supermultiplet as the mesons: $u\bar{d}$ is the antiparticle of $d\bar{u}$, and vice versa.

You may have noticed that I avoided talking about the baryon *octet*—and it is far from obvious how we are going to get *eight* baryons by putting together three quarks. In truth, the procedure is perfectly straightforward, but it does call for some facility in handling spins, and I would rather save it until Chapter 5. For now, I'll just tantalize you with the mysterious observation that if you take the decuplet and knock off the three corners (where the quarks are identical—*uuu*, *ddd*, and *sss*), and double the center (where all three are different—*uds*), you obtain precisely the eight states in the baryon octet. So the same set of quarks can account for the octet; it's just that some combinations do not appear at all, and one appears twice.

Indeed, all the Eightfold Way supermultiplets emerge in a natural way from the quark model. Of course, the same combination of quarks can go to make a number of different particles: The delta-plus and the proton are both composed of two *u*'s and a *d*; the pi-plus and the rho-plus are both $u\bar{d}$; and so on. Just as the hydrogen atom (electron plus proton) has many different energy levels, so a given collection of quarks can bind together in many different ways. But whereas the various energy levels in the electron/proton system are relatively close together (the spacings are typically several electron volts, in an atom whose rest energy is nearly 10^9 electron volts), so that we naturally think of them all as "hydrogen," the energy spacings for different states of a bound quark system are very large, and we normally regard them as distinct particles. Thus we can, in principle, construct an infinite number of hadrons out of only three quarks. Notice, however, that *some* things are absolutely excluded in the quark model: For example, a baryon with $S = 0$ and $Q = -2$; no combination of the three quarks can produce these numbers. Nor can there be a *meson* with a charge of $+2$ (like the Δ^{++} baryon) or a strangeness of -3 (like the Ω^-). For a long time there were major experimental searches for these so-called "exotic" particles; their discovery would be devastating for the quark model, but none has ever been found (see Problem 1.11).

The quark model *does*, however, suffer from one profound embarrassment: In spite of the most diligent search over a period of 20 years, no one has ever seen an individual quark. Now, if a proton is really made out of three quarks, you'd think that if you hit one hard enough, the quarks ought to come popping out. Nor would they be hard to recognize, carrying as they do the conspicuous label of fractional charge; an ordinary Millikan oil drop experiment would clinch the identification. Moreover, at least one of the quarks should be absolutely stable; what could it decay into, since there is no lighter particle with fractional charge? So quarks ought to be *easy* to produce, *easy* to identify, and *easy* to store, and yet, no one has ever found one.

The failure of experiments to produce isolated quarks occasioned widespread skepticism about the quark model in the late sixties and early seventies. Those who clung to the model tried to conceal their disappointment by introducing the notion of *quark confinement:* perhaps, for reasons not yet understood, quarks are *absolutely confined* within baryons and mesons, so that no matter how hard you try, you cannot get them out. Of course, this doesn't explain anything, it just gives a name to our frustration. But at least it poses sharply what has become a crucial theoretical problem for the eighties: to discover the mechanism responsible for quark confinement. There are some indications that the solution may be at hand.[25]

Even if all quarks are stuck inside hadrons, this does not mean they are inaccessible to experimental study. One can probe the inside of a proton in much the same way as Rutherford probed the inside of an atom—by firing something into it. Such experiments were carried out in the late sixties using high-energy electrons at the Stanford Linear Accelerator Center (SLAC). They were repeated in the early seventies using neutrino beams at CERN, and later still using protons. The results of these so-called "deep inelastic scattering" experiments were strikingly reminiscent of Rutherford's (Fig. 1.12): *Most* of the incident particles pass right through, whereas a small number bounce back sharply. This means that the charge of the proton is concentrated in small lumps, just as Rutherford's results indicated that the positive charge in an atom is concentrated at the nucleus.[26] However, in the case of the proton the evidence suggests *three* lumps,

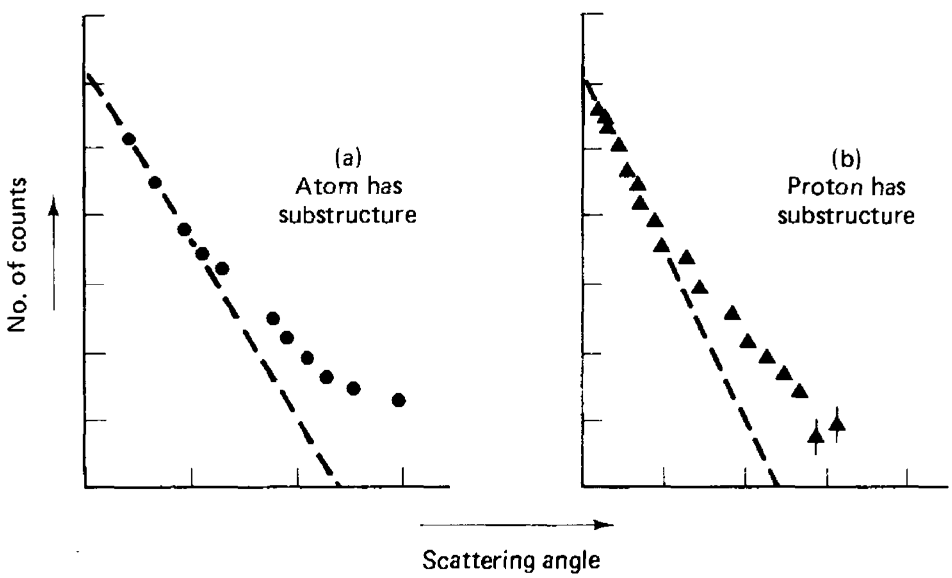


Figure 1.12 (a) In Rutherford scattering the number of particles deflected through large angles indicates that the atom has internal structure (a nucleus). (b) In deep inelastic scattering the number of particles deflected through large angles indicates that the proton has internal structure (quarks). The dashed lines show what you would expect if the positive charge were uniformly distributed over the volume of (a) the atom, (b) the proton. [*Source:* F. Halzen and A. D. Martin, *Quarks and Leptons* (New York: Wiley, 1984), p. 17. Copyright © John Wiley & Sons, Inc. Reprinted by permission.]

instead of *one*. This is strong support for the quark model, obviously, but still not conclusive.

Finally, there was a theoretical objection to the quark model: It appears to violate the Pauli exclusion principle. In Pauli's original formulation the exclusion principle stated that no two electrons can occupy the same state. However, it was later realized that the same rule applies to all particles of half-integer spin (the proof of this is one of the most important achievements of quantum field theory). In particular, the exclusion principle should apply to quarks, which, as we shall see, must carry spin $\frac{1}{2}$. Now the Δ^{++}, for instance, is supposed to consist of three identical u quarks in the same state; it (and also the Δ^- and the Ω^-) appear to be inconsistent with the Pauli principle. In 1964, O. W. Greenberg proposed a way out of this dilemma:[27] He suggested that quarks not only come in three *flavors* (u, d, and s) but each of these also comes in three *colors* ("red," "green," and "blue," say). To make a baryon, we simply take one quark of each color, then the three u's in Δ^{++} are no longer identical (one's red, one's green, and one's blue). Since the exclusion principle only applies to *identical* particles, the problem evaporates.

The color hypothesis sounds like sleight of hand, and many people initially considered it the last gasp of the quark model. As it turned out, the introduction of color was one of the most fruitful ideas of our time. I need hardly say that the term "color" here has absolutely no connection with the ordinary meaning of the word. Redness, blueness, and greenness are simply *labels* used to denote three new properties that, in addition to charge and strangeness, the quarks possess. A *red* quark carries one unit of redness, zero blueness, and zero greenness; its antiparticle carries *minus* one unit of redness, and so on. We could just as well call these quantities X-ness, Y-ness, and Z-ness, for instance. However, the color terminology has one especially nice feature: It suggests a delightfully simple characterization of the particular quark combinations that are found in nature.

All naturally occurring particles are colorless.

By "colorless" I mean that *either* the total amount of each color is zero *or* all three colors are present in equal amounts. (The latter case mimics the optical fact that light beams of three primary colors combine to make white.) This clever rule "explains" (if that's the word for it) why you can't make a particle out of *two* quarks, or *four* quarks, and for that matter why *individual* quarks do not occur in nature. The only colorless combinations you can make are $q\bar{q}$ (the mesons), qqq (the baryons), and $\bar{q}\bar{q}\bar{q}$ (the antibaryons). (You could have *six* quarks, of course, but we would interpret that as a bound state of two baryons.)

1.9 THE NOVEMBER REVOLUTION AND ITS AFTERMATH (1974–1983)

The decade from 1964 to 1974 was a barren time for elementary particle physics. The quark model, which had seemed so promising at the beginning, was in an uncomfortable state of limbo by the end. It had had some striking successes: It neatly explained the Eightfold Way, and correctly predicted the lumpy structure of the proton. But it had two conspicuous defects: the experimental absence of free quarks and inconsistency with the Pauli principle. Those who liked the model papered over these failures with what seemed at the time to be rather transparent rationalizations: the idea of quark confinement and the color hypothesis. But I think it is safe to say that by 1974 most elementary particle physicists felt queasy, at best, about the quark model. The lumps inside the proton were called *partons,* and it was unfashionable to identify them explicitly with quarks.

Curiously enough, what rescued the quark model was not the discovery of free quarks, or an explanation of quark confinement, or confirmation of the color hypothesis, but something entirely different and (almost)[28] completely unexpected: the discovery of the psi meson. The ψ was first observed at Brookhaven by a group under C. C. Ting, in the summer of 1974. But Ting wanted to check his results before announcing them publicly, and the discovery remained an astonishingly well-kept secret until the weekend of November 10–11, when the new particle was discovered independently by Burton Richter's group at SLAC. The two teams then published simultaneously,[29] Ting naming the particle J, and Richter calling it ψ. The J/ψ was an electrically neutral, extremely heavy meson—more than three times the weight of a proton (the original notion that mesons are "middle-weight" and baryons "heavy-weight" had long since gone by the boards). But what made this particle so unusual was its extraordinarily long lifetime. For the ψ lasted fully 10^{-20} seconds before disintegrating. Now, 10^{-20} seconds may not impress you as a particularly long time, but you must understand that the *typical* lifetimes for hadrons in this mass range are on the order of 10^{-23} seconds. So the ψ has a lifetime about a thousand times longer than any comparable particle. It's as though someone came upon an isolated village in Peru or the Caucasus where people live to be 70,000 years old. That wouldn't just be some actuarial anomaly; it would be a sign of fundamentally new biology at work. And so it was with the ψ: its long lifetime, to those who understood, spoke of fundamentally new physics. For good reason, the events precipitated by the discovery of the ψ came to be known as the *November Revolution.*[30]

In the months that followed, the true nature of the ψ meson was the subject of lively debate, but the explanation that won was provided by the quark model. It is now universally accepted that the ψ represents a bound state of a new (fourth) quark, the c (for *charm*) and its antiquark: $\psi = (c\bar{c})$. Actually, the idea of a fourth flavor, and even the whimsical name, had been introduced many years earlier, by Bjorken and Glashow.[31] Indeed, there was an intriguing parallel between the leptons and the quarks:

Leptons: $e,\ \nu_e,\ \mu,\ \nu_\mu$

Quarks: $d,\ u,\ s$

If all mesons and baryons are made out of quarks, these two families are left as

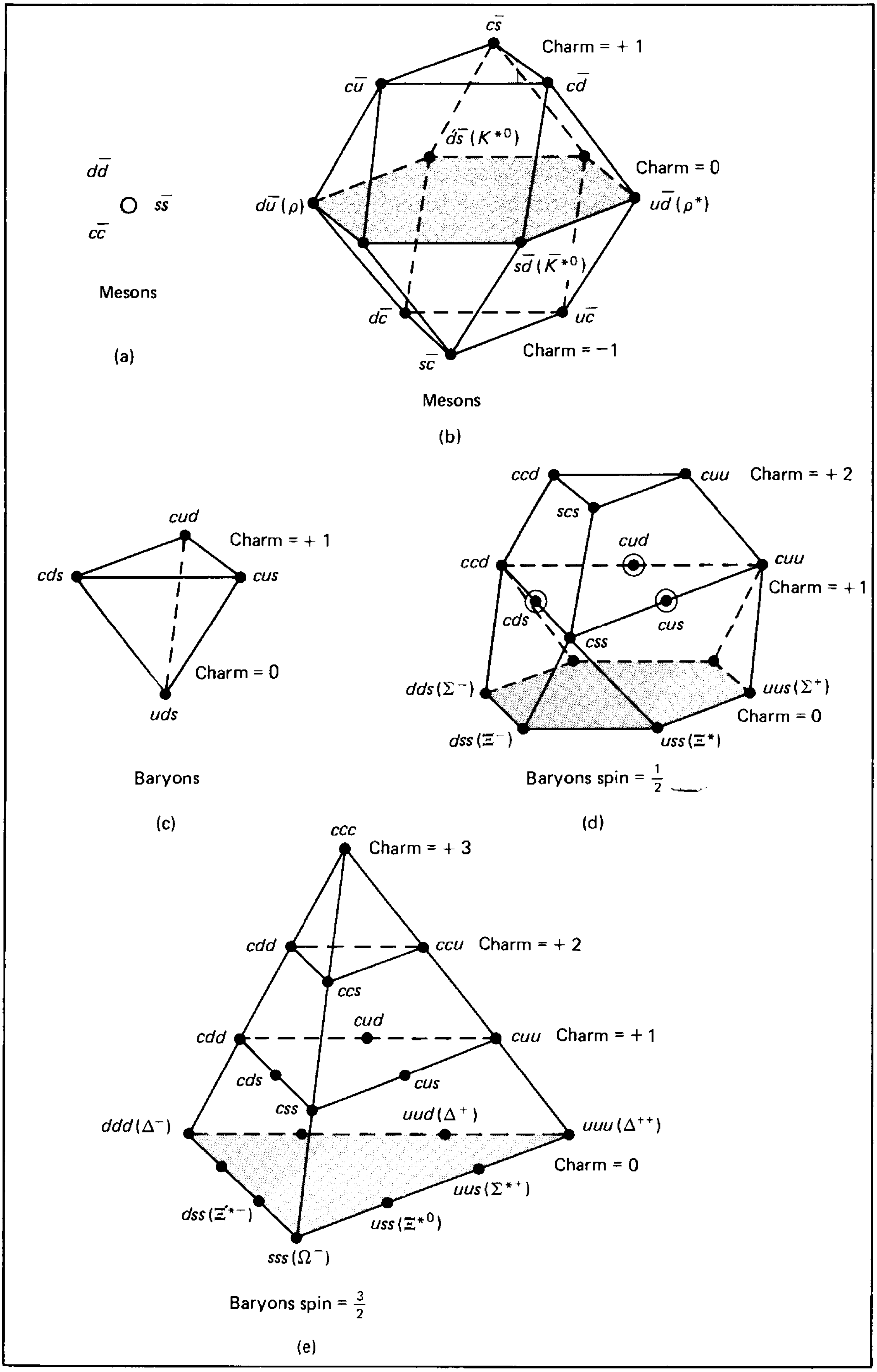


Figure 1.13 Supermultiplets constructed with four quarks. (From "Quarks with Color and Flavor," by S. Glashow. Copyright © Oct. 1975 by Scientific American, Inc. All rights reserved.)

the *truly* fundamental particles. But why *four* leptons and only *three* quarks? Wouldn't it be nicer if there were four of each? Later, Glashow, Iliopoulos, and Maiani[32] offered more compelling technical reasons for wanting a fourth quark, but the simple idea of a parallel between quarks and leptons is another of those farfetched speculations that turned out to have more substance than their authors could have imagined.

So when the ψ was discovered, the quark model was ready and waiting with an explanation. Moreover, it was an explanation pregnant with implications. For if a fourth quark exists, there should be all kinds of new baryons and mesons, carrying various amounts of charm. Some of these are shown in Figure 1.13; you can work out the possibilities for yourself (Problems 1.14 and 1.15). Notice that the ψ itself carries no *net* charm, for if the c is assigned a charm of +1, then $\bar{c}$ will have a charm of −1; the charm of the ψ is, if you will, "hidden." To confirm the charm hypothesis it was important to produce a particle with "naked" (or "bare") charm.[33] The first evidence for charmed baryons ($\Lambda_c^+ = udc$ and possibly $\Sigma_c^{++} = uuc$) appeared already in 1975 (Fig. 1.14);[34] the first charmed mesons ($D^0 = c\bar{u}$ and $D^+ = c\bar{d}$) were found in 1976,[35] and the charmed strange meson ($F^+ = c\bar{s}$) in 1977.[36] (The F meson was recently renamed D_s. There is also some evidence for usc and ssc.) With these discoveries the interpretation of the ψ as $c\bar{c}$ was established beyond reasonable doubt. More important, the quark model itself was put back on its feet.

However, the story does not end there, for in 1975 a new *lepton* was discovered,[37] spoiling Glashow's symmetry. This new particle (the tau) presumably has its own neutrino, so we are up to six leptons, and only four quarks. But don't dispair, because two years later a new heavy meson (the *upsilon*) was discovered,[38] and quickly recognized as the carrier of a fifth quark, b (for *beauty*, or *bottom*, depending on your taste): $\Upsilon = b\bar{b}$. Immediately the search began for mesons and hadrons exhibiting "naked beauty" (or "bare bottom"). (I'm sorry. *I* didn't invent this terminology. In a way, its silliness is a reminder of how wary people were of taking the quark model seriously, in the early days.) The first *beautiful* baryon, $\Lambda_b = udb$, may have been observed in 1981[39] (the claim is hotly contested[40]); the first beautiful mesons ($B^0 = b\bar{d}$ and $B^- = b\bar{u}$) were found in 1983.[41] At this point it doesn't take much imagination to predict that a sixth quark will eventually be found; it already has a name: t (for *truth*, of course, or *top*). If and when the t quark is discovered (there were some indications in the summer of 1984 that it may have been seen at CERN), Glashow's symmetry will be restored, with six leptons and six quarks. And there (knock on wood) the proliferation stops.

1.10 INTERMEDIATE VECTOR BOSONS (1983)

In his original theory of beta decay (1933) Fermi treated the process as a contact interaction, occurring at a single point, and therefore requiring no mediating

Figure 1.14 The charmed baryon. The probable interpretation of this event is $\nu_\mu + p \rightarrow \Lambda_c^+ + \mu^- + \pi^+ + \pi^-$. The charmed baryon decays ($\Lambda_c^+ \rightarrow \Lambda + \pi^+$) too soon to leave a track, but the subsequent decay of the Λ is clearly visible. (Photo courtesy of N. P. Samios, Brookhaven National Laboratory.)

particle. As it happens, the weak force (which is responsible for beta decay) *is* of extremely short range, so that Fermi's model was not far from the truth, and yields excellent approximate results at low energies. However, it was widely recognized that this approach was bound to fail at high energies, and would eventually have to be supplanted with a theory in which the interaction was mediated by the exchange of some particle. The mediator came to be known by the prosaic name *intermediate vector boson.* The challenge for theorists was to predict the properties of the intermediate vector boson, and for experimentalists, to produce one in the laboratory. You may recall that Yukawa, faced with the analogous problem for the strong force, was able to estimate the mass of the pion in terms of the range of the force, which he took to be roughly the same as the size of a nucleus. But we have no corresponding way to measure the range of the weak force; there are no "weak bound states" whose size would inform us—the weak force is simply too feeble to bind particles together. For many years predictions of the intermediate vector boson mass were little more than educated guesses (the "education" coming largely from the failure of experiments at progressively higher energies to detect the particle). By 1962 it was known that the mass had to be at least half the proton mass; 10 years later the experimental lower limit had grown to 2.5 proton masses.

But it was not until the emergence of the electroweak theory of Glashow, Weinberg, and Salam that a really firm prediction of the mass was possible. In this theory there are in fact *three* intermediate vector bosons, two of them charged ($W^{\pm}$) and one neutral (Z). Their masses were calculated to be[42]

$$M_W = 82 \pm 2\ \text{GeV}/c^2, \qquad M_Z = 92 \pm 2\ \text{GeV}/c^2 \tag{1.30}$$

In the late seventies, CERN began construction of a proton–antiproton collider designed specifically to produce these extremely heavy particles (bear in mind that the mass of the proton is 0.94 GeV/c^2, so we're talking about something nearly 100 times as heavy). In January 1983 the discovery of the W (at 81 $\pm$ 5 GeV/c^2) was reported by Carlo Rubbia's group,[43] and five months later the same team announced discovery of the Z (at 95 $\pm$ 3 GeV/c^2).[44] These experiments represent an extraordinary technical triumph,[45] and they were of fundamental importance in confirming a crucial aspect of the Standard Model, to which the physics community was by that time heavily committed (and for which a Nobel Prize had already been awarded). Unlike the strange particles or the ψ, however, the intermediate vector bosons were long awaited and universally expected, so the general reaction was a sigh of relief, not shock or surprise.

1.11 THE STANDARD MODEL (1978–?)

In the current view, then, all matter is made out of three kinds of elementary particles: leptons, quarks, and mediators. There are six leptons, classified according to their charge (Q), electron number (L_e), muon number (L_μ), and tau number (L_τ). They fall naturally into three *families* (or *generations*):

LEPTON CLASSIFICATION

	l	Q	L_e	L_μ	L_τ
First generation	e	−1	1	0	0
	ν_e	0	1	0	0
Second generation	μ	−1	0	1	0
	ν_μ	0	0	1	0
Third generation	τ	−1	0	0	1
	ν_τ	0	0	0	1

There are also six antileptons, with all the signs reversed. The positron, for example, carries a charge of +1 and an electron number −1. So there are really 12 leptons, all told.

Similarly, there are six "flavors" of quarks, which are classified according to charge, strangeness (S), charm (C), beauty (B), and truth (T). [For consistency, I suppose we should include "upness" (U) and "downness" (D), although these terms are seldom used. They are redundant, inasmuch as the only quark with $S = C = B = T = 0$ and $Q = \frac{2}{3}$, for instance, is the up quark, so it is not necessary to specify $U = 1$ and $D = 0$ as well.] The quarks, too, fall into three generations:

QUARK CLASSIFICATION

	q	Q	D	U	S	C	B	T
First generation	d	$-\frac{1}{3}$	−1	0	0	0	0	0
	u	$\frac{2}{3}$	0	1	0	0	0	0
Second generation	s	$-\frac{1}{3}$	0	0	−1	0	0	0
	c	$\frac{2}{3}$	0	0	0	1	0	0
Third generation	b	$-\frac{1}{3}$	0	0	0	0	−1	0
	t	$\frac{2}{3}$	0	0	0	0	0	1

Again, all signs would be reversed on the table of antiquarks. Meanwhile, each quark and antiquark comes in three colors, so there are 36 of them in all.

Finally, every interaction has its mediators: the photon for the electromagnetic force, two W's and a Z for the weak force, the graviton (presumably) for gravity, . . . but what about the strong force? In Yukawa's original theory (1934) the mediator of strong forces was the pion, but with the discovery of heavy mesons this simple picture could not stand; protons and neutrons could now exchange rho's and eta's and K's and phi's and all the rest of them. The

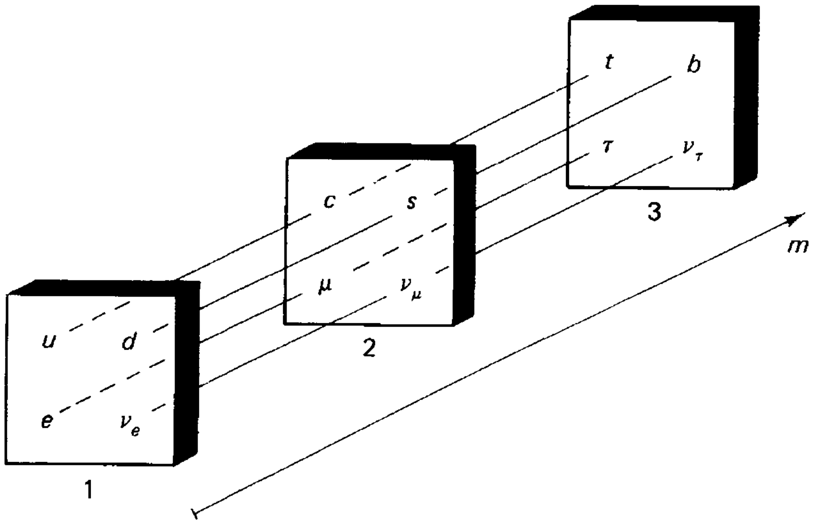


Figure 1.15 The three generations of quarks and leptons, in order of increasing mass.

quark model brought an even more radical revision, for if protons, neutrons, and mesons are complicated composite structures, there is no reason to believe their interaction *should* be simple. To study the strong force at the fundamental level, one should look, rather, at the interaction between individual quarks. So the question becomes: What particle is exchanged between two quarks, in a strong process? This mediator is called the *gluon,* and in the Standard Model there are eight of them. As we shall see, the gluons themselves carry color, and therefore (like the quarks) should not exist as isolated particles. We can hope to detect gluons only within hadrons, or in colorless combinations with other gluons (*glueballs*). Nevertheless, there is substantial indirect experimental evidence for the existence of gluons: The deep inelastic scattering experiments showed that roughly half the momentum of a proton is carried by electrically neutral constituents, presumably gluons; the *jet* structure characteristic of proton scattering at high energies can be explained in terms of the disintegration of quarks and gluons in flight;[46] and glueballs may conceivably have been observed.[47] But no one would say that the experimental evidence is really *compelling,* at this stage.

This is all adding up to an embarrassingly large number of supposedly "elementary" particles: 12 leptons, 36 quarks, 12 mediators (I won't count the graviton, since gravity is not included in the Standard Model). And, as we shall see later, the Glashow-Weinberg-Salam theory calls for at least one *Higgs* particle, so we have a minimum of 61 particles to contend with. Informed by our experience first with atoms and later with hadrons, many people have suggested that some, at least, of these 61 must be composites of more elementary subparticles (see Problem 1.17).[48] Such speculations lie beyond the Standard Model and outside the scope of this book. Personally, I do not think the large number of "elementary" particles in the Standard Model is by itself alarming, for they are tightly interrelated. The eight gluons, for example, are identical except for their colors, and the second and third generations mimic the first (Fig. 1.16). In the next chapter we shall see how this structure leads to the first systematic and comprehensive theory of elementary particle dynamics.

REFERENCES AND NOTES

1. There are many good discussions of the history of elementary particle physics. My own favorite is a delightful little book by C. N. Yang, *Elementary Particles* (Princeton, N.J.: Princeton University Press, 1961). More up-to-date accounts are J. S. Trefil's *From Atoms to Quarks* (New York: Scribners, 1980) and F. E. Close's *The Cosmic Onion* (London: Heinemann Educational Books, 1983). The early days are treated well in A. Keller's *The Infancy of Atomic Physics* (Oxford: Oxford University Press, 1983) and S. Weinberg's *Subatomic Particles* (New York: Scientific American Library, 1983). For a fascinating and comprehensive account, see A. Pais, *Inward Bound* (Oxford: Clarendon Press, 1986).
2. The story is beautifully told by A. Pais in his biography of Einstein, *Subtle is the Lord* (Oxford: Clarendon Press, 1982).
3. R. A. Millikan, *Phys. Rev.* **7,** 18 (1916). Quoted in reference [2].
4. M. Conversi, E. Pancini, and O. Piccioni, *Phys. Rev.* **71,** 209 (1947).
5. C. M. G. Lattes et al., *Nature* **159,** 694 (1947); **160,** 453, 486 (1947).
6. R. E. Marshak and H. A. Bethe, *Phys. Rev.* **72,** 506 (1947).
7. For an informal history of this discovery, see C. D. Anderson, *Am. J. Phys.* **29,** 825 (1961).
8. O. Chamberlain et al., *Phys. Rev.* **100,** 947 (1955); B. Cork et al., *Phys. Rev.* **104,** 1193 (1956). See also the *Scientific American* articles by E. Segrè and C. E. Wiegand (June 1956) and by G. Burbridge and F. Hoyle (April 1958).
9. The history of the neutrino is a fascinating story in its own right. See, for instance, J. Bernstein's *The Elusive Neutrino* (AEC Publications), L. M. Brown, *Physics Today* (Sept. 1978), or P. Morrison, *Scientific American* (January 1956). An extensive and useful bibliography on the neutrino is provided by L. M. Lederman, *Am. J. Phys.* **38,** 129 (1970).
10. F. Reines and C. L. Cowan, Jr., *Phys. Rev.* **92,** 8301 (1953); C. L. Cowan et al., *Science* **124,** 103 (1956).
11. R. Davis and D. S. Harmer, *Bull. Am. Phys. Soc.* **4,** 217 (1959). See also C. L. Cowan, Jr., and F. Raines, *Phys. Rev.* **106,** 825 (1957).
12. E. J. Konopinski and H. M. Mahmoud, *Phys. Rev.* **92,** 1045 (1953).
13. B. Pontecorvo, *Soviet Physics JETP* **37,** 1236 (1960) [translated from **37,** 1751 (1959)]; T. D. Lee, Rochester Conference 1960, p. 567.
14. G. Danby et al., *Phys. Rev. Lett.* **9,** 36 (1962). See also L. Lederman, *Scientific American* (March 1963).
15. G. D. Rochester and C. C. Butler, *Nature* **160,** 855 (1947). See also G. D. Rochester's memoir in Y. Sekido and H. Elliot (eds.), *Early History of Cosmic Ray Studies* (Dordrecht: Reidel, 1985), p. 299.
16. Stückelberg himself did not use the term *baryon,* which was introduced by A. Pais in 1953, *Prog. Theor. Phys.* **10,** 457 (1953).
17. A. Pais, *Phys. Rev.* **86,** 663 (1952). The same (copious production, slow decay) could be said for the pion (and for that matter the neutron). But their decays produce neutrinos, and people were used to the idea that neutrino interactions are weak. What was new was a purely hadronic decay whose rate was characteristic of neutrino-type processes.
18. M. Gell-Mann, *Phys. Rev.* **92,** 883 (1953), *Nuovo Cimento* **4,** Suppl. 2, 848 (1956).
19. T. Nakano and K. Nishijima, *Prog. Theor. Phys.* **10,** 581 (1953).
20. Since the early sixties, the Particle Data Group at Berkeley has periodically issued a listing of the established particles and their properties. These are published in *Reviews*

of Modern Physics and summarized in a booklet that can be obtained by writing to Technical Information Department, Lawrence Berkeley Laboratory, Berkeley, CA 94720. In the early days this summary took the form of "wallet cards," but by 1984 it had grown to a densely packed 163 pages. I shall refer to it as the Particle Data Booklet. No student of elementary particle physics should be without it.

21. The original papers are collected in M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (New York: Benjamin, 1964).
22. Actually, there is a possibility that it was seen in a cosmic ray experiment in 1954 [Y. Eisenberg, *Phys. Rev.* **96,** 541 (1954)], but the identification was ambiguous.
23. V. E. Barnes et al., *Phys. Rev. Lett.* **12,** 204 (1964) (reprinted in ref. 21). See also W. B. Fowler and N. P. Samios, *Sci. Am.* (October 1964).
24. An extensive bibliography on the quark model, and useful commentary, is given by O. W. Greenberg, *Am. J. Phys.* **50,** 1074 (1982). Many of the classic papers (including the original unpublished one by G. Zweig) are reprinted in D. B. Lichtenberg and S. P. Rosen, eds., *Developments in the Quark Theory of Hadrons* (Nonantum: Hadronic Press, 1980).
25. Y. Nambu, *Scientific American* (November 1976); K. Johnson, *Sci. Am.* (July 1979); C. Rebbi, *Sci. Am.* (February 1983).
26. M. Jacob and P. Landshoff, *Sci. Am.* (March 1980).
27. O. W. Greenberg, *Phys. Rev. Lett.* **13,** 598 (1964). Greenberg did not use this language; the terminology was introduced by D. B. Lichtenberg, in *Unitary Symmetry and Elementary Particles* (New York: Academic Press, 1970).
28. An exception was J. Iliopoulos. At an international conference of particle physicists held in London in the summer of 1974, he remarked, "I am ready to bet now a whole case [of wine] that the entire next Conference will be dominated by the discovery of charmed particles."
29. J. J. Aubert et al., *Phys. Rev. Lett.* **33,** 1404 (1974); J.-E. Augustin et al., *Phys. Rev. Lett.* **33,** 1406 (1974).
30. A useful bibliography on this material, and reprints of the major articles, is given in *New Physics,* J. Rosner, ed., published by the American Association of Physics Teachers, New York (1981). The excitement of the November Revolution is captured in the SLAC publication *Beam Line,* Volume 7, No. 11, November 1976. See also the *Scientific American* articles by S. D. Drell (June 1975) and S. L. Glashow (October 1975).
31. B. J. Bjorken and S. L. Glashow, *Phys. Lett.* **11,** 255 (1964).
32. S. L. Glashow, J. Iliopoulos, and L. Maiani, *Phys. Rev.* **D2,** 1285 (1970).
33. R. F. Schwitters, *Sci. Am.* (October 1977).
34. E. G. Cazzoli et al., *Phys. Rev. Lett.* **34,** 1125 (1975).
35. G. Goldhaber et al., *Phys. Rev. Lett.* **37,** 255 (1976); I. Peruzzi, *Phys. Rev. Lett.* **37,** 569 (1976).
36. R. Brandelik et al., *Phys. Lett.* **B70,** 132 (1977).
37. M. Perl et al., *Phys. Rev. Lett.* **35,** 1489 (1975). See also M. Perl and W. Kirk, *Sci. Am.* (March 1978).
38. S. W. Herb et al., *Phys. Rev. Lett.* **39,** 252 (1977). See also L. M. Lederman, *Sci. Am.* (October 1978). It is an indication of how anxious people were to find the fifth quark that the discoverers of the upsilon jumped the gun [D. C. Hom et al., *Phys. Rev. Lett.* **36,** 1236 (1976)], announcing a spurious particle now known fondly as the "oops-Leon" (after Leon Lederman, head of the group).
39. M. Basile et al., *Nuovo Cimento Lett.* **31,** 97 (1981).

40. D. Drijard et al., *Phys. Lett.* **108B,** 361 (1982).
41. S. Behrends et al., *Phys. Rev. Lett.* **50,** 881 (1983).
42. The formula for the W and Z masses was first obtained by S. Weinberg, *Phys. Rev. Lett.* **19,** 1264, (1967). It involved a parameter θ_W whose value was unknown at that time, and all Weinberg could say for sure was that $M_W \geqslant 37$ GeV/c^2 and $M_Z \geqslant 75$ GeV/c^2. In the next 15 years θ_W was measured in a variety of experiments, and by 1982 the predictions had been refined, as indicated in equation (1.30).
43. G. Arnison et al., *Phys. Lett.* **122B,** 103 (1983).
44. G. Arnison et al., *Phys. Lett.* **126B,** 398 (1983).
45. See D. B. Cline and C. Rubbia, *Phys. Today* (August 1980) page 44, and D. B. Cline, C. Rubbia, and S. van der Meer, *Sci. Am.* (March 1982). Already two books have been written about the discovery of the W and the Z: Christine Sutton, *The Particle Connection* (New York: Simon & Schuster, 1984), and Peter Watkins, *Story of the W and Z* (Cambridge: Cambridge University Press, 1986).
46. See M. Jacob and P. Landshoff, ref. 26.
47. K. Ishikawa, *Sci. Am.* (November 1982).
48. See, for instance, the review article by H. Terezawa in XXII International Conference on High-Energy Physics, Leipzig, Vol. I, (1984), p. 63.

PROBLEMS

1.1. If a charged particle is undeflected in passing through uniform crossed electric and magnetic fields **E** and **B** (mutually perpendicular, and both perpendicular to the direction of motion), what is its velocity? If we now turn off the electric field, and the particle moves in an arc of radius R, what is its charge-to-mass ratio?

1.2. The mass of Yukawa's meson can be estimated as follows. When two protons in a nucleus exchange a meson (mass m) they must temporarily violate the conservation of energy by an amount mc^2 (the rest energy of the meson). The Heisenberg uncertainty principle says that you may "borrow" an energy ΔE, provided you "pay it back" in a time Δt given by $\Delta E\, \Delta t = \hbar$ (where $\hbar \equiv h/2\pi$). In this case we need to borrow $\Delta E = mc^2$ long enough for the meson to make it from one proton to the other. It has to cross the nucleus (size r_0), and it travels, presumably, at some substantial fraction of the speed of light, so, roughly speaking, $\Delta t = r_0/c$. Putting this all together, we have

$$m = \frac{\hbar}{r_0 c}$$

Using $r_0 = 10^{-13}$ cm (the size of a typical nucleus), calculate the mass of Yukawa's meson. Express your answer as a multiple of the electron's mass, and compare the observed mass of the pion. [If you find that argument compelling, I can only say that you're pretty gullible. Try it for an *atom*, and you'll conclude that the mass of the photon is about 7×10^{-30} g, which is nonsense. Nevertheless, it is a useful device for "back-of-the-envelope" calculations, and it does very well for the pi meson. Unfortunately, many books present it as though it were a rigorous derivation, which it certainly is *not*. The uncertainty principle does *not* license violation of conservation of energy (nor does any such violation occur in this process; we shall see later on how this comes about). Moreover, it's an *inequality*, $\Delta E\, \Delta t \geqslant \hbar$, which at most could give you a *lower bound* on m. It *is* typically true that the *range* of a force is inversely proportional to the mass of the mediator, but the size of a bound state is not always a good measure of the range (that's why the argument fails for the photon: The range of the electromagnetic force is infinite, but the size of an atom is not). In general, when you hear a physicist invoke the uncertainty principle, keep a hand on your wallet.]

1.3. In the period before the discovery of the neutron many people thought the nucleus consisted of protons and *electrons,* with the atomic number equal to the excess number of protons. Beta decay seemed to support this idea—after all, electrons come popping out; doesn't that imply that there were electrons inside? Use the position-momentum uncertainty relation, $\Delta x\, \Delta p \geqslant \hbar$, to estimate the minimum momentum of an electron confined to a nucleus (radius 10^{-13} cm). From the relativistic energy-momentum relation, $E^2 - \mathbf{p}^2c^2 = m^2c^4$, determine the corresponding energy, and compare it with that of an electron emitted in, say, the beta decay of tritium (Fig. 1.6). (This result convinced some people that the beta-decay electron could *not* have been rattling around inside the nucleus, but must be produced in the disintegration itself.)

1.4. The *Gell-Mann/Okubo mass formula* relates the masses of members of the baryon octet (ignoring small differences between p and n; Σ^+, Σ^0, and Σ^-; and Ξ^0 and Ξ^-):

$$2(m_N + m_\Xi) = 3m_\Lambda + m_\Sigma$$

Using this formula, together with the known masses of the *nucleon* N (use the average of p and n), Σ (again, use the average), and Ξ (ditto), "predict" the mass of the Λ. How close do you come to the observed value?

1.5. The same formula applies to the mesons (with $\Sigma \to \pi$, $\Lambda \to \eta$, etc.); only, for reasons that remain something of a mystery, in this case you must use the *squares* of the masses. Use this to "predict" the mass of the η. How close do you come?

1.6. The mass formula for decuplets is much simpler—equal spacing between the rows:

$$M_\Delta - M_{\Sigma^*} = M_{\Sigma^*} - M_{\Xi^*} = M_{\Xi^*} - M_\Omega$$

Use this formula (as Gell-Mann did) to predict the mass of the Ω^-. (Use the average of the first two spacings to estimate the third.) How close is your prediction to the observed value?

1.7. **(a)** Members of the baryon decuplet typically decay after 10^{-23} sec into a lighter baryon (from the baryon octet) and a meson (from the pseudo-scalar meson octet). Thus, for example, $\Delta^{++} \to p^+ + \pi^+$. List all decay modes of this form for the Δ^-, Σ^{*+}, and Ξ^{*-}. Remember that these decays must conserve charge and strangeness (they are *strong* interactions).

(b) In any decay, there must be sufficient mass in the original particle to cover the masses of the decay products. (There may be *more* than enough; the extra will be "soaked up" in the form of kinetic energy in the final state.) Check each of the decays you proposed in part (a) to see which ones meet this criterion. The others are kinematically forbidden.

1.8. **(a)** Analyze the possible decay modes of the Ω^-, just as you did in Problem 1.7 for the Δ, Σ^*, and Ξ^*. See the problem? Gell-Mann predicted that the Ω^- would be "metastable" (i.e., much longer lived than the other members of the decuplet), for precisely this reason. (The Ω^- *does* in fact decay, but by the much slower *weak* interaction, which does not conserve strangeness.)

(b) From the bubble chamber photograph (Fig. 1.11, measure the length of the Ω^- track, and use this to estimate the lifetime of the Ω^-. (Of course, you don't know how fast it was going, but it's a safe bet that the speed was less than the velocity of light; let's say it was going about $0.1c$. Also, you don't know if the reproduction has enlarged or shrunk the scale, but never mind: this is quibbling over factors of 2, or 5, or maybe even 10. The important point is that the lifetime is many orders of magnitude longer than the 10^{-23} sec characteristic of all other members of the decuplet).

1.9. Check the *Coleman-Glashow relation* [*Phys. Rev.* **B134**, 671 (1964)]:

$$\Sigma^+ - \Sigma^- = p - n + \Xi^0 - \Xi^-$$

(the particle names stand for their masses).

1.10. Look up the table of "known" mesons compiled by M. Roos in *Rev. Mod. Phys.* **35**, 314 (1963), and compare the current Particle Data Booklet[20] to determine which of the 1963 mesons have stood the test of time. (Some of the names have been changed, so you will have to work from other properties, such as mass, charge, strangeness, etc.)

1.11. Of the spurious particles you identified in Problem 1.10, which are "exotic" (i.e., inconsistent with the quark model)? How many of the surviving mesons are exotic?

1.12. How many different *meson* combinations can you make with 1, 2, 3, 4, 5, or 6 different quark flavors? What's the general formula for n flavors?

1.13. How many different *baryon* combinations can you make with 1, 2, 3, 4, 5, or 6 different quark flavors? What's the general formula for n flavors?

1.14. Using four quarks (u, d, s, and c), construct a table of all the possible baryon species. How many combinations carry a charm of +1? How many carry charm +2, and +3?

1.15. Same as Problem 1.14, but this time for *mesons*.

1.16. De Rujula, Georgi, and Glashow [*Phys. Rev.* **D12**, 147 (1975)] estimated the quark masses to be: $m_u = m_d = 336$ MeV/c^2, $m_s = 540$ MeV/c^2, and $m_c = 1500$ MeV/c^2 (the bottom quark is about 4500 MeV/c^2). If they are right, the average binding energy for members of the baryon octet is -62 MeV. If they all had *exactly* this binding energy, what would their masses be? Compare the *actual* values, and give the percent error. (Don't try this on the other supermultiplets, however. There really is no reason to suppose the binding energy is the same for all members of the group. The problem of hadron masses is a thorny issue, to which we shall return in Chapter 5.)

1.17. M. Shupe [*Phys. Lett.* **86B**, 87 (1979)] has proposed that all quarks and leptons are composed of two even more elementary constituents: c (with charge $-1/3$) and n (with charge zero)—and their respective antiparticles, $\bar{c}$ and $\bar{n}$. You're allowed to combine them in groups of three particles or three antiparticles (ccn, for example, or $\bar{n}\bar{n}\bar{n}$). Construct all of the eight quarks and leptons in the first generation in this manner. (The other generations are supposed to be excited states.) Notice that each of the *quark* states admits three possible permutations (ccn, cnc, ncc, for example)—these correspond to the three colors. Mediators can be constructed from three particles plus three antiparticles. $W^{\pm}$, Z^0, and γ involve three *like* particles and three like antiparticles ($W^- = ccc\bar{n}\bar{n}\bar{n}$, for instance). Construct W^+, Z^0, and γ in this way. Gluons involve mixed combinations ($ccn\bar{c}\bar{n}\bar{n}$, for instance). How many possibilities are there in all? Can you think of a way to reduce this down to eight?

Chapter 2

Elementary Particle Dynamics

*This chapter introduces the fundamental forces by which elementary particles interact, and the **Feynman diagrams** we use to represent these interactions. The treatment is entirely qualitative and can be read quickly to get a sense of the "lay of the land." The quantitative details will come in Chapters 6 through 10.*

2.1 THE FOUR FORCES

As far as we know, there are just four fundamental forces in nature: *strong, electromagnetic, weak,* and *gravitational.* They are listed in the following table in order of decreasing strength:*

Force	Strength	Theory	Mediator
Strong	10	Chromodynamics	Gluon
Electromagnetic	10^{-2}	Electrodynamics	Photon
Weak	10^{-13}	Flavordynamics	W and Z
Gravitational	10^{-42}	Geometrodynamics	Graviton

To each of these forces there belongs a physical theory. The classical theory of gravity is, of course, Newton's law of universal gravitation. Its relativistic generalization is Einstein's general theory of relativity ("geometrodynamics" would be a better term). A completely satisfactory quantum theory of gravity has yet to be worked out; for the moment, most people assume that gravity is simply

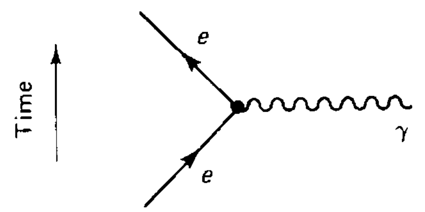
* The "strength" of a force is an intrinsically ambiguous notion—after all, it depends on the nature of the source and on how far away you are. So the numbers in this table should not be taken too literally, and (especially in the case of the weak force) you will see quite different figures quoted elsewhere.

too weak to play a significant role in elementary particle physics. The physical theory that describes electromagnetic forces is called *electrodynamics*. It was given its classical formulation by Maxwell over one hundred years ago. Maxwell's theory was already consistent with special relativity (for which it was, in fact, the main inspiration). The quantum theory of electrodynamics was perfected by Tomonaga, Feynman, and Schwinger in the 1940s. The weak forces, which account for nuclear beta decay (and also, as we have seen, the decay of the pion, the muon, and many of the strange particles) were unknown to classical physics; their theoretical description was given a relativistic quantum formulation right from the start. The first theory of the weak forces was presented by Fermi in 1933; it was refined by Lee and Yang, Feynman and Gell-Mann, and many others, in the fifties, and put into its present form by Glashow, Weinberg, and Salam, in the sixties. For reasons that will appear in due course, the theory of weak interactions is sometimes called *flavordynamics*;[1] in this book I refer to it simply as the Glashow–Weinberg–Salam (GWS) theory. (The GWS model treats weak and electromagnetic interactions as different manifestations of a single *electroweak* force, and in this sense the four forces reduce to three.) As for the strong forces, beyond the pioneering work of Yukawa in 1934 there really *was* no theory until the emergence of chromodynamics in the mid-seventies.

Each of these forces is mediated by the exchange of a particle. The gravitational mediator is called the *graviton*, electromagnetic forces are mediated by the *photon*, strong forces by the *gluon*, and weak forces by the *intermediate vector bosons*, W and Z. These mediators transmit the force between one quark or lepton and another. In principle, the force of impact between a bat and a baseball is nothing but the combined interaction of the quarks and leptons in one with the quarks and leptons in the other. More to the point, the strong force between two protons, say, which Yukawa took to be a fundamental and irreducible process, must be regarded as a complicated interaction of six quarks. This is clearly not the place to look for simplicity. Rather, we must begin by analyzing the force between one truly elementary particle and another. In this chapter I will show you qualitatively how each of the relevant forces acts on individual quarks and leptons. Subsequent chapters develop the machinery needed to make the theory quantitative.

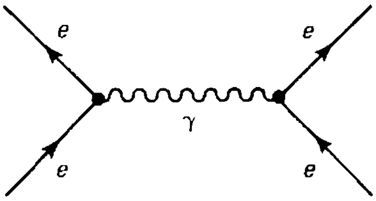
2.2 QUANTUM ELECTRODYNAMICS (QED)

Quantum electrodynamics is the oldest, the simplest, and the most successful of the dynamical theories; the others are self-consciously modeled on it. So I'll begin with a description of QED. *All electromagnetic phenomena are ultimately reducible to the following elementary process:*



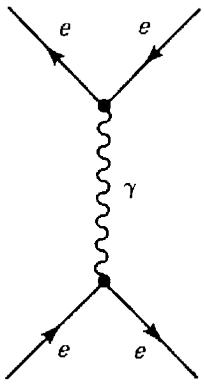
This diagram reads: Charged particle e enters, emits (or absorbs) a photon, γ, and exits.* For the sake of argument, I'll assume the charged particle is an electron; it could just as well be a quark, or any lepton except a neutrino (the latter is neutral, of course, and does not experience an electromagnetic force).

To describe more complicated processes, we simply patch together two or more replicas of this *primitive vertex*. Consider, for example, the following:



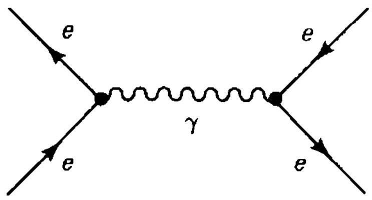
Here, two electrons enter, a photon passes between them (I need not say which one emits the photon and which one absorbs it; the diagram represents *both* orderings), and the two then exit. This diagram, then, describes the interaction between two electrons; in the classical theory we would call it the Coulomb repulsion of like charges (if the two are at rest). In QED this process is called *Møller scattering;* we say that the interaction is "mediated by the exchange of a photon," for reasons that should now be apparent.

Now, you're allowed to twist these "Feynman diagrams" around into any topological configuration you like—for example, we could stand the previous picture on its side:



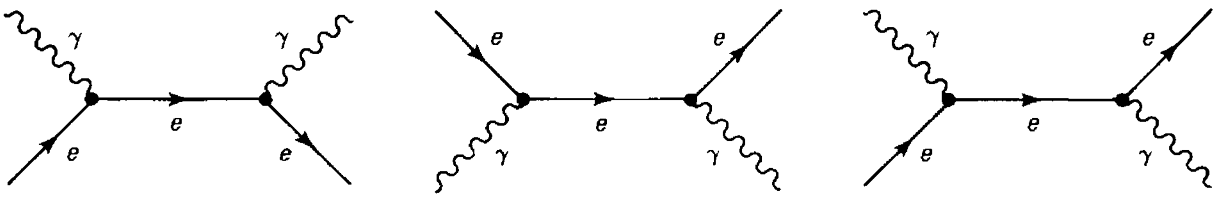
The rule of the game is that a particle line running "backward in time" (as indicated by the arrow) is to be interpreted as the corresponding *antiparticle* going *forward* (the photon is its own antiparticle, that's why I didn't need an arrow on the photon line). So in this process an electron and a positron annihilate to form a photon, which in turn produces a new electron-positron pair. An electron and a positron went *in,* an electron and a positron came *out* (not the *same* ones, but then, since all electrons are identical, it hardly matters). This represents the interaction of two opposite charges: their Coulomb attraction. In QED this process is called *Bhabha scattering.* There is a quite different diagram which also contributes:

* In this book time always flows *upward;* the traditional convention. Particle physicists tend increasingly to let t run *horizontally* (to the right), but there is no established consensus on the matter.

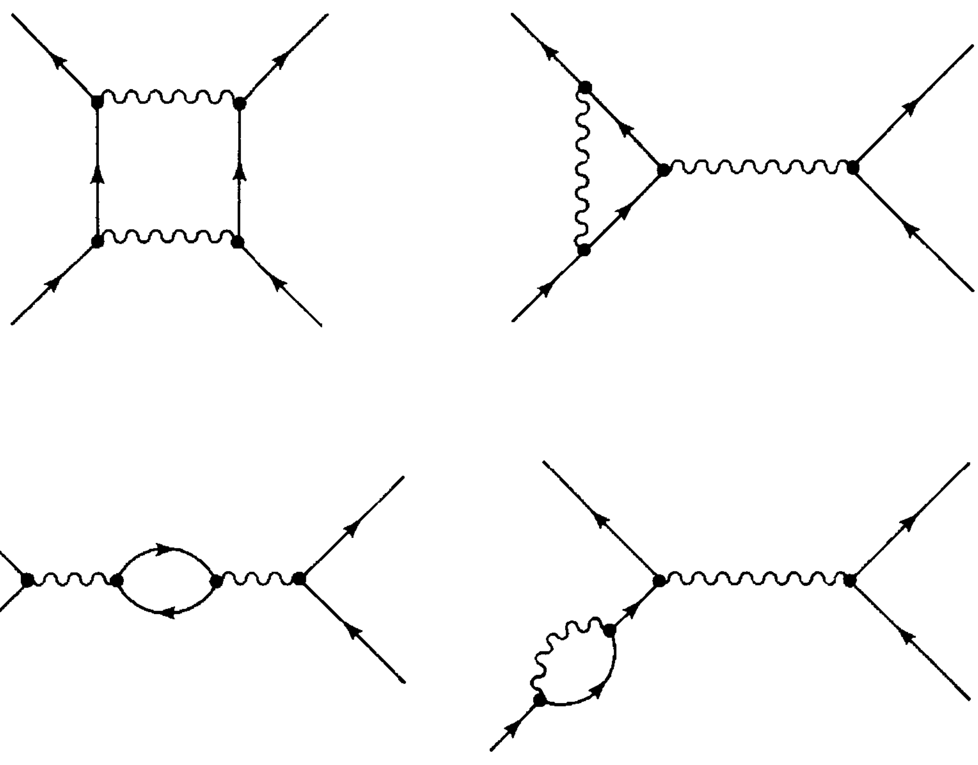


Both diagrams must be included in the analysis of Bhabha scattering.

Using just two vertices we can also construct the following diagrams, describing, respectively, pair annihilation, $e^- + e^+ \rightarrow \gamma + \gamma$; pair production, $\gamma + \gamma \rightarrow e^- + e^+$; and Compton scattering, $e^- + \gamma \rightarrow e^- + \gamma$:



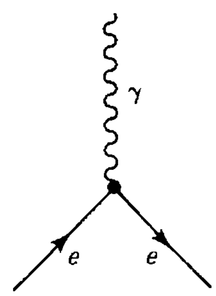
[Notice that Bhabha and Møller scattering are related by crossing symmetry (see Section 1.4); as are the three processes shown here. In terms of Feynman diagrams, crossing symmetry corresponds to twisting or rotating the figure.] If we allow more vertices, the possibilities rapidly proliferate; for example, with four vertices we obtain, among others, the following diagrams:



In each of these figures two electrons went in and two electrons came out. They too describe the repulsion of like charges (Møller scattering). The "innards" of the diagram are irrelevant as far as the observed process is concerned. Internal lines (those which begin and end within the diagram) represent particles that are not observed—indeed, that *cannot* be observed without entirely changing the process. We call them "virtual" particles. Only the *external* lines (those which enter or leave the diagram) represent "real" (observable) particles. The external lines, then, tell you what physical process is occurring; the internal lines describe the *mechanism* involved.

Please understand: these Feynman diagrams are purely symbolic; they do *not* represent particle trajectories (as you might see them in, say, a bubble chamber photograph). The vertical dimension is *time,* and horizontal spacings do *not* correspond to physical separations. For instance, in Bhabha scattering the electron and positron are *attracted,* not repelled (as the diverging lines might seem to suggest). All the diagram says is: "Once there was an electron and a positron; they exchanged a photon; then there was an electron and a positron again." Each Feynman diagram actually stands for a particular *number,* which can be calculated using the so-called *Feynman rules* (you'll learn how to do this in Chapter 6). Suppose you want to analyze a certain physical process (say, Møller scattering). First you draw all the diagrams that have the appropriate external lines (the one with two vertices, all the ones with four vertices, and so on), then you evaluate the contribution of each diagram, using the Feynman rules, and add it all up. The *sum total* of all Feynman diagrams with the given external lines represents the actual physical process. Of course, there's a problem here: there are infinitely many Feynman diagrams for any particular reaction! Fortunately, each vertex within a diagram introduces a factor of $\alpha = (e^2/\hbar c) = \frac{1}{137}$, the *fine structure constant.* Because this is such a small number, diagrams with more and more vertices contribute less and less to the final result, and, depending on the accuracy you need, may be ignored. In fact, in QED it is rare to see a calculation that includes diagrams with more than four vertices. The answers are only approximate, to be sure, but when the approximation is valid to six significant digits, only the most fastidious are likely to complain.

The Feynman rules enforce conservation of energy and momentum at each vertex, and hence for the diagram as a whole. It follows that the primitive QED vertex *by itself* does not represent a possible physical process. We can draw the diagram, but calculation would assign to it the number *zero.* The reason is purely kinematical: $e^- \to e^- + \gamma$ would violate conservation of energy. (In the center-of-mass frame the electron is initially at rest, so its energy is mc^2. It cannot decay into a photon plus a recoiling electron because the latter alone would require an energy greater than mc^2.) Nor, for instance, is $e^- + e^+ \to \gamma$ kinematically possible, although it is easy enough to draw the diagram:

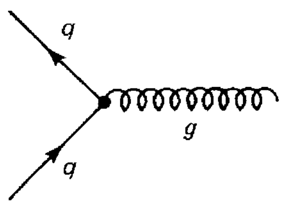


In the center-of-mass system the electron and positron enter symmetrically with equal and opposite velocities, so the total momentum before the collision is obviously zero. But the *final* momentum *cannot* be zero, since photons always travel at the speed of light; an electron-positron pair can annihilate to make *two* photons, but not *one*. *Within a larger diagram*, however, these figures are perfectly acceptable, because, although energy and momentum must be conserved at each vertex, *a virtual particle does not carry the same mass* as the corresponding free particle. In fact, a virtual particle can have *any* mass—whatever the conservation laws require.* In the business, we say that virtual particles do not lie on their *mass shell*. *External* lines, by contrast, represent *real* particles, and these do carry the "correct" mass.

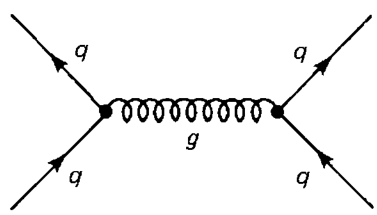
[Actually, the *physical* distinction between real and virtual particles is not quite as sharp as I have implied. If a photon is emitted on Alpha Centauri, and absorbed in your eye, it is technically a virtual photon, I suppose. However, in general, the farther a virtual particle is from its mass shell the shorter it lives, so a photon from a distant star would have to be extremely close to its "correct" mass; it would have to be *very close* to "real." As a calculational matter, you would get essentially the same answer if you treated the process as two separate events (emission of a real photon by star, followed by absorption of a real photon by eye). You might say that a real particle is a virtual particle which lasts long enough that we don't care to inquire how it was produced, or how it is eventually absorbed.]

2.3 QUANTUM CHROMODYNAMICS (QCD)

In chromodynamics *color* plays the role of charge, and the fundamental process (analogous to $e^- \to e^- + \gamma$) is quark $\to$ quark-plus-gluon (since leptons do not carry color, they do not participate in the strong interactions):



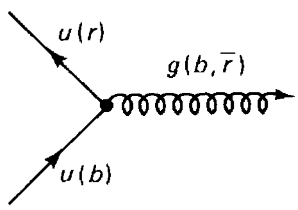
As before, we combine two or more such "primitive vertices" to represent more complicated processes. For example, the force between two quarks (which is responsible in the first instance for binding quarks together to make baryons, and indirectly for holding the neutrons and protons together to form a nucleus) is described in lowest order by the diagram:



* In special relativity, the energy E, momentum, $\mathbf{p}$, and mass m of a free particle are related by the equation $E^2 - \mathbf{p}^2c^2 = m^2c^4$. But for a *virtual* particle $E^2 - \mathbf{p}^2c^2$ can take on *any* value. Many authors interpret this to mean that virtual processes violate conservation of energy (see Problem 1.2). Personally, I consider this misleading, at best. Energy is *always* conserved.

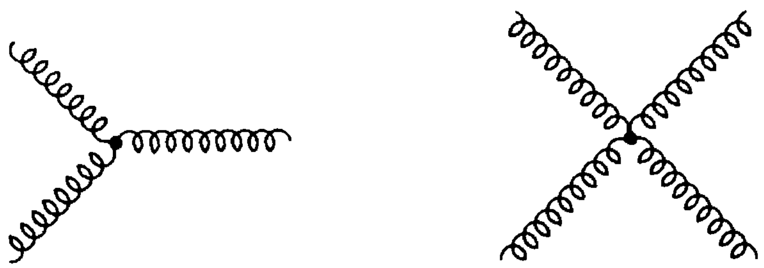
We say that the force between two quarks is "mediated" by the exchange of gluons.

At this level chromodynamics is very similar to electrodynamics. However, there are also important differences, most conspicuously, the fact that whereas there is only one kind of electric charge (it can be positive or negative, to be sure, but a *single number* suffices to characterize the charge of a particle), there are *three* kinds of color (red, green, and blue). In the process $q \rightarrow q + g$, the color of the quark (but not its flavor) may change. For example, a blue up-quark may convert into a red up-quark. Since color (like charge) is always conserved, this means that the gluon must carry away the difference—in this instance, one unit of blueness and *minus* one unit of redness:



Gluons, then, are "bicolored," carrying one positive unit of color and one negative unit. There are evidently $3 \times 3 = 9$ possibilities here, and you might expect there to be 9 kinds of gluons. For technical reasons, which we'll come to in Chapter 9, there are actually only 8.

Since the gluons *themselves* carry color (unlike the photon, which is electrically neutral), they couple directly to other gluons, and hence in addition to the fundamental quark-gluon vertex, we also have primitive gluon-gluon vertices, in fact, two kinds: three gluon vertices and four gluon vertices:



This direct gluon–gluon coupling makes chromodynamics a lot more complicated than electrodynamics, but also far richer, allowing, for instance, the possibility of *glueballs* (bound states of interacting gluons, with no quarks on the scene at all).

Another difference between chromodynamics and electrodynamics is the size of the *coupling constant.* Remember that each vertex in QED introduces a factor of $\alpha = \frac{1}{137}$, and the smallness of this number means that we need only consider Feynman diagrams with a small number of vertices. Experimentally, the corresponding coupling constant for the strong forces, α_s, as determined, say, from the force between two protons, is greater than 1, and the *bigness* of this number plagued particle physics for decades. For instead of contributing less and less, the more complex diagrams contribute *more and more,* and Feynman's procedure, which worked so well in QED, is apparently worthless. One

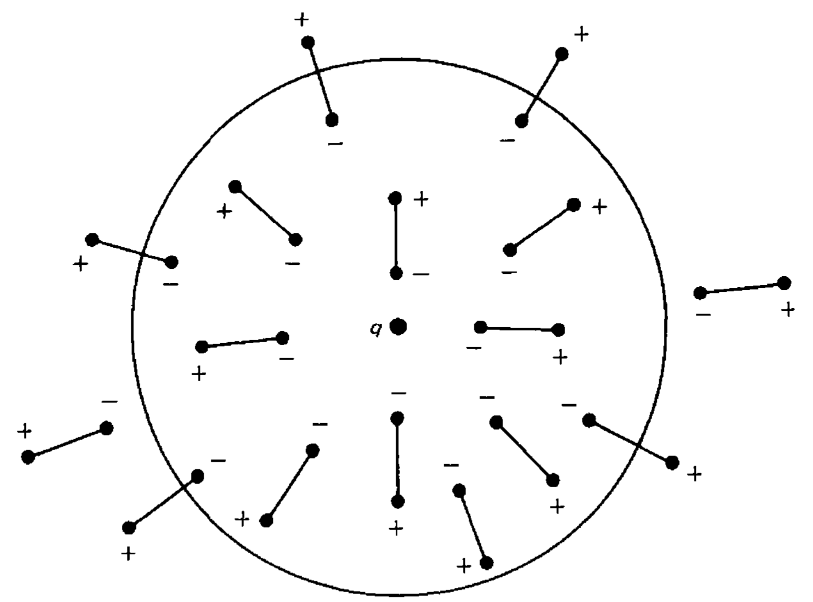


Figure 2.1 Screening of a charge q by a dielectric medium.

of the great triumphs of quantum chromodynamics (QCD) was the discovery that in this theory the number that plays the role of coupling "constant" is in fact not constant at all, but depends on the separation distance between the interacting particles (we call it a "running" coupling constant). Although at the relatively *large* distances characteristic of nuclear physics it is big at very short distances (less than the size of a proton) it becomes quite small. This phenomenon is known as *asymptotic freedom;* it means that within a proton or a pion, say, the quarks rattle around without interacting much. Just such behavior was found experimentally in the deep inelastic scattering experiments. From a theoretical point of view, the discovery of asymptotic freedom rescued the Feynman calculus as a legitimate tool for QCD, in the high-energy regime.

Even in electrodynamics, the effective coupling depends somewhat on how far you are from the source. This can be understood qualitatively as follows. Picture first a positive point charge q embedded in a dielectric medium (i.e., a substance whose molecules become polarized in the presence of an electric field). The negative end of each molecular dipole will be attracted toward q, and the positive end repelled away, as shown in Figure 2.1. As a result, the particle acquires a "halo" of negative charge, which partially cancels its field. In the

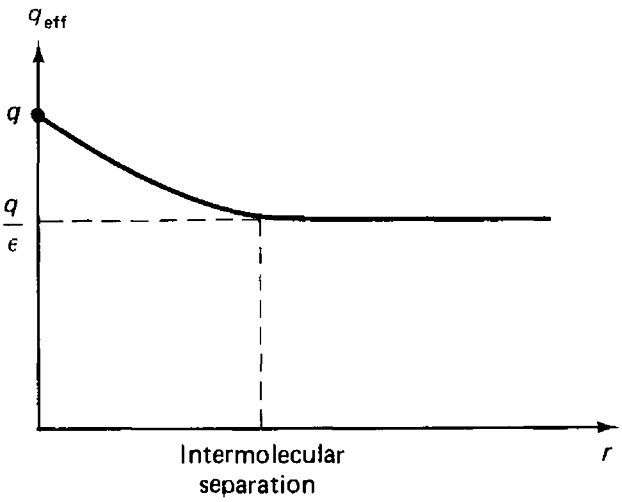


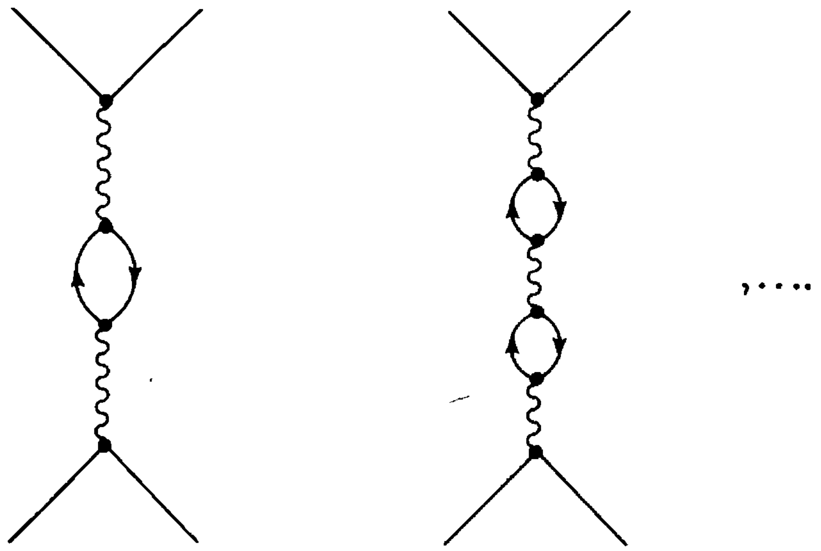
Figure 2.2 Effective charge as a function of distance.

presence of the dielectric, then, the *effective* charge of any particle is somewhat reduced:

$$q_{\text{eff}} = q/\epsilon \tag{2.1}$$

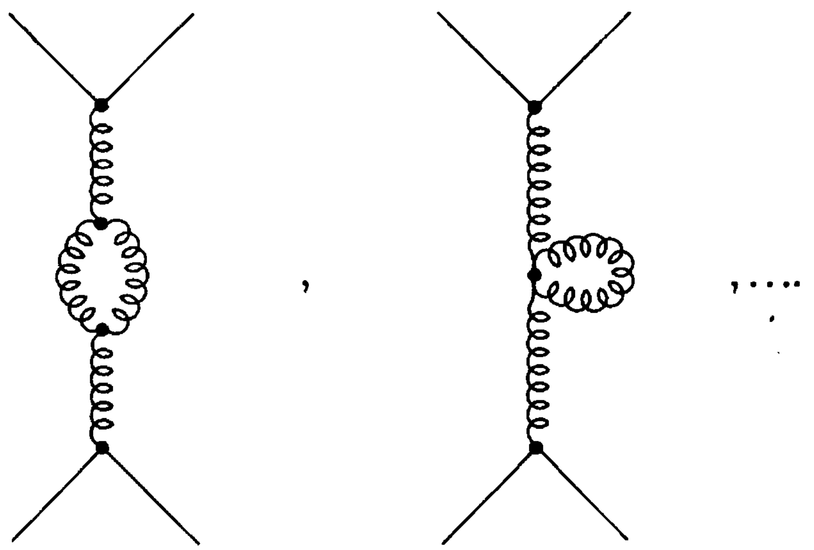
(The factor ϵ by which the field is reduced is called the *dielectric constant* of the material; it is a measure of the ease with which the substance can be polarized.[2]) Of course, if you are in closer than the nearest molecule, then there is no such screening, and you "see" the full charge q. Thus if you were to make a graph of the effective charge, as a function of distance, it would look something like Figure 2.2. The effective charge *increases* at very small distances.

Now, it so happens that in quantum electrodynamics the vacuum itself behaves like a dielectric; it sprouts positron-electron pairs, as shown in Feynman diagrams such as these:



The virtual electron in each "bubble" is attracted toward q, and the virtual positron is repelled away; the resulting *vacuum polarization* partially screens the charge and reduces its field. Once again, however, if you get *too* close to q, the screening disappears. What plays the role of the "intermolecular spacing" in this case is the Compton wavelength of the electron, $\lambda_c = h/mc = 2.43 \times 10^{-10}$ cm. For distances smaller than this the effective charge increases, just as it did in Figure 2.2. Notice that the *unscreened* ("close-up") charge, which you might regard as the "true" charge of the particle, is *not* what we measure in any ordinary experiment, since we are seldom working at such minute separation distances. [An exception is the Lamb shift—a tiny perturbation in the spectrum of hydrogen—in which the influence of vacuum polarization (or rather, its *absence* at short distances) is clearly discernible.] What we have always called "the charge of the electron" is actually the fully screened *effective* charge.

So much for electrodynamics. The same thing happens in QCD, but with an important added ingredient. Not only do we have the quark-quark-gluon vertex (which, by itself, would again lead to an increasing coupling strength at short distances), but now there are *also* the direct gluon-gluon vertices. So in addition to the diagrams analogous to vacuum polarization in QED, we must now also include gluon loops, such as these:

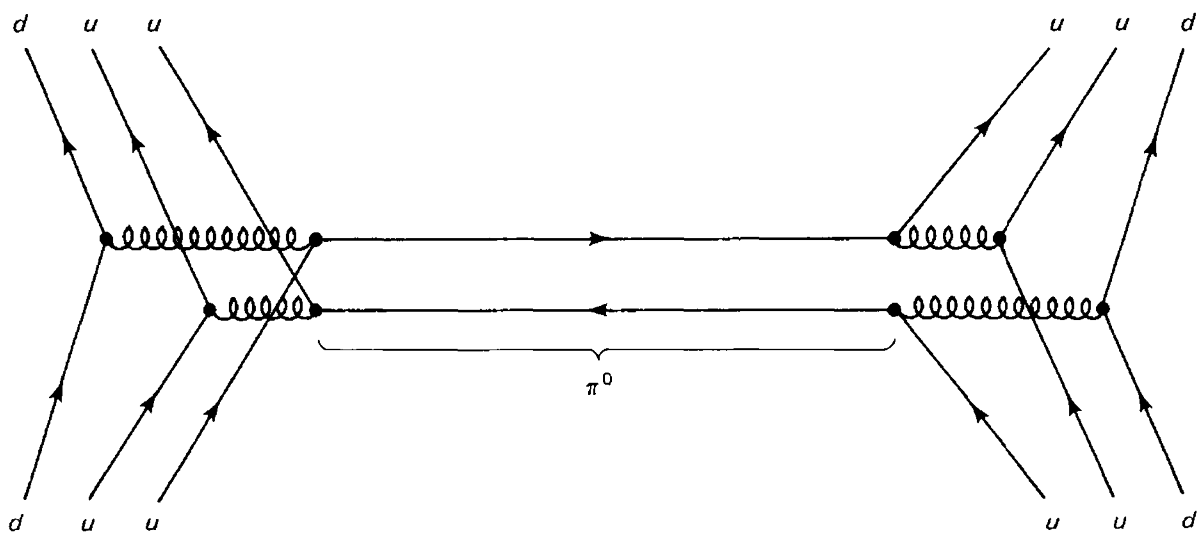


It is not clear a priori what influence these diagrams will have on the story;[3] as it turns out, their effect is the *opposite:* There occurs a kind of competition between the quark polarization diagrams (which drive α_s *up* at short distances) and gluon polarization (which drives it *down*). Since the former depends on the number of quarks in the theory (hence on the number of *flavors*, f), whereas the latter depends on the number of gluons (hence on the number of *colors*, n), the winner in the competition depends on the relative number of flavors and colors. The critical parameter turns out to be

$$a \equiv 2f - 11n \tag{2.2}$$

If this number is *positive,* then, as in QED, the effective coupling *increases* at short distances; if it is *negative,* the coupling *decreases.* In the Standard Model, $f = 6$ and $n = 3$, so $a = -21$, and the QCD coupling decreases at short distances. Qualitatively, this is the origin of asymptotic freedom.

The final distinction between QED and QCD is that whereas many particles carry electric charge, no naturally occurring particles carry color. Experimentally, it seems that quarks are confined in colorless packages of two (mesons) and three (baryons). As a consequence, the processes we actually observe in the laboratory are necessarily indirect and complicated manifestations of chromodynamics. It is as though our only access to electrodynamics came from the van der Waals forces between neutral molecules. For example, the (strong) force between two protons involves (among many others) the following diagram:



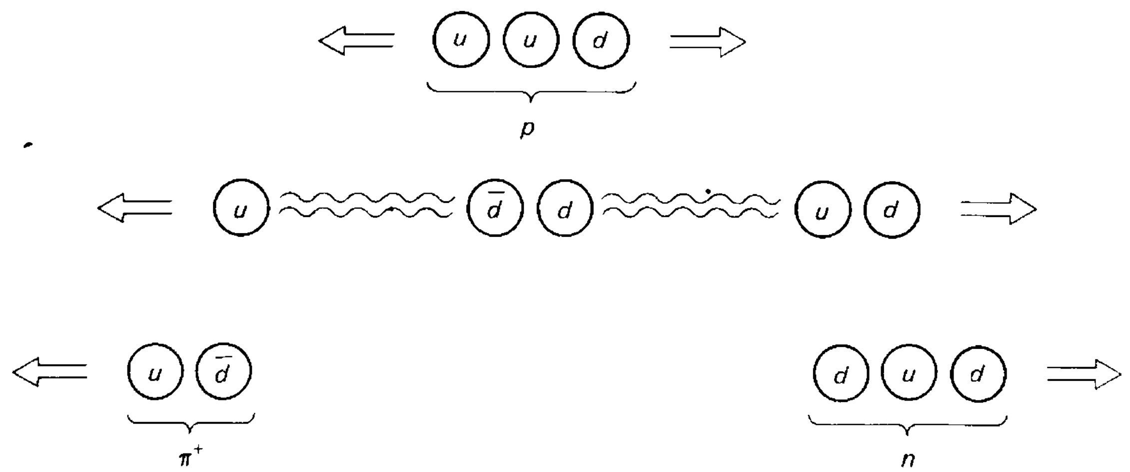


Figure 2.3 A possible scenario for quark confinement: As we pull a u quark out of the proton a pair of quarks is created, and instead of a free quark, we are left with a pion and a neutron.

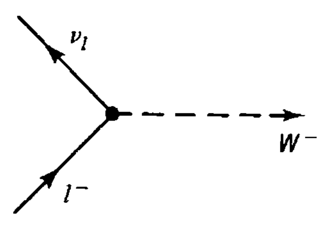
You will recognize here the remnants of Yukawa's pion-exchange model, but the entire process is enormously more complex that Yukawa ever imagined. If QCD is correct, it must contain the explanation for quark confinement; that is, it must be possible to *prove,* as a consequence of this theory, that quarks can only exist in the form of colorless combinations. Presumably this proof will take the form of a demonstration that the potential energy increases without limit as the quarks are pulled farther and farther apart, so that it would take an infinite energy (or at any rate, enough to create new quark–antiquark pairs) to separate them completely (see Fig. 2.3). So far, no one has provided a conclusive proof that QCD implies confinement (see, however, ref. 25 in Chapter 1). The difficulty is that confinement involves the *long-range* behavior of the quark-quark interaction, but this is precisely the regime in which the Feynman calculus fails.

2.4 WEAK INTERACTIONS[4]

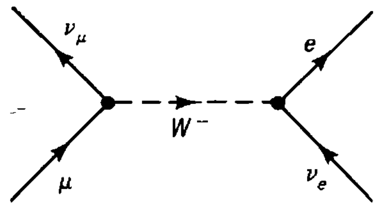
There is no particular name for the "stuff" that produces weak forces, in the sense that electric charge produces electromagnetic forces and color produces strong forces. Some people call it "weak charge." Whatever word you use, all quarks and all leptons carry it. (Leptons have no color, so they do not participate in the strong interactions; neutrinos have no charge, so they experience no electromagnetic forces; but *all* of them join in the weak interactions.) There are two kinds of weak interactions: *charged* (mediated by the W's) and *neutral* (mediated by the Z). The theory is cleaner for leptons than it is for quarks, so let us begin with the leptons.

2.4.1 Leptons

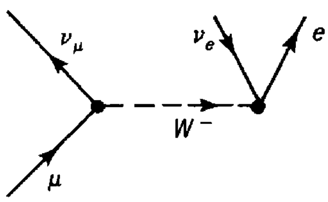
The fundamental charged vertex looks like this:



A negative lepton (it could be e^-, μ^-, or τ^-) converts into the corresponding neutrino, with emission of a W^- (or absorption of a W^+): $l^- \rightarrow \nu_l + W^-$.* As always, we combine the primitive vertices together to generate more complicated reactions. For example, the process $\mu^- + \nu_e \rightarrow e^- + \nu_\mu$ would be represented by the diagram:

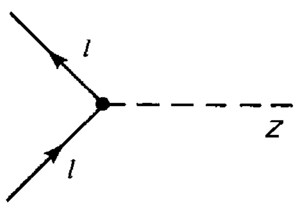


Such a neutrino-muon scattering event would be hard to set up in the laboratory, but with a slight twist essentially the same diagram describes the decay of the muon, $\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e$, which happens all the time:

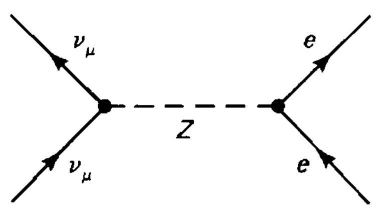


(Technically, this is only the lowest-order contribution to muon decay, but in weak interaction theory one almost never needs to consider higher-order corrections.)

The fundamental neutral vertex is:



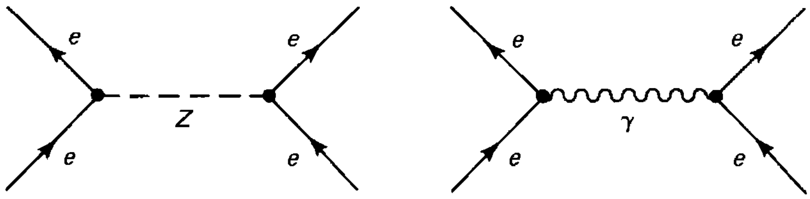
In this case l can be *any* lepton (including neutrinos). The Z mediates such processes as neutrino-electron scattering ($\nu_\mu + e^- \rightarrow \nu_\mu + e^-$):



Although charged weak processes were recognized from the start (beta decay itself is a charged process), the theoretical possibility of neutral weak processes was not appreciated until 1958. The Glashow–Weinberg–Salam (GWS) model

* This implies, of course that $l^+ \rightarrow \bar{\nu}_l + W^+$ is also an allowed vertex.

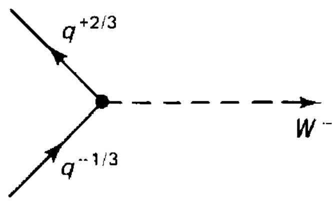
includes neutral weak processes as essential ingredients, and their existence was confirmed experimentally at CERN in 1973.[5] The reason it took so long for neutral weak processes to be discovered is twofold: (1) nobody was looking for them and (2) they tend to be masked by much stronger electromagnetic effects. For example, the Z can be exchanged between two electrons, but so can the photon:



Presumably there is a minute correction to Coulomb's law that's attributable to the first diagram, but the photon-mediated process overwhelmingly dominates. Experiments at DESY (in Hamburg) studied the reaction $e^- + e^+ \to \mu^- + \mu^+$ at very high energy and found unmistakable evidence of a contribution from the Z.[6] But to observe a *pure* neutral weak interaction one has to go to neutrino scattering, in which there is no competing electromagnetic mechanism, and neutrino experiments are notoriously difficult.

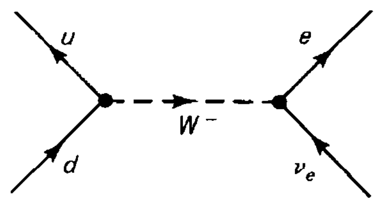
2.4.2 Quarks

Notice that the leptonic weak vertices connect members of the *same generation:* e^- converts to ν_e (with emission of W^-), or $\mu^- \to \mu^-$ (emitting a Z), but e^- never goes to μ^- nor μ^- to ν_e. In this way the theory enforces the conservation of electron number, muon number, and tau number. It is tempting to suppose that the same rule applies to the quarks, so that the fundamental charged vertex is:

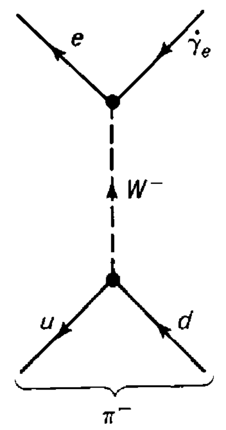


A quark with charge $-\frac{1}{3}$ (which is to say: d, s, or b) converts into the corresponding quark with charge $+\frac{2}{3}$ (u, c, or t, respectively), with the emission of a W^-. The outgoing quark carries the same color as the ingoing one, but a different flavor. It's not that the W carries off the "missing" flavor—after all, the W must be capable of coupling to leptons, which *have* no flavor; rather, *flavor is simply not conserved in weak interactions.* (Because quark flavor typically changes at a weak vertex, as quark color changes at a strong vertex, weak interaction theory is sometimes called "flavordynamics.")

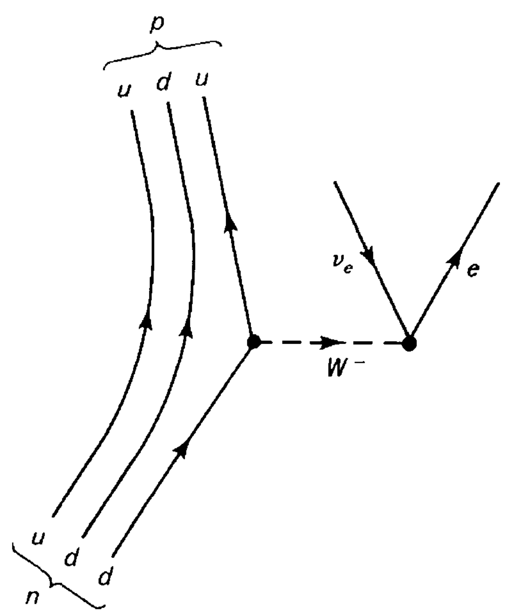
The far end of the W line can couple to leptons (a "semileptonic" process), or to other quarks (a purely hadronic process). The most important semileptonic process is undoubtedly $d + \nu_e \to u + e$:



Because of quark confinement, this process would never occur in nature as it stands. However, turned on its side, and with the $\bar{u}$ and d bound together (by the strong force), this diagram represents a possible decay of the pion, $\pi^- \rightarrow e^- + \bar{\nu}_e$:

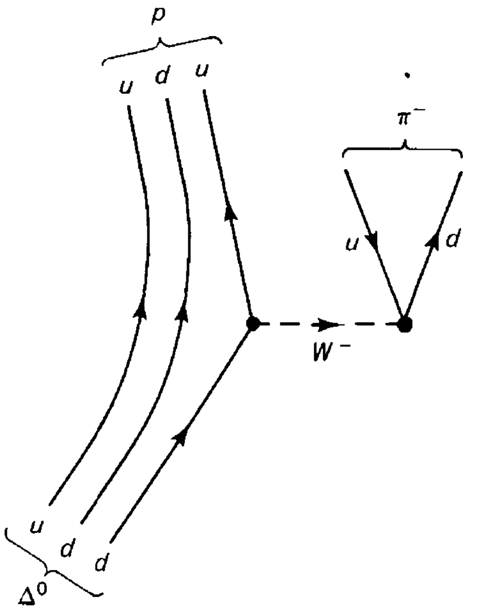


(For reasons to be discussed later, the more common decay is actually $\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$, but the diagram is the same, with e replaced by μ.) Moreover, essentially the same diagram accounts for the beta decay of the neutron ($n \rightarrow p^+ + e^- + \bar{\nu}_e$):

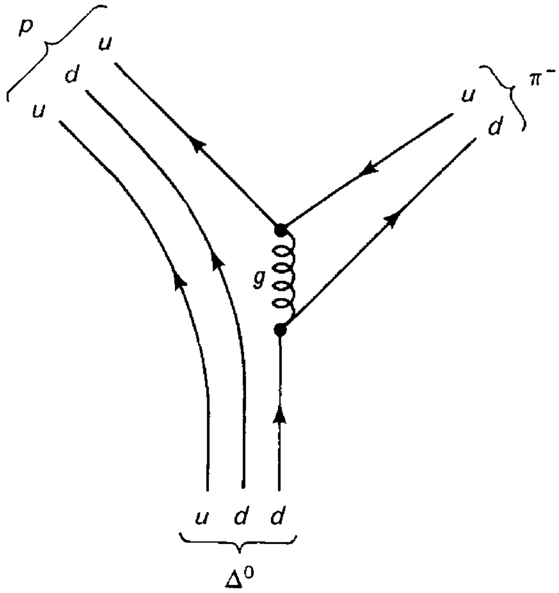


Thus, apart from strong interaction contamination (in the form of the "spectator" u and d quarks), the decay of the neutron is identical in structure to the decay of the muon, and closely related to the decay of the pion. In the days before the quark model, these appeared to be three very different processes.

Eliminating the electron-neutrino vertex in favor of a second quark vertex we obtain a purely hadronic weak interaction, $\Delta^0 \longrightarrow p^+ + \pi^-$:*

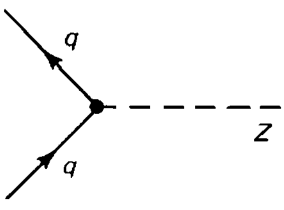


Actually, this particular decay also proceeds by the strong interaction:



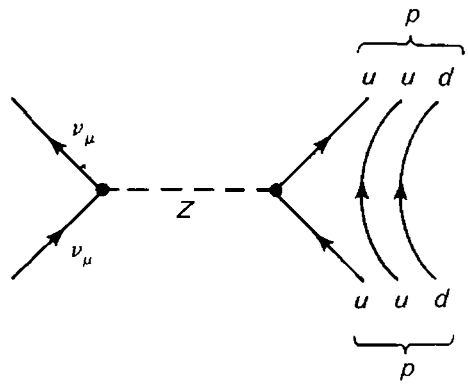
The weak mechanism is an immeasurably small contribution. We'll see more realistic examples of nonleptonic weak interactions in a moment.

The fundamental neutral vertex for leptons ($l \longrightarrow l + Z$) leaves the lepton species unchanged; again, it is natural to suppose that the same applies to quarks:



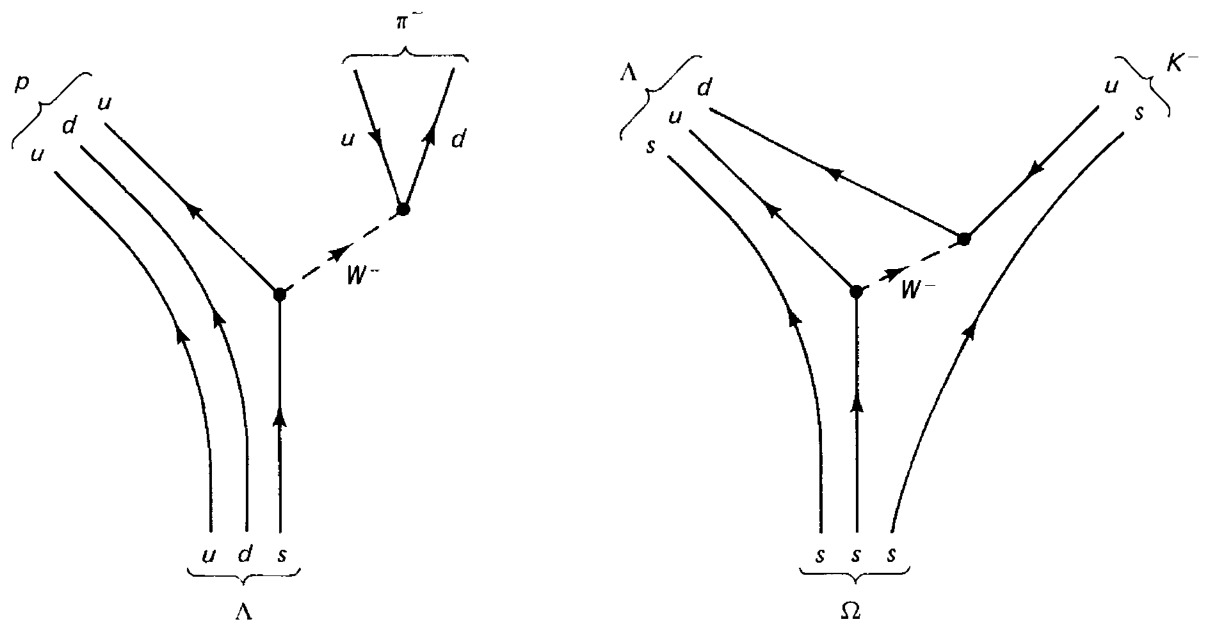
* The Δ^0 has the same quark content as the neutron, but this decay is not possible for neutrons because they are not heavy enough to make a proton and a pion.

This leads to neutrino-scattering processes such as $\nu_\mu + p \to \nu_\mu + p$:



Z exchange also makes a tiny contribution to the electron-proton force within an atom. As before, this contribution is masked by the dominant electromagnetic force, but it is detectable in certain carefully chosen atomic transitions.

So far, it's all pretty simple: The quarks mimic the leptons, as far as the weak interactions are concerned. The only difference is that because of the confining property of the *strong* force, there are generally spectator quarks present, which go along for the ride. Sad to say, this picture is a little *too* simple. For as long as the fundamental quark vertex is allowed to operate only *within each generation,* we can never hope to account for strangeness-changing weak interactions, such as the decay of the lambda ($\Lambda \to p^+ + \pi^-$) or the omega-minus ($\Omega^- \to \Lambda + K^-$), which involve the conversion of a strange quark into an up-quark:



The solution to this dilemma was suggested by Cabibbo in 1963, applied to neutral processes by Glashow, Illiopoulos, and Maiani (GIM) in 1970, and extended to three generations by Kobayashi and Maskawa (KM) in 1973.* The essential idea is that the quark generations are "skewed," for the purposes of weak interactions. Instead of

* The Cabibbo/GIM/KM mechanism will be discussed more fully in Chapter 10.

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix} \tag{2.3}$$

the weak force couples the pairs

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \quad \begin{pmatrix} c \\ s' \end{pmatrix}, \quad \begin{pmatrix} t \\ b' \end{pmatrix} \tag{2.4}$$

where d', s', and b' are *linear combinations* of the physical quarks d, s, and b:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.5}$$

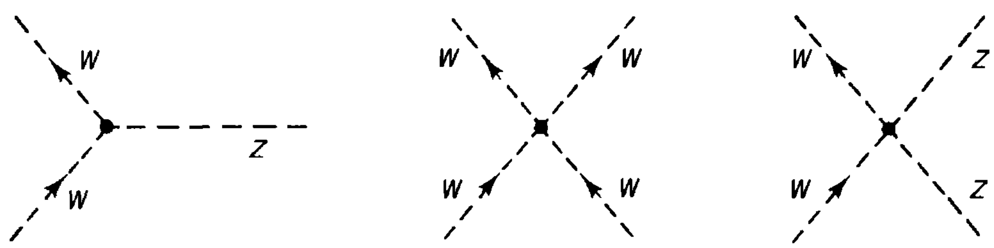
If this 3×3 *Kobayashi-Maskawa matrix* were the *unit* matrix, then d', s', and b' would be the same as d, s, and b, and no "cross-generational" transitions could occur. "Upness-plus-downness" would be absolutely conserved (just as the electron number is); "strangeness-plus-charm" would be conserved (like muon number); and so would "topness-plus-bottomness" (like tau number). But it's *not* the unit matrix (although it's pretty close); experimentally, the magnitudes of the matrix elements are[7]

$$\begin{pmatrix} 0.9705 \text{ to } 0.9770 & 0.21 \text{ to } 0.24 & 0. \text{ to } 0.014 \\ 0.21 \text{ to } 0.24 & 0.971 \text{ to } 0.973 & 0.036 \text{ to } 0.070 \\ 0. \text{ to } 0.024 & 0.036 \text{ to } 0.069 & 0.997 \text{ to } 0.999 \end{pmatrix} \tag{2.6}$$

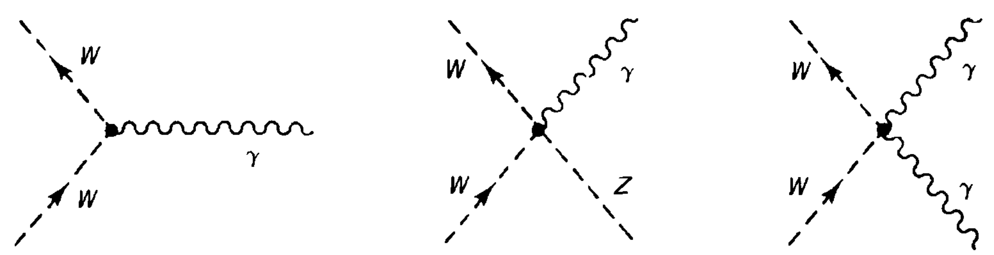
V_{ud} measures the coupling of u to d, V_{us} the coupling of u to s, and so on. The fact that the latter is nonzero is what permits strangeness-changing processes, such as the decay of the Λ and the Ω^-, to occur.

2.4.3 Weak and Electromagnetic Couplings of *W* and *Z*

There are also direct couplings of W and Z to one another, in GWS theory (just as there are direct gluon-gluon couplings in QCD):



Moreover, because the W is charged, it couples to the photon:



Although these interactions are critical for the internal consistency of the theory, as we shall see in Chapter 11, they are of limited practical importance at this point in time (see Problem 2.6).

2.5 DECAYS AND CONSERVATION LAWS

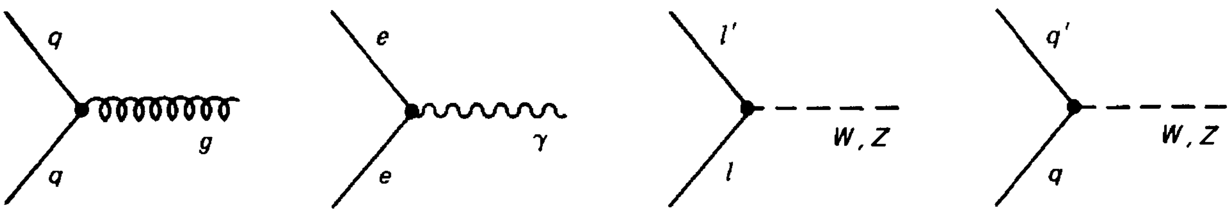
One of the most striking general properties of elementary particles is their tendency to disintegrate; we might almost call it a universal principle that *every particle decays into lighter particles, unless prevented from doing so by some conservation law.* The neutrinos and the photon are stable (having zero mass, there is nothing lighter for them to decay into); the electron is stable (it's the lightest charged particle, so conservation of charge prevents its decay); and the proton is presumably stable (it's the lightest baryon, and the conservation of baryon number saves it). By the same token, the positron and the antiproton are stable. But apart from these, all particles spontaneously disintegrate, even the neutron, although it becomes stable in the protective environment of many atomic nuclei. In practice, our world is populated mainly by protons, neutrons, electrons, photons, and neutrinos; more exotic things are created now and then (by collisions) but they do not last long. Each unstable species has a characteristic mean lifetime,* τ: for the muon it's 2.2×10^{-6} sec; for the π^+ it's 2.6×10^{-8} sec; for the π^0 it's 8.3×10^{-17} sec. Most particles exhibit several different decay *modes;* 64% of all K^+'s, for example, decay into $\mu^+ + \nu_\mu$, but 21% go to $\pi^+ + \pi^0$, 6% to $\pi^+ + \pi^+ + \pi^-$, 5% to $(e^+ + \nu_e + \pi^0)$, and so on. One of the goals of elementary particle theory is to calculate these lifetimes and *branching ratios.*

A given decay is governed by one of the three fundamental forces: $\Delta^{++} \rightarrow p^+ + \pi^+$, for example, is a strong decay; $\pi^0 \rightarrow \gamma + \gamma$ is electromagnetic; and $\Sigma^- \rightarrow n + e^- + \bar{\nu}_e$ is weak. How can we tell? Well, if a photon comes out, the process is certainly electromagnetic, and if a neutrino emerges, the process is certainly weak. But if neither a photon nor a neutrino is present, it's a little harder to say. For example, $\Sigma^- \rightarrow n + \pi^-$ is weak, but $\Delta^- \rightarrow n + \pi^-$ is strong. I'll show you in a moment how to figure that out, but first I want to mention the most dramatic *experimental* difference between strong, electromagnetic, and weak decays: A typical strong decay involves a lifetime around 10^{-23} sec, a typical electromagnetic decay takes about 10^{-16} sec, and weak decay times range from around 10^{-13} sec (for the τ) up to 15 min (for the neutron). For a given type of interaction, the decay generally proceeds more rapidly the larger the mass difference between the original particle and the decay products, just as a ball rolls faster down a steeper hill. There are exceptions: $\pi^+ \rightarrow \mu^+ + \nu_\mu$, for example, is faster by a factor of 10^4 than $\pi^+ \rightarrow e^+ + \nu_e$, but such cases demand special explanations. It is this kinematic effect that accounts for the enormous range in weak interaction lifetimes. In particular, the proton and electron together are *so*

* The *lifetime* τ is related to the *half-life* $t_{1/2}$ by the formula $t_{1/2} = (\ln 2)\tau = 0.693\tau$. The half-life is the time it takes for half the particles in a large sample to disintegrate (see Ch. 6, Sect. 6.1).

close to the neutron's mass that the decay $n \rightarrow p^+ + e^- + \bar{\nu}_e$ *barely* makes it at all, and the lifetime of the neutron is greater by far than that of any other unstable particle. Experimentally, then, there is a vast separation in lifetime between strong and electromagnetic decays (a factor of about 10 million), and again between electromagnetic and weak decays (a factor of at least a thousand). Indeed, particle physicists are so used to thinking in terms of 10^{-23} sec as the "normal" unit of time that the handbooks generally classify anything with a lifetime greater than 10^{-17} sec or so as a "stable" particle!* [8]

Now, what about the conservation laws which, as I say, permit certain reactions and forbid others? To begin with there are the purely kinematic conservation laws—conservation of energy and momentum (which we shall study in Chapter 3) and conservation of angular momentum (which comes in Chapter 4). The fact that a particle cannot spontaneously decay into particles heavier than itself is actually a consequence of conservation of energy (although it may seem so "obvious" as to require no explanation at all). The kinematic conservation laws apply to *all* interactions—strong, electromagnetic, weak, and for that matter anything else that may come along in the future—since they derive from special relativity itself. However, our concern right now is with the *dynamical* conservation laws that govern the three relevant interactions. Ten years ago I would simply have *stated* them as empirical rules coming from experiment, which you just have to memorize. It is in that spirit that we encountered them in Chapter 1. But now that we have a workable model for each of the basic forces, it becomes a question of examining the fundamental vertices:



Strong Electromagnetic Weak

Since all physical processes are obtained by sticking these together in elaborate combinations, anything that is conserved at each vertex must be conserved for the reactions as a whole. So, what do we have?

1. *Conservation of charge:* All three interactions, of course, conserve electric charge. In the case of the weak interactions the lepton (or quark) that comes out may not have the same charge as the one that went in, but if so, the difference is carried away by the *W*.

* Incidentally, 10^{-23} sec is about the time it takes a light signal to cross a hadron (diameter $\sim 10^{-15}$ m). You obviously cannot determine the lifetime of such a particle by measuring the length of its track [as we did for the Ω^- in Problem 1.8(b)]. Instead, you make a histogram of *mass* measurements, and invoke the uncertainty principle: $\Delta E\, \Delta t = \hbar$. Here $\Delta E = (\Delta m)c^2$, and $\Delta t = \tau$, so we get

$$\tau = \frac{\hbar}{(\Delta m)c^2}$$

Thus the *spread in mass* is a measure of the particle's lifetime.

2. *Conservation of color:* The electromagnetic and weak interactions do not affect color. At a strong vertex the quark color does change, but the difference is carried off by the gluon. (The direct gluon-gluon couplings also conserve color.) However, since naturally occurring particles are always colorless, the observable manifestation of color conservation is pretty trivial: zero in, zero out.

3. *Conservation of baryon number:* In all the primitive vertices, if a quark goes *in,* a quark comes *out,* so the total number of quarks present is a constant. In this arithmetic we count *anti*quarks as *negative,* so that, for example, at the vertex $q + \bar{q} \to g$ the quark number is zero before and zero after. Of course, we never see individual quarks, only baryons (with quark number 3), antibaryons (quark number -3), and mesons (quark number zero). So, in practice, it is more convenient to speak of the conservation of baryon number ($A = 1$ for baryons, $A = -1$ for antibaryons, and $A = 0$ for everything else). The baryon number is just $\frac{1}{3}$ the quark number. Notice that there is no analogous conservation of *meson number;* since mesons carry zero quark number, a given collision or decay can produce as many mesons as it likes, consistent with conservation of energy.

4. *Electron number, muon number, and tau number:* The strong forces do not touch leptons at all; in an electromagnetic interaction the same particle comes out (accompanied by a photon) as went in; and the weak interactions only mix together leptons from the same generation. So, the electron number, muon number, and tau number are all conserved. If it weren't for Cabibbo mixing, there would be a similar conservation of generation type for quarks (upness-plus-downness, strangeness-plus-charm, and beauty-plus-truth), but the fact that the generations are skewed in the weak interactions spoils things, and there is no hadronic analog to conservation of the individual lepton numbers.

5. *Approximate conservation of flavor:* So far, all the conservation laws we have considered are *absolute,* in the sense that they hold for all three interactions, as presently understood. An observed violation of any of them would be big news, calling for a major overhaul in our view of the subatomic world. But what about quark *flavor?* Flavor is conserved at a strong or electromagnetic vertex, but *not* at a weak vertex, where an up quark may turn into a down quark or a strange quark, with nothing at all picking up the lost upness or supplying the "gained" downness or strangeness. Because the weak forces *are* so weak, we say that the various flavors are *approximately* conserved. In fact, as you may remember, it was precisely this approximate conservation that led Gell-Mann to introduce the notion of strangeness in the first place. He "explained" the fact that strange particles are always produced in pairs:

$$\pi^-(d\bar{u}) + p^+(uud) \to K^+(u\bar{s}) + \Sigma^-(dds) \tag{2.7}$$

for instance, but

$$\pi^-(d\bar{u}) + p^+(uud) \not\to \pi^+(u\bar{d}) + \Sigma^-(dds) \tag{2.8}$$

by arguing that the latter violates conservation of strangeness. (Actually, this *is* a possible *weak* interaction, but it will never be seen in the laboratory, because it must compete against enormously more probable strong processes that do conserve strangeness.) In *decays,* however, the nonconservation of strangeness is very conspicuous, because for many particles this is the only way they *can* decay; there *is* no competition from strong or electromagnetic processes. The Λ, for instance, is the lightest strange baryon; if it is to decay, it must go to n (or p) plus something. But the lightest strange meson is the K, and n (or p) plus K weighs substantially more than the Λ. If the Λ decays at all (and it *does,* as we know: $\Lambda \rightarrow p^+ + \pi^-$ 64% of the time, and $\Lambda \rightarrow n + \pi^0$ 36% of the time), then strangeness cannot be conserved, and the reaction must proceed by the weak interaction. By contrast, the Δ^0 (with a strangeness of zero) can go to $p^+ + \pi^-$ or $n + \pi$ by the *strong* interaction, and its lifetime is accordingly much shorter.

6. *The OZI Rule:* Finally, I must tell you about one very peculiar case that has been on my conscience since Chapter 1. I have in mind the decay of the psi, which, you will recall, is a bound state of the charmed quark and its antiquark: $\psi = c\bar{c}$. The ψ has an anomalously long lifetime ($\sim 10^{-20}$ sec); the question is, *why?* It has nothing to do with conservation of charm; the net charm of the psi is zero. The ψ lifetime is short enough so that the decay is clearly due to the strong interactions. But why is it a thousand times slower than a strong decay "ought" to be? The explanation (if you call it that) goes back to an old observation by Okubo, Zweig, and Iizuka, known as the "OZI rule." These authors were puzzled by the fact that the ϕ meson (whose quark content, $s\bar{s}$, makes it the strange analog to the ψ) decays much more often into two K's than into three π's (the *two* pion decay is forbidden for other reasons, which we will come to in Chapter 4), in spite of the fact that the three pion decay is energetically favored (the mass of two K's is 990 MeV/c^2; three π's weigh only 415 MeV/c^2). In Figure 2.4, we see that the three-pion diagram can be cut in two by *snipping*

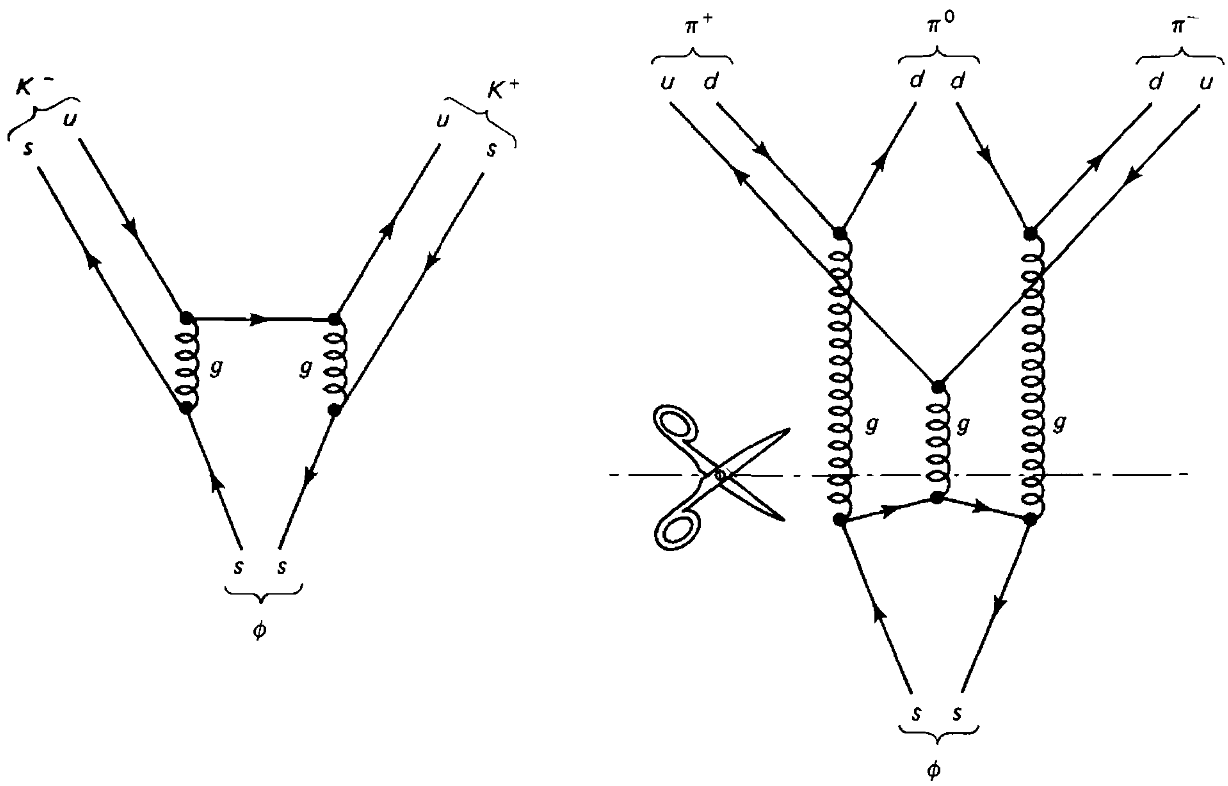


Figure 2.4 The OZI rule: If the diagram can be cut in two by slicing only gluon lines (and not cutting open any external particles), the process is suppressed.

only gluon lines. The OZI rule states that such processes are "suppressed." Not absolutely forbidden, mind you, for the decay $\phi \rightarrow 3\pi$ *does* in fact occur, but far *less likely* than one would otherwise have supposed. The OZI rule is related to asymptotic freedom, in the following sense: In an OZI-suppressed diagram the gluons must be "hard" (high energy), since they carry the energy necessary to make the hadrons into which they fragment. But asymptotic freedom says that gluons couple weakly at high energies (short ranges). By contrast, in OZI-allowed processes the gluons are typically "soft" (low energy), and in this regime the coupling is strong. Qualitatively, at least, this accounts for the OZI rule. (The quantitative details will have to await a more complete understanding of QCD.)

But what does all this have to do with the ψ? Well, presumably the same rule applies, suppressing $\psi \rightarrow 3\pi$, and leaving the decay into two charmed D mesons (analogs to the K, but with the charmed quarks in place of the strange quarks) as the favored route. Only there's a new twist in the ψ system, for the D's turn out to be too heavy: A pair of D's weighs more than the ψ. So the decay $\psi \rightarrow D^+ + D^-$ (or $D^0 + \bar{D}^0$) is *kinematically* forbidden, while $\psi \rightarrow 3\pi$ is *OZI suppressed,* and it is to this happy combination that the ψ owes its unusual longevity.

2.6 UNIFICATION SCHEMES

At one time electricity and magnetism were two distinct subjects, the one dealing with pith balls, batteries, and lightning; the other with lodestones, bar magnets, and the North Pole. But in 1820 Oersted noticed that an electric current could deflect a magnetic compass needle, and 10 years later Faraday discovered that a moving magnet could generate an electric current in a loop of wire. By the time Maxwell put the whole theory together in its final form, electricity and magnetism were properly regarded as two aspects of a single subject: electromagnetism.

Einstein dreamed of going a step further, combining gravity with electrodynamics in a single *unified field theory.* Although this program was not successful, a similar vision inspired Glashow, Weinberg, and Salam to join the *weak* and electromagnetic forces. Their theory starts out with four massless mediators, but, as it develops, three of them acquire mass (by the so-called *Higgs mechanism*), becoming the W's and the Z, while one remains massless: the photon. Although experimentally a reaction mediated by W or Z is quite different from one mediated by the γ, if the GWS theory is right they are all manifestations of a single *electroweak* interaction. The relative *weakness* of the weak force is attributable to the enormous mass of the intermediate vector bosons; its *intrinsic* strength is in fact somewhat *greater* than that of the electromagnetic force, as we shall see in Chapter 10.

Beginning in the early seventies, many people have been working on the obvious next step: combining the strong force (in the form of chromodynamics) with the electroweak force (à la GWS). Several different schemes for implementing this *grand unification* are now on the table, and although it is too soon to draw any definitive conclusions, some of the early results are promising. You will recall that the strong coupling constant α_s *decreases* at short distances (which is to say, for very high-energy collisions). So too does the weak coupling α_w, but at a slower rate. Meanwhile, the electromagnetic coupling constant, α_e, which is the smallest of the three, *increases*. Could it be that they all converge to a common limiting value, at extremely high energy? (See Fig. 2.5.) Such is the promise of the grand unified theories (GUTs). Indeed, from the functional form of the running coupling constants it is possible to estimate the energy at which this unification occurs: around 10^{15} GeV. This is, of course, astronomically higher than any currently accessible energy (remember, the mass of the Z is 90 GeV/c^2). Nevertheless, it is an exciting idea, for it means that the observed difference in strength among the three interactions is an "accident" resulting from the fact that we are obliged to work at low energies, where the unity of the forces is obscured. If we could just get in close enough to see the "true" strong, electric, and weak charges, without any of the screening effects of vacuum polarization, we would find that they are all equal. How nice!

Another suggestion of the GUTs is that the proton may be unstable, although its half-life is fantastically long (at least 10^{20} times the age of the universe). It has often been remarked that conservation of charge and color are in a sense more "fundamental" than the conservation of baryon number and lepton number, because charge is the "source" for electrodynamics, and color for chromodynamics. If these quantities were not conserved, QED and QCD would have to be completely reformulated. But baryon number and lepton number do not function as sources for any interaction, and their conservation has no deep dynamical significance. In the grand unified theories new interactions are contemplated, permitting decays such as

$$p^+ \to e^+ + \pi^0 \quad \text{or} \quad p^+ \to \bar{\nu}_\mu + \pi^+ \tag{2.9}$$

in which baryon number and lepton number change. Several major experiments are now under way to search for these proton decays. So far, the results are negative.[9]

If grand unification works, all of elementary particle physics will be reduced to the action of a single force. The final step, then, will be to bring in gravity, vindicating at last Einstein's dream. Indeed, many theorists are already working

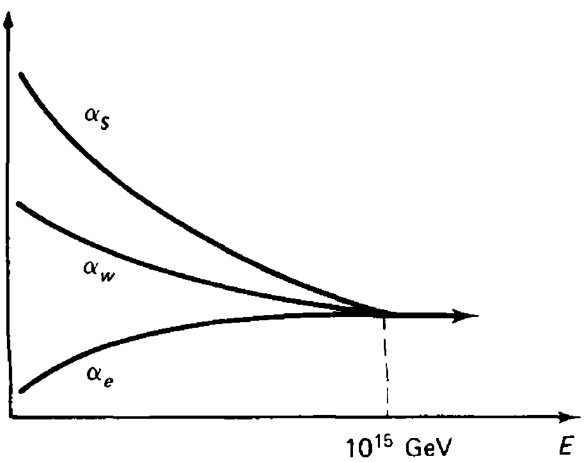


Figure 2.5 Evolution of the three fundamental coupling constants.

on this, the ultimate unification. But it is probably safe to say that a detailed theory is still years off—after all, we hardly know how to carry out the most rudimentary calculations in chromodynamics, and here we are speculating about a theory two generations more sophisticated!

REFERENCES AND NOTES

1. Consistent etymology would call for *geusidynamics,* from the Greek word for "flavor"; see *Phys. Today* (April 1981), p. 74. M. Gaillard suggests *asthenodynamics,* from the Greek word for *weak.*
2. See, for example, E. M. Purcell, *Electricity and Magnetism,* 2d Ed. (New York: McGraw-Hill, 1985), Sec. 10.1.
3. C. Quigg, in *Sci. Am.* (April 1985) gives a qualitative interpretation of the effect of gluon polarization, but I do not find it entirely persuasive. Quigg's article is an outstanding and accessible introduction to the current state of elementary particle physics.
4. The classic papers on weak interaction theory up to 1960 are collected in P. K. Kabir, ed., *The Development of Weak Interaction Theory* (New York: Gordon & Breach, 1963). A similar collection covering the modern era is contained in C. H. Lai, ed., *Gauge Theory of Weak and Electromagnetic Interactions* (Singapore: World Scientific, 1981).
5. F. J. Hasert et al., *Phys. Lett.* **46B,** 138 (1973), and *Nucl. Phys.* **B73,** 1 (1974). See also D. B. Cline, A. K. Mann, and C. Rubbia, *Sci. Am.* (December 1974).
6. S. L. Wu, *Phys. Rep.* **107,** 59 (1984), Section 5.6. See also M.-A. Bouchiat and L. Pottier, *Sci. Am.* (June 1984).
7. F. J. Gilman, *Rev. Mod. Phys.* **56,** S296 (1984).
8. See D. T. Gillespie, *A Quantum Mechanics Primer* (London: International Textbook Co.), p. 78, for a careful justification of the procedure discussed in this footnote.
9. J. M. LoSecco, F. Reines, and D. Sinclair, *Sci. Am.* (June 1985).

PROBLEMS

2.1. Calculate the ratio of the gravitational attraction to the electrical repulsion between two stationary electrons. (Do I need to tell you how far apart they are?)

2.2. Sketch the lowest-order Feynman diagram representing *Delbruck scattering:* $\gamma + \gamma \rightarrow \gamma + \gamma$. (This process, the scattering of light by light, has no analog in classical electrodynamics.)

2.3. Draw all the fourth-order (four vertex) diagrams for Compton scattering. (There are 17 of them; disconnected diagrams don't count.)

2.4. Determine the mass of the virtual photon in each of the lowest-order diagrams for Bhabha scattering (assume the electron and positron are at rest). What is its velocity? (Note that these answers would be impossible for *real* photons.)

2.5. **(a)** Which decay do you think would be more likely,

$$\Xi^- \rightarrow \Lambda + \pi^- \quad \text{or} \quad \Xi^- \rightarrow n + \pi^-$$

Explain your answer, and confirm it by looking up the experimental data.

(b) Which decay of the $D^0(c\bar{u})$ meson is more likely,

$$D^0 \rightarrow K^- + \pi^+, \qquad D^0 \rightarrow \pi^- + \pi^+, \quad \text{or} \quad D^0 \rightarrow K^+ + \pi^-$$

Which is *least* likely? Draw the Feynman diagrams, explain your answer and check the experimental data. (One of the successful predictions of the Cabibbo/GIM/KM model was that charmed mesons should decay preferentially into strange mesons, even though *energetically* the 2π mode is favored.)

(c) How about the "beautiful" (B) mesons? Should they go to the D's, K's, or π's? How about "truthful" mesons?

2.6. Draw all the lowest-order diagrams contributing to the process $e^+ + e^- \rightarrow W^+ + W^-$. [One of them involves the direct coupling of Z to W's and another the coupling of γ to W's, so if a positron-electron collider is ever built with sufficient energy to make two W's, these interactions will be directly observable.]

2.7. Examine the following processes, and state for each one whether it is *possible* or *impossible,* according to the Standard Model (which does not include GUTs, with their potential violation of the conservation of lepton number and baryon number). In the former case, state which interaction is responsible—strong, electromagnetic, or weak; in the latter case cite a conservation law that prevents it from occurring. (Following the usual custom, I will not indicate the charge when it is unambiguous, thus γ, Λ, and n are neutral; p is positive, e is negative; etc.)

(a) $p + \bar{p} \rightarrow \pi^+ + \pi^0$
(b) $\eta \rightarrow \gamma + \gamma$
(c) $\Sigma^0 \rightarrow \Lambda + \pi^0$
(d) $\Sigma^- \rightarrow n + \pi^-$
(e) $e^+ + e^- \rightarrow \mu^+ + \mu^-$
(f) $\mu^- \rightarrow e^- + \bar{\nu}_e$
(g) $\Delta^+ \rightarrow p + \pi^0$
(h) $\bar{\nu}_e + p \rightarrow n + e^+$
(i) $e + p \rightarrow \nu_e + \pi^0$
(j) $p + p \rightarrow \Sigma^+ + n + K^0 + \pi^+ + \pi^0$
(k) $p \rightarrow e^+ + \gamma$
(l) $p + p \rightarrow p + p + p + \bar{p}$
(m) $n + \bar{n} \rightarrow \pi^+ + \pi^- + \pi^0$
(n) $\pi^+ + n \rightarrow \pi^- + p$
(o) $K^- \rightarrow \pi^- + \pi^0$
(p) $\Sigma^+ + n \rightarrow \Sigma^- + p$
(q) $\Sigma^0 \rightarrow \Lambda + \gamma$
(r) $\Xi^- \rightarrow \Lambda + \pi^-$
(s) $\Xi^0 \rightarrow p + \pi^-$
(t) $\pi^- + p \rightarrow \Lambda + K^0$
(u) $\pi^0 \rightarrow \gamma + \gamma$
(v) $\Sigma^- \rightarrow n + e + \bar{\nu}_e$

2.8. Some decays involve two (or even all three) different forces. Draw possible Feynman diagrams for the following processes:

(a) $K^+ \rightarrow \mu^+ + \nu_\mu + \gamma$
(b) $\Sigma^+ \rightarrow p + \gamma$

What interactions are involved? (Both these decays have been observed, by the way.)

2.9. The upsilon meson, $b\bar{b}$, is the bottom-quark analog to the ψ, $c\bar{c}$. Its mass is 9460 MeV/c^2, and its lifetime is 1.5×10^{-20} sec. From this information, what can you say about the mass of the B meson, $u\bar{b}$? (The observed mass is 5270 MeV/c^2.)

2.10. The ψ' meson, at 3685 MeV/c^2, has the same quark content as the ψ (i.e., $c\bar{c}$). Its principal decay mode is $\psi' \rightarrow \psi + \pi^+ + \pi^-$. Is this a strong interaction? Is it OZI-suppressed? What lifetime would you expect for the ψ'? (The observed value is 3×10^{-21} sec.)

Chapter 3

Relativistic Kinematics

In this chapter I summarize the basic principles, notation, and terminology of relativistic kinematics. This is material you must know cold *in order to understand Chapters 6 through 11 (it is* not *needed for Chapters 4 and 5, however, and if you prefer you can read them first). Although the treatment is reasonably self-contained, I do assume that you have encountered special relativity before—if not, you should pause here and read the appropriate chapter in any introductory physics text before proceeding. If you are already quite familiar with relativity, this chapter will be an easy review—but read through it anyway because some of the notation may be new to you.*

3.1 LORENTZ TRANSFORMATIONS

According to the special theory of relativity,[1] the laws of physics apply just as well in a reference system moving at constant velocity as they do in one at rest. An embarrassing implication of this is that there's no way of telling which system (if any) *is* at rest, and hence there is no way of knowing what "the" velocity of any other system might be. So perhaps I had better start over. Ahem.

According to the special theory of relativity,[1] the laws of physics are equally valid in all *inertial* reference systems. An inertial system is one in which Newton's first law (the law of inertia) is obeyed: objects keep moving in straight lines at constant speeds unless acted upon by some force.* It's easy to see that any two inertial systems must be moving at constant velocity with respect to one another, and conversely, that any system moving at constant velocity with respect to an inertial system is itself inertial.

* If you are wondering whether a freely falling system in a uniform gravitational field is "inertial," you know more than is good for you. Let's just keep gravity out of it.

Imagine, then, that we have two inertial frames, S and S', with S' moving at uniform velocity $\mathbf{v}$ (magnitude v) with respect to S (S, then, is moving at velocity $-\mathbf{v}$ with respect to S'). We may as well lay out our coordinates in such a way that the motion is along the common x/x' axis (Fig. 3.1), and set the master clocks at the origin in each system so that both read zero at the instant the two coincide (that is, $t = t' = 0$ when $x = x' = 0$). Suppose, now, that some event occurs at position (x, y, z) and time t in S. Question: What are the spacetime coordinates (x', y', z') and t' of this *same event* in S'? The answer is provided by the Lorentz transformations:

$$\begin{aligned} &\text{i.} \quad x' = \gamma(x - vt) \\ &\text{ii.} \quad y' = y \\ &\text{iii.} \quad z' = z \\ &\text{iv.} \quad t' = \gamma\left(t - \frac{v}{c^2}x\right) \end{aligned} \tag{3.1}$$

where

$$\gamma \equiv \frac{1}{\sqrt{1 - v^2/c^2}} \tag{3.2}$$

The inverse transformations, which take us back from S' to S, are obtained by simply changing the sign of v (see Problem 3.1):

$$\begin{aligned} &\text{i}'. \quad x = \gamma(x' + vt') \\ &\text{ii}'. \quad y = y' \\ &\text{iii}'. \quad z = z' \\ &\text{iv}'. \quad t = \gamma\left(t' + \frac{v}{c^2}x'\right) \end{aligned} \tag{3.3}$$

The Lorentz transformations have a number of immediate consequences, of which I mention briefly the most important:

1. *The relativity of simultaneity:* If two events occur at the same time in S, but at different locations, then they do *not* occur at the same time in S'. Specifically, if $t_A = t_B$, then

$$t'_A = t'_B + \frac{\gamma v}{c^2}(x_B - x_A) \tag{3.4}$$

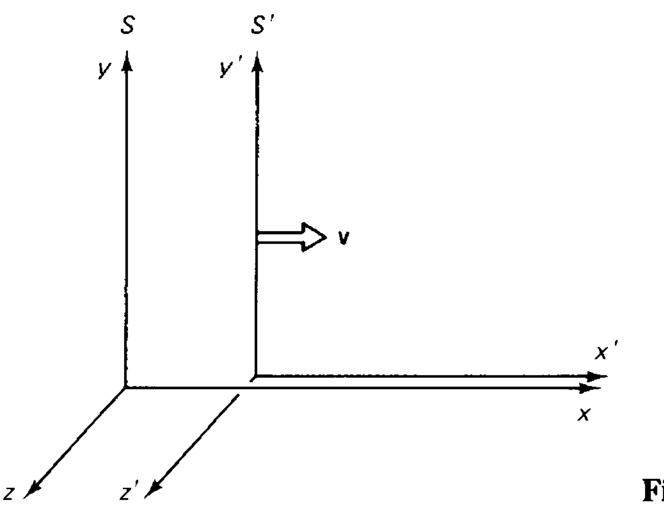


Figure 3.1 The inertial systems S and S'.

(see Problem 3.2). Events that are simultaneous in one inertial system, then, are not simultaneous in others.

2. *Lorentz contraction:* Suppose a stick lies on the x' axis, at rest in S'. Say one end is at the origin ($x' = 0$) and the other is at L' (so its length in S' is L'). What is its length as measured in S? Since the stick is *moving* with respect to S, we must be careful to record the positions of its two ends at the same instant, say $t = 0$. At that moment the left end is at $x = 0$ and the right end, according to equation (i), is at $x = L'/\gamma$. Thus the length of the stick is $L = L'/\gamma$, in S. Notice that γ is always greater than or equal to 1. It follows that a *moving object is shortened* by a factor of γ, as compared with its length in the system in which it is at rest. Notice that Lorentz contraction only applies to lengths *along the direction of motion;* perpendicular dimensions are not affected.

3. *Time dilation:* Suppose the clock at the origin in S' ticks off an interval T'; for simplicity, say it runs from $t' = 0$ to $t' = T'$. How long is this period as measured in S? Well, it begins at $t = 0$, and it ends when $t' = T'$ at $x' = 0$, so [according to eq. (iv')] $t = \gamma T'$. Evidently the clocks in S tick off a *longer* interval, $T = \gamma T'$, by that same factor of γ; or, put it the other way around: *moving clocks run slow.* Unlike Lorentz contraction, which is only indirectly relevant to elementary particle physics, time dilation is a commonplace in the laboratory. For in a sense every unstable particle has a built-in clock: whatever it is that tells the particle when its time is up. And these internal clocks do indeed run slow when the particle is moving. That is to say, a moving particle lasts longer (by a factor of γ) than it would at rest.* (The tabulated lifetimes are, of course, for particles at rest.) In fact, the cosmic ray muons produced in the upper atmosphere would never make it to ground level were it not for time dilation (see Problem 3.4).

4. *Velocity addition.* Suppose a particle is moving in the x direction at speed u', with respect to S'. What is its speed, u, with respect to S? Well, it travels a distance $\Delta x = \gamma(\Delta x' + v\,\Delta t')$ in a time $\Delta t = \gamma[\Delta t' + (v/c^2)\Delta x']$, so

$$\frac{\Delta x}{\Delta t} = \frac{\Delta x' + v\,\Delta t'}{\Delta t' + (v/c^2)\,\Delta x'} = \frac{(\Delta x'/\Delta t') + v}{1 + (v/c^2)(\Delta x'/\Delta t')}.$$

But $\Delta x/\Delta t = u$, and $\Delta x'/\Delta t' = u'$, so

$$u = \frac{u' + v}{1 + (u'v/c^2)} \tag{3.5}$$

The numerator represents the classical answer to the same question, $u = u' + v$; the denominator introduces a relativistic correction that is small unless u' and v are close to c. Notice that if $u' = c$, then $u = c$ also: the speed of light is the same in all inertial systems.

* Actually, the disintegration of an individual particle is a random process; when we speak of a "lifetime" we really mean the *average* lifetime of that particle type. When I say that a moving particle lasts longer, I really mean that the *average* lifetime of a *group* of moving particles is longer.

3.2 FOUR-VECTORS

It is convenient at this point to introduce some simplifying notation. We define the *position-time four-vector* x^μ, $\mu = 0, 1, 2, 3$, as follows:

$$x^0 = ct, \qquad x^1 = x, \qquad x^2 = y, \qquad x^3 = z \tag{3.6}$$

In terms of x^μ, the Lorentz transformations take on a more symmetrical appearance:

$$\begin{aligned} x^{0\prime} &= \gamma(x^0 - \beta x^1) \\ x^{1\prime} &= \gamma(x^1 - \beta x^0) \\ x^{2\prime} &= x^2 \\ x^{3\prime} &= x^3 \end{aligned} \tag{3.7}$$

where

$$\beta \equiv \frac{v}{c} \tag{3.8}$$

More compactly:

$$x^{\mu\prime} = \sum_{\nu=0}^{3} \Lambda^\mu_\nu x^\nu \qquad (\mu = 0, 1, 2, 3) \tag{3.9}$$

The coefficients Λ^μ_ν may be regarded as the elements of a matrix Λ:

$$\Lambda = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.10}$$

(i.e., $\Lambda^0_0 = \Lambda^1_1 = \gamma$; $\Lambda^1_0 = \Lambda^0_1 = -\gamma\beta$; $\Lambda^2_2 = \Lambda^3_3 = 1$; and all the rest are zero). To avoid writing lots of Σ's, we shall follow Einstein's "summation convention," which says that repeated Greek indices (one as subscript, one as superscript) are to be summed from 0 to 3. Thus equation (3.9) becomes, finally,*

$$x^{\mu\prime} = \Lambda^\mu_\nu x^\nu \tag{3.11}$$

A special virtue of this tidy notation is that the same form describes Lorentz transformations which are *not* along the x direction; in fact, the S and S' axes need not even be parallel; the Λ matrix is more complicated, naturally, but equation (3.11) still holds. [On the other hand, there is no real loss of generality in using expression (3.10), since we are always free to *choose* parallel axes, and to align the x axis along the direction of **v**.]

* In an expression such as this the Greek letter used for the summation index, ν, is of course completely arbitrary. The same goes for the index μ, although it must match on the two sides of the equation. Thus equation (3.11) could just as well be written $x^{\kappa\prime} = \Lambda^\kappa_\lambda x^\lambda$. Either expression stands for the set of four equations:

$$\begin{aligned} x^{0\prime} &= \Lambda^0_0 x^0 + \Lambda^0_1 x^1 + \Lambda^0_2 x^2 + \Lambda^0_3 x^3 \\ x^{1\prime} &= \Lambda^1_0 x^0 + \Lambda^1_1 x^1 + \Lambda^1_2 x^2 + \Lambda^1_3 x^3 \\ x^{2\prime} &= \Lambda^2_0 x^0 + \Lambda^2_1 x^1 + \Lambda^2_2 x^2 + \Lambda^2_3 x^3 \\ x^{3\prime} &= \Lambda^3_0 x^0 + \Lambda^3_1 x^1 + \Lambda^3_2 x^2 + \Lambda^3_3 x^3 \end{aligned}$$

Although the individual coordinates of an event change, in accordance with equation (3.11), when we go from S to S', there is a particular *combination* of them that remains the same (Problem 3.7):

$$I \equiv (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = (x^{0\prime})^2 - (x^{1\prime})^2 - (x^{2\prime})^2 - (x^{3\prime})^2 \quad (3.12)$$

Such a quantity, which has the same value in any inertial system, is called an *invariant*. (In the same sense, the quantity $r^2 = x^2 + y^2 + z^2$ is invariant under *rotations*.) Now, I would *like* to write this invariant in the form of a sum: $\sum_{\mu=0}^{3} x^\mu x^\mu$, but unfortunately there are those three irritating minus signs. To keep track of them, we introduce the *metric*, $g_{\mu\nu}$, whose components can be displayed as a matrix $\mathbf{g}$:

$$\mathbf{g} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad (3.13)$$

(i.e., $g_{00} = 1$; $g_{11} = g_{22} = g_{33} = -1$; all the rest are zero).* With the help of $g_{\mu\nu}$, the invariant I can be written as a double sum:

$$I = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} g_{\mu\nu} x^\mu x^\nu = g_{\mu\nu} x^\mu x^\nu \quad (3.14)$$

Carrying things a step further, we define the *covariant* four-vector x_μ (index *down*) as follows:

$$x_\mu \equiv g_{\mu\nu} x^\nu \quad (3.15)$$

(i.e., $x_0 = x^0$, $x_1 = -x^1$, $x_2 = -x^2$, $x_3 = -x^3$). To emphasize the distinction we call the "original" four-vector x^μ (index *up*) a *contravariant* four-vector. The invariant I can then be written in its cleanest form:

$$I = x_\mu x^\mu \quad (3.16)$$

All this will no doubt seem like monstrous notational overkill, just to keep track of three minus signs, but it's actually very simple, once you get used to it. (What's more, it generalizes nicely to non-Cartesian coordinate systems and to the curved spaces encountered in general relativity, though neither of these is relevant to us here.)

The position-time four-vector x^μ is the archetype for all four-vectors. We define a four-vector, a^μ, as a four-component object that transforms in the same way x^μ does when we go from one inertial system to another, to wit:

$$a^{\mu\prime} = \Lambda^\mu_\nu a^\nu \quad (3.17)$$

with the same coefficients Λ^μ_ν. To each such (contravariant) four-vector we as-

* I should warn you that some physicists define the metric with the opposite signs (−1, 1, 1, 1). It doesn't *matter* much—if I is invariant, so too is $-I$. But it does mean you must be on the lookout for unfamiliar signs. Fortunately, most *particle* physicists nowadays use the convention in equation (3.13).

sociate a *co*variant four-vector a_μ, obtained by simply changing the signs of the spatial components, or, more formally

$$a_\mu = g_{\mu\nu} a^\nu \tag{3.18}$$

Of course, we can go back from covariant to contravariant by reversing the signs again:

$$a^\mu = g^{\mu\nu} a_\nu \tag{3.19}$$

where $g^{\mu\nu}$ are technically the elements in the matrix $\mathbf{g}^{-1}$ (however, since our metric is its own inverse, $g^{\mu\nu}$ is the same as $g_{\mu\nu}$). Given any two four-vectors, a^μ and b^μ, the quantity

$$a^\mu b_\mu = a_\mu b^\mu = a^0 b^0 - a^1 b^1 - a^2 b^2 - a^3 b^3 \tag{3.20}$$

is invariant (the same number in any inertial system). We shall refer to it as the *scalar product* of a and b; it is the four-dimensional analog to the dot product of two three-vectors (there is no four-vector analog to the cross product).* If you get tired of writing indices, feel free to use the dot notation:

$$a \cdot b \equiv a_\mu b^\mu \tag{3.21}$$

However, you will then need a way to distinguish this four-dimensional scalar product from the ordinary dot product of two three-vectors. The best way is to be scrupulously careful to put an arrow over all three-vectors (except perhaps the velocity, **v**, which, since it is not part of a four-vector, is not subject to ambiguity). In this book, I use boldface for three-vectors. Thus

$$a \cdot b = a^0 b^0 - \mathbf{a} \cdot \mathbf{b} \tag{3.22}$$

We also use the notation a^2 for the scalar product of a^μ with itself:

$$a^2 \equiv a \cdot a = (a^0)^2 - \mathbf{a}^2 \tag{3.23}$$

Notice, however, that a^2 need not be positive. Indeed, we can classify all four-vectors according to the sign of a^2:

$$\begin{aligned} &\text{If } a^2 > 0, \quad && a^\mu \text{ is called } \textit{timelike} \\ &\text{If } a^2 < 0, \quad && a^\mu \text{ is called } \textit{spacelike} \\ &\text{If } a^2 = 0, \quad && a^\mu \text{ is called } \textit{lightlike} \end{aligned} \tag{3.24}$$

From *vectors* it is a short step to *tensors:* a second-rank tensor, $s^{\mu\nu}$, carries two indices, has $4^2 = 16$ components, and transforms with *two* factors of Λ:

$$s^{\mu\nu\prime} = \Lambda^\mu_\kappa \Lambda^\nu_\sigma s^{\kappa\sigma} \tag{3.25}$$

a third-rank tensor, $t^{\mu\nu\lambda}$, has three indices, $4^3 = 64$ components, and transforms with *three* factors of Λ:

$$t^{\mu\nu\lambda\prime} = \Lambda^\mu_\kappa \Lambda^\nu_\sigma \Lambda^\lambda_\tau t^{\kappa\sigma\tau} \tag{3.26}$$

* The closest thing is $(a^\mu b^\nu - a^\nu b^\mu)$, but this is a second-rank *tensor,* not a four-vector (see below).

and so on. In this hierarchy a vector is a tensor of rank 1, and a scalar (invariant) is a tensor of rank zero. We construct covariant and "mixed" tensors by lowering indices (at cost of a minus sign for each spatial index), for example

$$s^{\mu}{}_{\nu} = g_{\nu\lambda}s^{\mu\lambda}; \qquad s_{\mu\nu} = g_{\mu\kappa}g_{\nu\lambda}s^{\kappa\lambda} \tag{3.27}$$

and so on. Notice that the product of two tensors is itself a tensor [$(a^{\mu}b^{\nu})$ is a tensor of second rank; $(a^{\mu}t^{\nu\lambda\sigma})$ is a tensor of fourth rank; and so on.] Finally, we can obtain from any tensor of rank $n + 2$ a "contracted" tensor of rank n, by summing like upper and lower indices. Thus $s^{\mu}{}_{\mu}$ is a scalar; $t^{\mu\nu}{}_{\nu}$ is a vector; $a_{\mu}t^{\mu\nu\lambda}$ is a second-rank tensor.

3.3 ENERGY AND MOMENTUM

Suppose you're driving down the highway, and pretend for the sake of argument that you're going at close to the speed of light. You might want to keep an eye on two different "times": if you're worried about making an appointment in San Francisco, you should check the stationary clocks posted now and then along the side of the road. But if you're wondering when would be an appropriate time to stop for a bite to eat, it would be more sensible to look at the watch on your wrist. For according to relativity, the moving clock (in this case, your watch) is running slow (relative to the "stationary" clocks on the ground), and so too is your heart rate, your metabolism, your speech and thought, *every*thing. Specifically, while the "ground" time advances by an infinitesimal amount dt, your own (or *proper*) time advances by the smaller amount $d\tau$:

$$d\tau = \frac{dt}{\gamma} \tag{3.28}$$

At normal driving speeds, of course, γ is so close to 1 that dt and $d\tau$ are essentially identical, but in elementary particle physics the distinction between laboratory time (read off the clock on the wall) and particle time (as it would appear on the particle's watch) is crucial. Although we can always get from one to the other, using equation (3.28), in practice it is usually most convenient to work with proper time, because τ is invariant. All observers can read the particle's watch, and at any given moment they must all agree on what it says, even though their *own* clocks may differ from it and from one another.

When we speak of the "velocity" of a particle (with respect to the laboratory), we mean, of course, the distance it travels (measured in the lab frame) divided by the time it takes (measured on the lab clock):

$$\mathbf{v} = \frac{d\mathbf{x}}{dt} \tag{3.29}$$

But in view of what has just been said, it is also useful to introduce the "proper" velocity, $\boldsymbol{\eta}$, which is the distance traveled (again, measured in the lab frame) divided by the *proper* time:*

$$\boldsymbol{\eta} \equiv \frac{d\mathbf{x}}{d\tau} \tag{3.30}$$

According to equation (3.28), the two velocities are related by a factor of γ:

$$\boldsymbol{\eta} = \gamma \mathbf{v} \tag{3.31}$$

However, $\boldsymbol{\eta}$ is much easier to work with, for if we want to go from the lab system, S, to a moving system, S', *both the numerator and the denominator in (3.29) must be transformed* [leading to the cumbersome velocity addition rule (3.5)], whereas in equation (3.30) *only the numerator transforms;* $d\tau$, as we have seen, is invariant. In fact, proper velocity is part of a four-vector:

$$\eta^\mu = \frac{dx^\mu}{d\tau} \tag{3.32}$$

whose zeroth component is

$$\eta^0 = \frac{dx^0}{d\tau} = \frac{d(ct)}{(1/\gamma)dt} = \gamma c \tag{3.33}$$

Thus

$$\eta^\mu = \gamma(c, v_x, v_y, v_z) \tag{3.34}$$

Incidentally, $\eta_\mu \eta^\mu$ should be invariant, and it *is:*

$$\eta_\mu \eta^\mu = \gamma^2(c^2 - v_x^2 - v_y^2 - v_z^2) = \gamma^2 c^2(1 - v^2/c^2) = c^2 \tag{3.35}$$

They don't make 'em more invariant than *that!*

Classically, momentum is mass times velocity. We would like to carry this over in relativity, but the question arises: Which velocity should we use—*ordinary* velocity or *proper* velocity? Classical considerations offer no clue, for the two are equal in the nonrelativistic limit. In a sense, it's just a matter of definition, but there is a subtle and compelling reason why ordinary velocity would be a *bad* choice, whereas proper velocity is a *good* choice. The point is this: If we defined momentum as $m\mathbf{v}$, then the law of conservation of momentum would be inconsistent with the principle of relativity (if it held in one inertial system, it would *not* hold in other inertial systems). But if we define momentum as $m\boldsymbol{\eta}$, then conservation of momentum *is* consistent with the principle of relativity (if it holds in one inertial system, it automatically holds in all inertial systems). I'll let you prove this for yourself in Problem 3.10. Mind you, this doesn't guarantee

* Proper velocity is a hybrid quantity, in the sense that distance is measured in the *lab* frame, whereas time is measured in the *particle* frame. Some people object to the adjective "proper" in this context, holding that this should be reserved for quantities measured entirely in the particle frame. Of course, in its *own* frame the particle never moves at all—its velocity is zero. If my terminology disturbs you, call η the "four-velocity." I should add that although proper velocity is the more convenient quantity to *calculate* with, ordinary velocity is still the more *natural* quantity from the point of view of an observer watching a particle fly past.

that momentum *is* conserved; that's a matter for *experiments* to decide. But it *does* say that if we're hoping to extend momentum conservation to the relativistic domain, we had better *not* define momentum as $m\mathbf{v}$, whereas $m\boldsymbol{\eta}$ is perfectly acceptable.

That's a tricky argument, and if you didn't follow it, try reading that last paragraph again. The upshot is that in relativity, momentum is defined as mass times *proper* velocity:

$$\mathbf{p} \equiv m\boldsymbol{\eta} \tag{3.36}$$

Since proper velocity is part of a four-vector, the same goes for momentum:

$$p^\mu = m\eta^\mu \tag{3.37}$$

The spatial components of p^μ constitute the (relativistic) momentum three-vector:

$$\mathbf{p} = \gamma m\mathbf{v} = \frac{m\mathbf{v}}{\sqrt{1 - v^2/c^2}} \tag{3.38}$$

Meanwhile, the "time" component is

$$p^0 = \gamma mc \tag{3.39}$$

For reasons that will appear in a moment, we define the "relativistic *energy*," E, as

$$E = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} \tag{3.40}$$

The zeroth component of p^μ, then, is E/c. Thus energy and momentum together make up a four-vector—the *energy-momentum four-vector:*

$$p^\mu = \left(\frac{E}{c}, p_x, p_y, p_z\right) \tag{3.41}$$

Incidentally, from equations (3.35) and (3.37) we have

$$p_\mu p^\mu = \frac{E^2}{c^2} - \mathbf{p}^2 = m^2c^2 \tag{3.42}$$

which, again, is manifestly invariant.

The relativistic momentum (3.38) reduces to the classical expression in the nonrelativistic regime ($v \ll c$), but the same cannot be said for relativistic energy (3.40). To see how this quantity comes to be called "energy," we expand the radical in a Taylor series:

$$E = mc^2\left(1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + \cdots\right) = mc^2 + \frac{1}{2}mv^2 + \frac{3}{8}m\frac{v^4}{c^2} + \cdots \tag{3.43}$$

Notice that the *second* term here corresponds to the classical kinetic energy, while the leading term (mc^2) is a constant. Now you may recall that in classical mechanics only *changes* in energy are physically significant—you can add a constant with impunity. In this sense the relativistic formula *is* consistent with the classical one, in the limit $v \ll c$ where the higher terms in the expansion are negligible. The constant term, which survives even when $v = 0$, is called the *rest energy;*

$$R \equiv mc^2 \tag{3.44}$$

the remainder, which is energy attributable to the motion of the particle, is the *relativistic kinetic energy:*

$$T \equiv mc^2(\gamma - 1) = \frac{1}{2} mv^2 + \frac{3}{8} m \frac{v^4}{c^2} + \cdots \tag{3.45}$$

(Notice that I have never mentioned *relativistic mass* in all this. It is a superfluous quantity that serves no useful function. In case you encounter it, the definition is $m_{\text{rel}} \equiv \gamma m$; it has died out because it differs from E only by a factor of c^2. Whatever can be said about m_{rel} could just as well be said about E, for instance, the "conservation of relativistic mass" is nothing but conservation of *energy*, with a factor of c^2 divided out.)

In classical mechanics there is no such thing as a massless particle; its momentum ($m\mathbf{v}$) would be zero, its kinetic energy ($\frac{1}{2}mv^2$) would be zero, it could sustain no force, since $F = ma$—it would be a dynamical cipher. At first glance you might suppose that the same would be true in relativity, but a careful inspection of the formulas

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - v^2/c^2}}, \qquad E = \frac{mc^2}{\sqrt{1 - v^2/c^2}} \tag{3.46}$$

reveals a loophole: When $m = 0$ the numerators are zero, but if $v = c$, the denominators *also* vanish, and these equations are indeterminate (0/0). So it is just possible that *we could allow $m = 0$, provided the particle always travels at the speed of light.* In this case equations (3.46) will not serve to define E and $\mathbf{p}$; nevertheless, equation (3.42) presumably still applies, so that

$$E = |\mathbf{p}|c \tag{3.47}$$

for massless particles. Personally, I would regard this "argument" as a joke, were it not for the fact that at least two types of massless particles (the photon and the neutrinos) are known to exist in nature. They do indeed travel at the speed of light, and their energy and momentum *are* related by equation (3.47). So evidently we must take the loophole seriously. You may well ask: If equations (3.46) do not define $\mathbf{p}$ and E, what *does* determine the momentum and energy of a massless particle? Not the mass (that's zero by assumption); not the speed (that's always c). How, then, *does* a photon with an energy of 2 eV differ from a photon with an energy of 3 eV? Relativity offers no answer to this question, but curiously enough *quantum* mechanics *does*, in the form of Planck's formula:

$$E = h\nu \tag{3.48}$$

It is the *frequency* of the photon that determines its energy and momentum: The 2 eV photon is *red*, and the 3 eV photon is *purple!*

3.4 COLLISIONS

The reason for introducing energy and momentum is, of course, that these quantities are *conserved* in any physical process. In relativity, as in classical mechanics, the cleanest application of these conservation laws is to *collisions*. Imagine first a classical collision, in which object A hits object B (perhaps they are both carts on an air table), producing objects C and D. (See Fig. 3.2.) Of course, C and D might be the same as A and B; but we may as well allow that some paint (or whatever) rubs off A onto B, so that the final masses are not the same as the original ones. (We *do* assume, however, that A, B, C, and D are the only actors in the drama; if some wreckage, W, is left at the scene, then we would be talking about a more complicated process: $A + B \rightarrow C + D + W$.) By its nature, a collision is something that happens so fast that no *external* force, such as gravity, or friction with the track, has an appreciable influence. Classically, mass and momentum are always conserved in such a process; kinetic energy may or may not be conserved.

> ***Classical Collisions***
> 1. Mass is conserved, $m_A + m_B = m_C + m_D$.
> 2. Momentum is conserved, $\mathbf{p}_A + \mathbf{p}_B = \mathbf{p}_C + \mathbf{p}_D$.
> 3. Kinetic energy may or may not be conserved.

In fact, we may distinguish three types of collisions: "sticky" ones, in which the kinetic energy *decreases* (typically, it is converted into heat); "explosive" ones, in which the kinetic energy *increases* (for example, suppose A has a compressed spring on its front bumper, and the catch is released in the course of the collision so that spring energy is converted into kinetic energy); and *elastic* ones, in which the kinetic energy is conserved.

> ***Types of Collisions (Classical)***
> **(a)** *Sticky:* Kinetic energy decreases, $T_A + T_B > T_C + T_D$.
> **(b)** *Explosive:* Kinetic energy increases, $T_A + T_B < T_C + T_D$.
> **(c)** *Elastic:* Kinetic energy conserved, $T_A + T_B = T_C + T_D$.

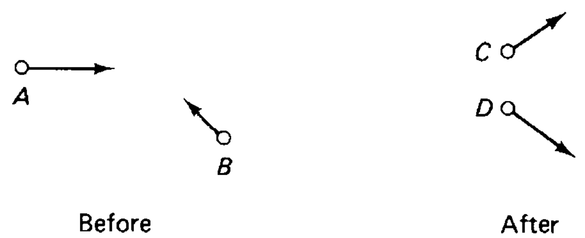


Figure 3.2 A collision in which $A + B \rightarrow C + D$.

In the extreme case of type (a), the two particles stick together, and there is really only *one* final object: $A + B \to C$. In the extreme case of type (b), a single object breaks in two: $A \to C + D$ (in the language of particle physics, A *decays* into $C + D$).

In a relativistic collision, *energy and momentum are always conserved.* In other words all four components of the energy-momentum four-vector are conserved. As in the classical case, *kinetic* energy may or may not be conserved.

Relativistic Collisions

1. Energy is conserved, $E_A + E_B = E_C + E_D$.
2. Momentum is conserved $\mathbf{p}_A + \mathbf{p}_B = \mathbf{p}_C + \mathbf{p}_D$.

$$\Rightarrow p_A^\mu + p_B^\mu = p_C^\mu + p_D^\mu$$

3. Kinetic energy may or may not be conserved.

Again, we may classify collisions as sticky, explosive, or elastic, depending on whether the kinetic energy decreases, increases, or remains the same. Since the *total* energy (rest plus kinetic) is *always* conserved, it follows that rest energy (and hence also mass) increases in a sticky collision, decreases in an explosive collision, and is unchanged in an elastic collision.

Types of Collisions (Relativistic)

(a) *Sticky:* Kinetic energy decreases, rest energy and mass increase.
(b) *Explosive:* Kinetic energy increases, rest energy and mass decrease.
(c) *Elastic:* Kinetic energy, rest energy, and mass are conserved.

Please note: *Except in elastic collisions, mass is not conserved;** conversely, if mass *is* conserved, the collision is elastic. In an explosive collision (or a particle decay), rest energy is converted into kinetic energy (or, in the absurd language of the popular press, infuriating to anyone with the slightest respect for dimensional consistency, "mass is converted into energy").

In spite of a certain structural parallel between the classical and relativistic analyses, there is a striking difference in the interpretation of inelastic collisions. In the classical case we say that energy is converted from kinetic form to some "internal" form (heat energy, spring energy, etc.), or vice versa. In the relativistic analysis we say that it goes from kinetic energy to *rest* energy, or vice versa. How can these possibly be consistent? After all, relativistic mechanics is supposed to reduce to classical mechanics in the limit $v \ll c$. The answer is that all "internal" forms of energy are reflected in the rest energy of an object. A hot potato *weighs more* than a cold potato; a compressed spring *weighs more* than a relaxed spring. On the macroscopic scale, rest energies are enormously greater than internal energies, so these mass differences are utterly negligible in everyday life, and very small even at the atomic level. Only in nuclear and particle physics are typical internal energies comparable to typical rest energies. Nevertheless, in principle, whenever you weigh an object, you are measuring not only the masses of its constituent parts, but all of their interaction energies as well.

* In the old terminology we would say that *relativistic* mass is conserved, but *rest* mass is not.

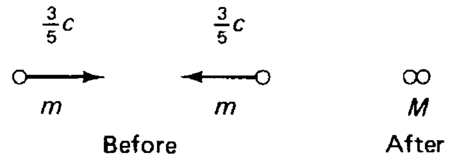


Figure 3.3 Sticky collision of two equal masses (Example 3.1).

3.5 EXAMPLES AND APPLICATIONS

Solving problems in relativistic kinematics is as much an art as a science. Although the *physics* involved is minimal—nothing but conservation of energy and conservation of momentum—the *algebra* can be formidable. Whether a given problem takes two lines or seven pages depends a lot on how skillful and experienced you are at manipulating the tools and the tricks of the trade. I now propose to work a few examples, pointing out as I go along some of the labor-saving devices that are available to you.[2]

EXAMPLE 3.1

Two lumps of clay, each of mass m, collide head-on at $\frac{3}{5}c$ (Fig. 3.3). They stick together. Question: What is the mass M of the final composite lump?

Question: What is the mass, M of the final composite lump?

Solution. Conservation of energy says $E_1 + E_2 = E_M$. Conservation of momentum says $\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_M$. In this case conservation of momentum is trivial: $\mathbf{p}_1 = -\mathbf{p}_2$, so the final lump is at rest (which was obvious from the start). The initial energies are equal, so conservation of energy yields

$$Mc^2 = 2E_m = \frac{2mc^2}{\sqrt{1-(3/5)^2}} = \frac{5}{4}(2mc^2)$$

Conclusion: $M = \frac{5}{2}m$. Notice that this is *greater* than the sum of the initial masses; in sticky collisions kinetic energy is converted into rest energy, so the mass increases.

EXAMPLE 3.2

A particle of mass M, initially at rest, decays into two pieces, each of mass m (Fig. 3.4). Question: What is the speed of each piece as it flies off?

Solution. This is, of course, the reverse of the process in Example 3.1. Conservation of momentum just says that the two lumps fly off in opposite directions at equal speeds. Conservation of energy requires that

$$M = \frac{2m}{\sqrt{1 - v^2/c^2}}, \quad \text{so} \quad v = c\sqrt{1-(2m/M)^2}$$

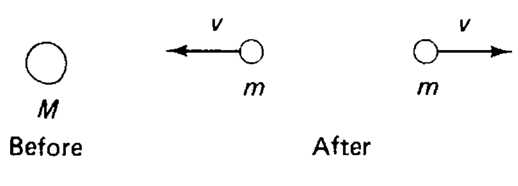


Figure 3.4 A particle decays into two equal pieces. (Example 3.2).

This answer makes no sense unless M exceeds $2m$; there has to be at least enough rest energy available to cover the rest energies in the final state (any *extra* is fine; it can be soaked up in the form of kinetic energy). We say that $M = 2m$ is the *threshold* for the process $M \longrightarrow 2m$ to occur. The deuteron, for example, is below the threshold for decay into proton plus neutron ($m_d = 1875.6$ Mev/c^2; $m_p + m_n = 1877.9$ MeV/c^2), and therefore is stable. A deuteron can be *pulled* apart, but only by pumping enough energy into the system to make up the difference. (If it puzzles you that a bound state of p and n should weigh *less* than the sum of its parts, the point is that the binding energy of the deuteron, which, like all internal energy, is reflected in its rest mass, is *negative*. Indeed, for *any* stable bound state the binding energy must be negative; if the composite particle weighs *more* than the sum of its constituents, it will spontaneously disintegrate.)

EXAMPLE 3.3

A pion at rest decays into a muon plus a neutrino (Fig. 3.5). Question: What is the speed of the muon?

Solution. Conservation of energy requires $E_\pi = E_\mu + E_\nu$. Conservation of momentum gives $\mathbf{p}_\pi = \mathbf{p}_\mu + \mathbf{p}_\nu$; but $\mathbf{p}_\pi = 0$, so $\mathbf{p}_\mu = -\mathbf{p}_\nu$. Thus the muon and the neutrino fly off back-to-back, with equal and opposite momenta.

To proceed, we need a formula relating the energy of a particle to its momentum; equation (3.42) does the job. [You might have been inclined to solve equation (3.38) for the velocity, and plug the result into equation (3.40). But that would be very poor strategy. In general, velocity is a bad parameter to work with, in relativity. Better to use equation (3.42), which takes you *directly* back and forth between E and $\mathbf{p}$.]

Suggestion 1. To get the energy of a particle, when you know its momentum (or vice versa), use the invariant

$$E^2 - \mathbf{p}^2c^2 = m^2c^4 \tag{3.49}$$

In the present case, then:

$$E_\pi = m_\pi c^2$$
$$E_\mu = c\sqrt{m_\mu^2c^2 + \mathbf{p}_\mu^2}$$
$$E_\nu = |\mathbf{p}_\nu|c = |\mathbf{p}_\mu|c$$

Putting these into the equation for conservation of energy, we have

$$m_\pi c^2 = c\sqrt{m_\mu^2c^2 + \mathbf{p}_\mu^2} + |\mathbf{p}_\mu|c$$

or

$$(m_\pi c - |\mathbf{p}_\mu|)^2 = m_\mu^2c^2 + \mathbf{p}_\mu^2$$

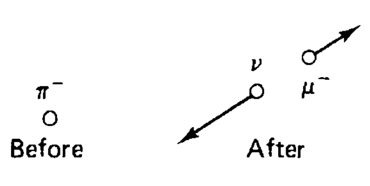


Figure 3.5 Decay of the charged pion (Example 3.3).